Sparse Signal Recovery Using Structured Total Maximum Likelihood

Jun-Jie Huang, and Pier Luigi Dragotti
Department of Electrical and Electronic Engineering, Imperial College London
South Kensington, London SW7 2AZ, UK
{j.huang15, p.dragotti}@imperial.ac.uk

***Abstract*—In this paper, we consider the sparse signal recovery problem when the dictionary is a Fourier frame. Based on the annihilation relation, the sparse signal recovery from noisy observations is posed as a structured total maximum likelihood (STML) problem. The recent structured total least squares (STLS) approach for finite rate of innovation signal recovery can be viewed as a particular version of our method. We transform the STML problem which has an additional log-det term into a form similar to the STLS problem. It can be effectively tackled using an iterative quadratic maximum likelihood like algorithm. From simulation results, our proposed STML approach outperforms the STLS based algorithm and the state-of-the-art sparse recovery algorithms.**

***Index Terms*—Sparse representation, Finite Rate of Innovation, Structured Total Least Squares, Structured Total Maximum Likelihood**

I. Introduction

Consider a standard sparse representation problem where the goal is to find a K-sparse signal $\boldsymbol{x} \in \mathbb{C}^M$ from noisy observation $\boldsymbol{y} \in \mathbb{C}^N$:

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{x} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{D} \in \mathbb{C}^{N \times M}$ is the dictionary with $N < M$, $||\boldsymbol{x}||_0 = K$ with $||\boldsymbol{x}||_0 \stackrel{\text{def}}{=} \#\{n : |x[n]| \neq 0\}$, and $\boldsymbol{n}$ is complex-valued additive white Gaussian noise.

As l_0 norm is not convex, the original problem is intractable. Instead, convex relaxation methods, such as Basis Pursuit (BP) [1] and LASSO [2], relax the non-convex l_0 norm to a convex l_1 norm. Greedy methods iteratively update non-zero elements of the solution based on the correlation between the residual signal and the dictionary. The state-of-the-art greedy algorithms include Orthogonal Matching Pursuit (OMP) [3], [4], [5], [6], Subspace Pursuit (SP) [7], and Compressive Sampling Matching Pursuit (CoSAMP) [8].

Sparse signal recovery is also related to the finite rate of innovation (FRI) theory [9], [10], [11], [12] and the super-resolution for line spectral estimation [13], [14], [15]. Both approaches focus on continuous time sparse signals. In particular, FRI theory shows that perfect reconstruction can be achieved for classes of non-bandlimited signals including streams of pulses, piecewise sinusoidal and piecewise polynomial signals. When the dictionary $\boldsymbol{D}$ is a Fourier frame or a Gabor frame, the observed signal is a superpostion of complex exponentials. As the columns of the dictionary are highly coherent, the conventional sparse recovery algorithms may not be applicable. However, it is possible to solve the sparse representation problem using variations of FRI signal recovery algorithms [10], [16], [17], [18]. The sparse representation problem can be considered as a discretized FRI signal recovery problem where the pulse locations can only be on a uniform grid.

Recently, a model fitting based algorithm [19] tries to recover the FRI signal using structured total least squares (STLS) [20], [21]. The FRI model fitting problem is formulated as minimizing the residual in the noisy observation with a low rank constraint represented by an annihilation relation. It achieves the best performance in FRI signal reconstruction compared with state-of-the-art algorithms [10], [16], [17], [22]. At the same time, Beck and Eldar [23] have proposed a structured total maximum likelihood (STML) method as a way to solve structured least square problems. They solved it by using BFGS algorithm which is an iterative gradient descent based method. However, the computational complexity is very high.

In this paper, we try to solve the sparse signal recovery problem when the dictionary is a Fourier frame using the STML framework. With an additional log-det term, the STML becomes more difficult to solve. Inspired by the linear constraint proposed in [19], we transform the STML formulation with the log-det term into a form similar to that of STLS. Using an iterative algorithm similar to the iterative quadratic maximum likelihood (IQML) [20] method, the sparse signal can be effectively estimated using our proposed STML method. From simulation results, our proposed STML algorithm achieves more robust performance compared with the STLS based methods and the state-of-the-art sparse recovery algorithms.

The rest of the paper is organized as follows: Section 2 formulates the problem of recovering sparse signal from noisy observations observed with a Fourier frame. Section 3 reviews STLS approach [19] for FRI signal recovery. Section 4 introduces our proposed sparse signal recovery with STML and a novel iterative algorithm. Section 5 presents simulation results and Section 6 draws conclusions.

II. Problem Formulation

We assume we observe $\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x}$ with $\boldsymbol{F} \in \mathbb{C}^{N \times M}$ being the Fourier frame and $\boldsymbol{x} \in \mathbb{C}^M$ is a K-sparse vector. Consequently, $\boldsymbol{y} \in \mathbb{C}^N$ is the sum of K exponentials:

$$y[n] = \frac{1}{\sqrt{N}} \sum_{k=0}^{K-1} a_k u_k^n, \tag{2}$$

where $0 \leq n < N$, $0 \leq m_0 < ... < m_{K-1} < M$, $u_k = e^{j\frac{2\pi}{N}m_k}$, and a_k are non-zero complex-valued weights. Hence, the indices m_k corresponds to the non-zero entries in $\boldsymbol{x}$ and a_k are the corresponding amplitudes.

Given observation $\boldsymbol{y}$, the sparse signal $\boldsymbol{x}$ is typically recovered using convex relaxation techniques [1], [2]. More recent work [24] has advocated the use of methods based on FRI theory [9], [10], [19], [25], [26] as an alternative to solve the sparsity recovery problem.

Central to FRI theory is the use of Prony's method. Prony's method is based on the observation that when $\boldsymbol{y}$ is a sum of exponentials as in Eqn. (2), then there exists an annihilating filter $\boldsymbol{h}$ such that:

$$\boldsymbol{y} * \boldsymbol{h} = 0. \tag{3}$$

The filter satisfying the above equality has z-transform:

$$H(z) \quad = \quad \prod_{k=0}^{K-1}(z - u_k). \tag{4}$$

That is, the roots of $H(z)$ correspond to the non-zero entries of $\boldsymbol{x}$. Therefore, the knowledge of $\boldsymbol{h}$ is sufficient to retrieve $\boldsymbol{x}$ if $2K \leq N$.

Eqn. (3) can be written in matrix form as follows:

$$\mathbf{B}(\boldsymbol{y})\boldsymbol{h} = \mathbf{0}, \tag{5}$$

where $\mathbf{B}(\boldsymbol{y}) \in \mathbb{C}^{(M-K)\times(K+1)}$ is the Toeplitz lifted matrix for observation $\boldsymbol{y}$ with $\mathbf{B}(\boldsymbol{y})_{m,i} = y[m-i]$.

In the noiseless case, $\mathbf{B}(\boldsymbol{y})$ is a rank-deficient matrix and has constant values along its diagonals. The annihilating filter $\boldsymbol{h} \in \mathbb{C}^{K+1}$ belongs to the null-space of $\mathbf{B}(\boldsymbol{y})$. However, if noise is present $\widetilde{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{n}$ then $\mathbf{B}(\widetilde{\boldsymbol{y}})$ becomes full rank. Here we assume complex-valued noise $\boldsymbol{n}$ with i.i.d. Gaussian distribution in both the real and the imaginary parts with zero mean and variance $\sigma_e^2/2$.

The objective is to denoise the Toeplitz matrix (i.e. find an approximated Toeplitz matrix which has minimum distance with $\mathbf{B}(\widetilde{\boldsymbol{y}})$ and satisfies the low rank property). With the corresponding annihilating filter, the non-zero elements in the sparse signal $\boldsymbol{x}$ can be identified by retrieving the roots of $H(z)$. The amplitudes a_k can then be estimated using least squares.

III. FRI Signal Recovery With STLS

In [19], FRI signal recovery is formulated as a model fitting problem using *structured total least squares* (STLS) [21]:

$$(\text{STLS}) : \min_{\boldsymbol{y},\boldsymbol{h}} ||\boldsymbol{y} - \widetilde{\boldsymbol{y}}||_2^2, s.t. \mathbf{B}(\boldsymbol{y})\boldsymbol{h} = 0, \, and \, ||\boldsymbol{h}||^2 = 1, \tag{6}$$

where $\widetilde{\boldsymbol{y}}$ and $\boldsymbol{y}$ are the observed noisy exponentials and the desired clean signal, and $\boldsymbol{h} \in \mathbb{C}^{K+1}$ is the annihilating filter for Toeplitz matrix $\mathbf{B}(\boldsymbol{y}) \in \mathbb{C}^{(M-K)\times(K+1)}$.

STLS is a non-convex problem. The constraints define a non-convex set due to the rank deficiency requirement on $\mathbf{B}(\boldsymbol{y})$. By introducing a vector-valued Lagrange multiplier $\boldsymbol{u} \in \mathbb{C}^{M-K}$ and a scalar Lagrange multiplier λ, the constrained optimization problem can be reformulated as an unconstrained one:

$$\min_{\boldsymbol{y},\boldsymbol{h}} \left\{ ||\boldsymbol{y} - \widetilde{\boldsymbol{y}}||^2 + 2\mathcal{R}\{\boldsymbol{u}^H \mathbf{B}(\boldsymbol{y})\boldsymbol{h}\} + \lambda(||\boldsymbol{h}||^2 - 1) \right\}, \tag{7}$$

where $\mathcal{R}(\cdot)$ represents the real part of the argument.

As $\mathbf{B}(\boldsymbol{y})\boldsymbol{h}$ represents $\boldsymbol{y} * \boldsymbol{h}$ and due to the commutativity property of convolution, the right dual matrix $\mathbf{R}(\cdot)$ of $\mathbf{B}(\cdot)$ is also a Toeplitz matrix and is defined as [19]:

$$\mathbf{B}(\boldsymbol{y})\boldsymbol{h} = \mathbf{R}(\boldsymbol{h})\boldsymbol{y}. \tag{8}$$

Let us further define $\mathbf{D}_{\boldsymbol{h}} = \mathbf{R}(\boldsymbol{h})\mathbf{R}(\boldsymbol{h})^H$. Based on the optimality conditions of Eqn. (7), the l_2 distance between $\boldsymbol{y}$ and $\widetilde{\boldsymbol{y}}$ can be expressed as:

$$||\boldsymbol{y} - \widetilde{\boldsymbol{y}}||^2 = \boldsymbol{h}^H \mathbf{S}(\boldsymbol{h})\boldsymbol{h}, \tag{9}$$

where $\mathbf{S}(\boldsymbol{h}) = \mathbf{B}(\widetilde{\boldsymbol{y}})^H \mathbf{D}_{\boldsymbol{h}}^{-1} \mathbf{B}(\widetilde{\boldsymbol{y}})$.

As $\mathbf{S}(\boldsymbol{h})$ depends on $\boldsymbol{h}$, this problem cannot be directly minimized. The iterative quadratic maximum likelihood (IQML) scheme [20], which assumes $\mathbf{S}(\boldsymbol{h})$ remains constant at each iteration and solves for $\boldsymbol{h}$, can be applied to find an approximate solution. Doğan *et al.* [19] propose to modify the quadratic constraint $||\boldsymbol{h}||^2 = 1$ to a linear constraint $\boldsymbol{l}^H\boldsymbol{h} = 1$ with $\boldsymbol{l} = [1, 0, ...0]^T$ which enforces the annihilating filter to be updated on a hyper-plane and excludes the need of solving a eigenvalue problem at each iteration. The annihilating filters update is then given by:

$$\boldsymbol{h}^{(i)} = \frac{\mathbf{S}(\boldsymbol{h}^{(i-1)})^{-1}\boldsymbol{l}}{\boldsymbol{l}^H\mathbf{S}(\boldsymbol{h}^{(i-1)})^{-1}\boldsymbol{l}}. \tag{10}$$

A randomized linear constraint where the coefficients of $\boldsymbol{l}$ are randomly drawn from $\mathcal{N}(0,1) + j\mathcal{N}(0,1)$ shows a superior performance compared to the linear constraint. The annihilating filter coefficients at the first iteration $\boldsymbol{h}^{(0)}$ are randomly generated. This randomized initialization scheme enables a higher flexibility and improves the robustness of the STLS algorithm.

IV. Sparse Signal Recovery With Structured Total Maximum Likelihood

Inspired by [23], a *structured total maximum likelihood* (STML) expression for sparse signal recovery is proposed in this section. It shows that STLS can be derived from STML and provides an easier derivation for STLS compared with that in [19]. Based on the (randomized) linear constraint scheme, we propose an effective IQML-like iterative algorithm for the STML formulation.

A. Structured Total Maximum Likelihood Formulation

By assumption, the residual $\boldsymbol{n} = \widetilde{\boldsymbol{y}} - \boldsymbol{y}$ is a zero mean normal random variable with variance σ_e^2 .

Proposition 1: If $\boldsymbol{h}$ is the optimal annihilating filter for $\mathbf{B}(\boldsymbol{y})$, $\mathbf{B}(\widetilde{\boldsymbol{y}})\boldsymbol{h}$ follows a normal distribution:

$$\mathbf{B}(\widetilde{\boldsymbol{y}})\boldsymbol{h} \sim \mathcal{N}\left(0, \sigma_e^2 \mathbf{D}_{\boldsymbol{h}}\right).$$

Algorithm 1 Sparse Signal Recovery with STML

Input: Noisy observation $\widetilde{\boldsymbol{y}}$, dictionary $\boldsymbol{D}$, and $K > 0$.
Output: K-sparse signal $\hat{\boldsymbol{x}}$

for $i = 1$ to $numberofinitializations$ **do**
    Randomly initialize $\boldsymbol{h}^{(0)}$, initialize $\boldsymbol{l}$, set $\sigma_e^2 = 0$
    for $j = 1$ to $numberofiterations$ **do**
        Build $\mathbf{S}_{\mathbf{E}}(\boldsymbol{h}^{(j-1)})$ as in Eqn.(15)
        Solve for $\boldsymbol{h}^{(j)}$ as in Eqn.(16)
        Estimate $\sigma_e^{(j)^2}$ as in Eqn.(17)
        Find the roots $\boldsymbol{u}^{(j)}$ of $\boldsymbol{h}^{(j)}$ as in Eqn.(4)
        Retrieve the K-sparse signal $\boldsymbol{x}^{(j)} = \mathcal{S}(\boldsymbol{u}^{(j)})$
        Reconstruction error $e^{(j)} = ||\widetilde{\boldsymbol{y}} - \boldsymbol{D}\boldsymbol{x}^{(j)}||^2$
        if $e^{(j)} < e_{min}$ **then**
            e_{min}=$e^{(j)}$
            $\hat{\boldsymbol{x}} = \boldsymbol{x}^{(j)}$
        end if
        if $|e^{(j-1)} - e^{(j)}|/e^{(j)} < \epsilon$ **then**
            Terminate the inner loop
        end if
    end for
end for

Proof: If the annihilating filter $\boldsymbol{h}$ is in the null-space of a rank-deficient matrix $\mathbf{B}(\boldsymbol{y})$, we can decompose $\mathbf{B}(\widetilde{\boldsymbol{y}})\boldsymbol{h}$ into $\mathbf{B}(\widetilde{\boldsymbol{y}})\boldsymbol{h} = \mathbf{B}(\boldsymbol{y})\boldsymbol{h} + \mathbf{B}(\boldsymbol{n})\boldsymbol{h} = \mathbf{R}(\boldsymbol{h})\boldsymbol{n}$.

$$Var(\mathbf{B}(\widetilde{\boldsymbol{y}})\boldsymbol{h}) = \mathbb{E}\{\mathbf{R}(\boldsymbol{h})\boldsymbol{n}\boldsymbol{n}^H\mathbf{R}(\boldsymbol{h})^H\} = \sigma_e^2\mathbf{D}_{\boldsymbol{h}}. \tag{11}$$

Thus, its mean and variance are $\mu(\mathbf{B}(\widetilde{\boldsymbol{y}})\boldsymbol{h}) = 0$, and $Var(\mathbf{B}(\widetilde{\boldsymbol{y}})\boldsymbol{h}) = \sigma_e^2\mathbf{D}_{\boldsymbol{h}}$, respectively. ■

The annihilating filter $\boldsymbol{h}$ which maximizes the likelihood of observation $\widetilde{\boldsymbol{y}}$ is:

$$\hat{\boldsymbol{h}}_{ML} = \underset{\boldsymbol{h}}{\operatorname{argmax}} f(\widetilde{\boldsymbol{y}}|\boldsymbol{h}), \tag{12}$$

where $f(\widetilde{\boldsymbol{y}}|\boldsymbol{h}) = \frac{\exp(\boldsymbol{h}^H\mathbf{B}(\widetilde{\boldsymbol{y}})^H\mathbf{D}_{\boldsymbol{e}}^{-1}\mathbf{B}(\widetilde{\boldsymbol{y}})\boldsymbol{h})}{\sqrt{\det(2\pi\boldsymbol{D}_{\boldsymbol{e}})}}$, and $\mathbf{D}_{\boldsymbol{e}} = \sigma_e^2\mathbf{D}_{\boldsymbol{h}}$.

The maximum likelihood estimator can be found by minimizing the negative log likelihood:

$$(\text{STML}) : \min_{\boldsymbol{h}}\{\boldsymbol{h}^H\mathbf{S}(\boldsymbol{h})\boldsymbol{h} + \sigma_e^2\text{logdet}(\sigma_e^2\mathbf{D}_{\boldsymbol{h}})\}. \tag{13}$$

Compared with Eqn.(9), there is a logarithmic regularization term in Eqn. (13) which could serve to stabilize the solution. If $\mathbf{D}_{\boldsymbol{e}}$ in the log-det term is regarded as a constant during each iteration, the optimization for STML and STLS become the same. As the log-det will be zero when there is no noise (i.e. $\sigma_e^2 = 0$), the STML formulation in Eqn. (13) reduces to the STLS formulation in noiseless case. Bresler and Macovski [20] gave the insight that the STLS formulation in Eqn. (9) is the negative log likelihood for $\boldsymbol{w} = \mathbf{D}_{\boldsymbol{h}}^{-\frac{1}{2}}\mathbf{B}(\widetilde{\boldsymbol{y}})\boldsymbol{h}$ which follows $\mathcal{N}(0, \sigma_e^2)$, and it is not injective for the mapping from $\widetilde{\boldsymbol{y}}$ to $\boldsymbol{w}$. The conditional density of $\widetilde{\boldsymbol{y}}$ could be different from that of $\boldsymbol{w}$, while our STML formulation considers the direct negative log likelihood for $\widetilde{\boldsymbol{y}}$.

B. Iterative Algorithm for STML

The log-det term is smooth yet non-convex and usually taken as the surrogate of rank. With one more non-convex term, STML seems more difficult to solve. We propose an IQML-like iterative algorithm using the (randomized) linear constraint. Based on the negative log likelihood obtained in Eqn. (13), and the linear constraint $\boldsymbol{l}^H\boldsymbol{h} = 1$, we modify the STML formula as follows:

$$\begin{aligned}&\boldsymbol{h}^H\mathbf{S}(\boldsymbol{h})\boldsymbol{h} + \sigma_e^2\text{logdet}(\sigma_e^2\mathbf{D}_{\boldsymbol{h}})\\ =&\boldsymbol{h}^H\left\{\mathbf{S}(\boldsymbol{h}) + \boldsymbol{l}\boldsymbol{l}^H\sigma_{\text{e}}^2\text{logdet}(\sigma_e^2\mathbf{D}_{\boldsymbol{h}})\right\}\boldsymbol{h}.\end{aligned} \tag{14}$$

The regularized optimization problem has been transformed into something similar to Eqn. (9) in STLS method:

$$\min_{\boldsymbol{h}}\boldsymbol{h}^H\mathbf{S}_{\mathbf{E}}(\boldsymbol{h})\boldsymbol{h}, s.t. \boldsymbol{l}^H\boldsymbol{h} = 1, \tag{15}$$

where $\mathbf{S}_{\mathbf{E}}(\boldsymbol{h}) = \mathbf{S}(\boldsymbol{h}) + \mathbf{E}(\boldsymbol{h})$ with $\mathbf{E}(\boldsymbol{h}) = \boldsymbol{l}\boldsymbol{l}^H\sigma_e^2\text{logdet}(\sigma_e^2\mathbf{D}_{\boldsymbol{h}})$ which is a weighted rank-1 matrix $\boldsymbol{l}\boldsymbol{l}^H$ weighted by the log-det term. Since $\text{logdet}(\mathbf{D}_{\boldsymbol{h}})$ is not invariant to the magnitude of $\boldsymbol{h}$, we use the normalized $\hat{\boldsymbol{h}} = \boldsymbol{h}/||\boldsymbol{h}||^2$ to construct $\mathbf{D}_{\hat{\boldsymbol{h}}}$ for $\mathbf{S}_{\mathbf{E}}(\boldsymbol{h})$.

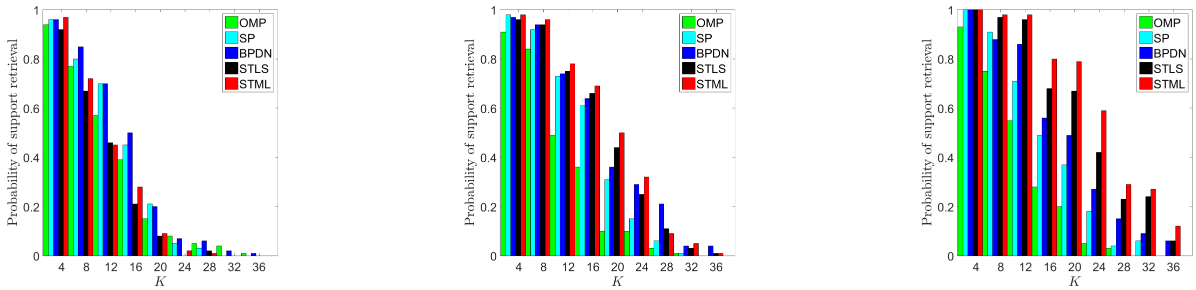
Similar to IQML algorithm [20] and Doğan *et al.*'s approach [19], we use an iterative method to consistently improve estimation for $\boldsymbol{h}$ by assuming that $\mathbf{S}_{\mathbf{E}}(\boldsymbol{h})$ remains constant during each iteration and is constructed using previous estimates of $\boldsymbol{h}$. The updated annihilating filter $\boldsymbol{h}^{(i)}$ at iteration i is therefore given by:

$$\boldsymbol{h}^{(i)} = \frac{\mathbf{S}_{\mathbf{E}}(\boldsymbol{h}^{(i-1)})^{-1}\boldsymbol{l}}{\boldsymbol{l}^H\mathbf{S}_{\mathbf{E}}(\boldsymbol{h}^{(i-1)})^{-1}\boldsymbol{l}}. \tag{16}$$

As σ_e^2 corresponds to the variance of the residual signal which cannot be annihilated by $\boldsymbol{h}$, the noise variance at iteration i can be updated using Eqn. (17). This helps our algorithm get rid of the requirement of a prior knowledge on σ_e^2:

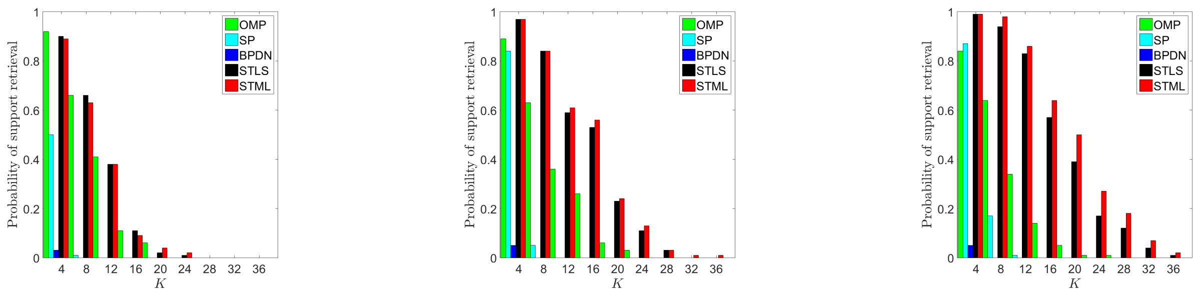
$$\sigma_e^{(i)^2} \quad = \quad \boldsymbol{h}^{(i-1)^H}\mathbf{S}(\boldsymbol{h}^{(i-1)})\boldsymbol{h}^{(i-1)}. \tag{17}$$

Algorithm 1 summarizes our proposed iterative approach for sparse signal recovery using structured total maximum likelihood. In [19], the use of multiple random initializations increases the probability of finding a good annihilating filter and enhances the robustness of the algorithm. The same strategy is also used here. For each random initialization, the initial annihilating filter is randomly generated. At each iteration, an annihilating filter is estimated according to (16) by using the filter estimated during the previous iteration. With the updated filter, its roots can be retrieved and the corresponding sparse signal can be estimated. Let us denote $\mathcal{S}(\boldsymbol{u})$ as the corresponding K-sparse signal of the roots $\boldsymbol{u}$. The reconstruction error has been taken as the selection criterion for the best K-sparse signal $\hat{\boldsymbol{x}}$. The inner loop will be terminated if the percentage of the difference between consecutive reconstruction error is smaller than a small number ϵ. By default, we set the number of initializations to 5, the



(a) SNR = 20 dB (b) SNR = 30 dB (c) SNR = 40 dB

Fig. 1. Probability of exact support retrieval of different algorithms when the dictionary is a Fourier frame with size of 128×256.



(a) SNR = 20 dB (b) SNR = 30 dB (c) SNR = 40 dB

Fig. 2. Probability of exact support retrieval of different algorithms when the dictionary is a Fourier frame with size of 128×512.

number of iterations to 50, and $\epsilon = 0.01$. If linear constraint is applied, we set $\boldsymbol{l} = [1, 0, ...0]^T$. For randomized linear constraint, $\boldsymbol{l}$ is randomly generated.

V. Simulation Results

We compare our proposed STML algorithm with the STLS based method [19] and the state-of-the-art sparse recovery algorithms including OMP, SP, and BPDN. We implemented STLS and OMP. The code of SP used for testing was downloaded from author's website. BPDN was realized using CVX package. Randomized linear constraint is applied for both STLS and STML as it provides better performance than linear constraint. For each sparsity level and (signal-to-noise ratio) SNR, 100 different sparse signals have been generated. The SNR (dB) is defined as $10\log_{10}(\frac{1}{M}||\boldsymbol{y}||^2)/\sigma^2$. Fig. 1 and Fig. 2 show the simulation results on the probability of retrieving the exact sparse signal by different algorithms. The dictionary is a Fourier frame with size 128×256 and size 128×512 in Fig.1 and in Fig.2, respectively. The amplitudes of the sparse signal are drawn from $\mathcal{N}(0, 1)$ for both real and imaginary part. Complex additive white Gaussian noise, which follows $\mathcal{N}(0, \sigma^2/2) + j\mathcal{N}(0, \sigma^2/2)$, is added into the clean superimposed exponential samples. Three different SNR scenarios were evaluated, including 20 dB, 30 dB and 40 dB.

In Fig. 1, the dictionary is a $2\times$ over-complete DFT matrix. All methods have a high probability of success when the sparsity level is low. Their performances will be significantly deteriorated as the sparsity level increases. In general, BPDN has better performance than OMP and SP. STML outperforms STLS in most cases and leads other methods with a large margin at SNR = 30 dB and 40 dB. At lower SNR, it is not as good as the conventional algorithms. In Fig. 2, the dictionary is a $4\times$ over-complete DFT matrix and has stronger coherence. Due to the increased coherence, performance of methods based on convex relaxation deteriorates. BPDN can hardly recover the correct sparse signal even at very low sparsity level. We found that the exact support retrieval probability of OMP gradually decreases as the sparsity level increases while SP has a sudden drop on the probability of support retrieval when the sparsity level exceeds 8. In general, STML achieves the best performance in most cases.

VI. Conclusions

In this paper, we considered the sparse signal recovery problem over an over-complete Fourier frame. A novel STML method has been presented for sparse signal recovery leveraging the idea from FRI signal recovery. It can be interpreted as a generalization of the recent STLS based method. Based on a linear constraint, STML problem has been converted into a similar form as the STLS which could be iteratively solved using an IQML-like method. From simulation results, our proposed STML method outperforms the state-of-the-art algorithms for highly coherent dictionaries.

For future works, the STML algorithm should be further improved at low SNR scenarios.

Acknowlegment

This work is supported by the European Research Council (ERC) starting investigator award Nr. 277800 (RecoSamp).

References

[1] Scott Shaobing Chen, David L Donoho, and Michael A Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[2] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[3] Stéphane G Mallat and Zhifeng Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on signal processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[4] Joel A Tropp and Anna C Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.

[5] Deanna Needell and Roman Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of computational mathematics*, vol. 9, no. 3, pp. 317–334, 2009.

[6] David L Donoho, Yaakov Tsaig, Iddo Drori, and Jean-Luc Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1094–1121, 2012.

[7] Wei Dai and Olgica Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.

[8] Deanna Needell and Joel A Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[9] Martin Vetterli, Pina Marziliano, and Thierry Blu, "Sampling signals with finite rate of innovation," *IEEE transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, 2002.

[10] Thierry Blu, Pier-Luigi Dragotti, Martin Vetterli, Pina Marziliano, and Lionel Coulot, "Sparse sampling of signal innovations," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 31–40, 2008.

[11] Pier Luigi Dragotti, Martin Vetterli, and Thierry Blu, "Exact sampling results for signals with finite rate of innovation using strang-fix conditions and local kernels," in *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*. IEEE, 2005, vol. 4, pp. iv–233.

[12] Jose Antonio Urigiien, Yonina C Eldar, P Dragotti, et al., "Sampling at the rate of innovation: Theory and applications," *Compressed Sensing: Theory and Applications*, p. 148, 2012.

[13] Emmanuel J Candès and Carlos Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[14] Emmanuel J Candès and Carlos Fernandez-Granda, "Super-resolution from noisy data," *Journal of Fourier Analysis and Applications*, vol. 19, no. 6, pp. 1229–1254, 2013.

[15] Gongguo Tang, Badri Narayan Bhaskar, and Benjamin Recht, "Near minimax line spectral estimation," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 499–512, 2015.

[16] James A Cadzow, "Signal enhancement-a composite property mapping algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 1, pp. 49–62, 1988.

[17] Yingbo Hua and Tapan K Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 5, pp. 814–824, 1990.

[18] Pier Luigi Dragotti and Yue M Lu, "On sparse representation in Fourier and local bases," *IEEE Transactions on Information Theory*, vol. 60, no. 12, pp. 7888–7899, 2014.

[19] Zafer Doğan, Christopher Gilliam, Thierry Blu, and Dimitri Van De Ville, "Reconstruction of finite rate of innovation signals with model-fitting approach," *IEEE Transactions on Signal Processing*, vol. 63, no. 22, pp. 6024–6036, 2015.

[20] Yoram Bresler and Albert Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 5, pp. 1081–1089, 1986.

[21] Bart De Moor, "Total least squares for affinely structured matrices and the noisy realization problem," *IEEE Transactions on Signal Processing*, vol. 42, no. 11, pp. 3104–3113, 1994.

[22] Donald W Tufts and Ramdas Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 975–989, 1982.

[23] Amir Beck and Yonina C Eldar, "Structured total maximum likelihood: An alternative to structured total least squares," *SIAM Journal on Matrix Analysis and Applications*, vol. 31, no. 5, pp. 2623–2649, 2010.

[24] Jon Onativia, Yue M Lu, and Pier Luigi Dragotti, "Finite dimensional FRI," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 1808–1812.

[25] Pier Luigi Dragotti, Martin Vetterli, and Thierry Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets strang–fix," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1741–1757, 2007.

[26] Jose Antonio Urigüen, Thierry Blu, and Pier Luigi Dragotti, "FRI sampling with arbitrary kernels," *IEEE Transactions on Signal Processing*, vol. 61, no. 21, pp. 5310–5323, 2013.