



Structured dictionary optimization: application to the modulated wideband converter

Hassan Mortada, Olivier Rabaste, Jonathan Bosse

► To cite this version:

Hassan Mortada, Olivier Rabaste, Jonathan Bosse. Structured dictionary optimization: application to the modulated wideband converter. IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP 2019), Dec 2019, Le Gosier, France. <hal-02906394>

HAL Id: hal-02906394

<https://hal.science/hal-02906394v1>

Submitted on 24 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Structured dictionary optimization: application to the modulated wideband converter

Hassan Mortada, Olivier Rabaste and Jonathan Bosse
ONERA, the French Aerospace Lab
91123 Palaiseau Cedex, France
Emails: surname.name@onera.fr

Abstract—In this paper we consider the minimization of dictionary mutual coherence for the Modulated Wideband Converter (MWC) architecture. This consists in optimizing the selection of the periodic sequences which are used to alias the spectrum of the input signal. In particular, we consider here the specific setting when the filter cutoff and sampling frequency is increased in order to obtain multiple mixtures of the aliased bands in each physical channel. This setting implies a structured dictionary yielding a challenging optimization problem which has not been covered in the literature. We propose and extend methods based on ℓ_2 and ℓ_∞ norms minimization methods to solve this particular problem.

I. INTRODUCTION

Compressed sensing framework provides an effective solution to sample wideband and sparse spectrum, *i.e.*, signals that have large bandwidth with only a few and unknown active bands that contain the information of interest. Such signals can be found in various applications such as in radar [1], medical imaging [2] and digital communications [3].

Sampling such signals may present a challenge. Indeed, the Nyquist sampling frequency (at least twice the bandwidth of a real signal) may exceed the capability of conventional analog-to-digital converters. Moreover, for sparse spectrum signals, Nyquist frequency presents a waste in hardware and computational costs and resources. Some strategies such as signal demodulation and non-uniform sampling [4] have been proposed to deal with this challenge and sample with a sub-Nyquist frequency. However, they require the knowledge, in advance, of the active bands.

The Modulated Wideband Converter (MWC) [5] strategy yields a sampling rate that is much lower than the Nyquist frequency without any knowledge on the active bands. This strategy is based on the compressed sensing framework [6] and is composed of two stages: sub-Nyquist signal acquisition and signal recovery. The first stage consists in splitting the input analog signal into I physical channels. In each channel i , the signal is multiplied by a T_p -periodic function. The resulting signals are then low-pass filtered with a cutoff frequency $f_s/2$ and sampled with sub-Nyquist frequency f_s . Indeed, the multiplication with a periodic function is used to alias the spectrum such that mixtures of the spectrum bands with bandwidth $f_p = 1/T_p$ are obtained around the frequencies which are multiples of f_p . An MWC architecture and example of spectrum aliasing is given in Figure 1.

The signal recovery stage consists in applying a sparse reconstruction method [7] to find the relatively small number

of active bands compared to the total band number. The signal recovery quality may be enhanced through two factors. A first factor is to increase the number of mixtures, allowing thus to recover more bands. This can be done by adding more physical channels. However, this solution is costly and increases the hardware burden. An alternative solution comes at the price of increasing the filter cutoff and sampling frequency f_s , *i.e.* $f_s = Q' f_p$, such that Q' mixtures are obtained in each channel. A second factor to enhance the signal recovery quality is to generate periodic functions such that the resulting dictionary (containing the Fourier coefficient of a sequence period) is as incoherent as possible. This allows to better separate the aliased bands and retrieve the active bands effectively. Multiple methods have been proposed in the literature to address this problem [8]–[10]. However, these methods cannot be directly applied to the multiple mixtures per channel case (when $Q' > 1$) since the dictionary has, in that case, a more specific and complex structure.

In this paper we consider both factors to enhance the signal recovery quality. To the best of our knowledge, sequence optimization has never been addressed in the multiple mixtures per channel case. We thus propose here to extend previously cited methods in order to deal with this more challenging case. This paper is organized as follows. In section II, we provide a novel problem formulation in case of multiple mixtures per channel. Then in Section III, we present three methods to optimize the selection of the periodic sequences. Finally, results and simulations are presented in Section IV to illustrate the effectiveness of the proposed methods on simulated wideband radar data.

The notations are as follows. The l th column of a matrix \mathbf{M} is denoted as \mathbf{M}_l , while the elements are denoted as M_{ml} where m and l are respectively the row and column indices.

II. PROBLEM FORMULATION

We consider an MWC setting with I physical channels, an ideal low-pass filter with cutoff frequency $f_s/2$ and sampling frequency f_s . Moreover, the sampling frequency is chosen as $f_s = Q' f_p$ with odd $Q' = 2Q + 1$, providing thus Q' band mixtures per channel.

The input analog signal $x(t)$ with Nyquist rate f_{NYQ} and J active bands is multiplied in each channel i by a periodic function $p_i(t) = \sum_{m=0}^{M-1} s_{im} \Pi\left(t - \frac{mT_p}{M}\right)$ where M is the sequence period length and $\Pi(t)$ is a rectangular window of length T_p/M . The Fourier transform of the analog

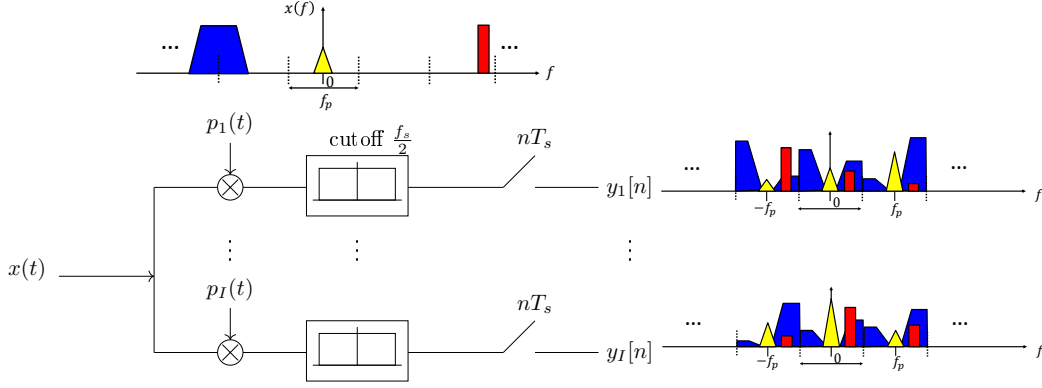


Fig. 1: The modulated wideband converted scheme. On the top, an example of a sparse spectrum with $J = 3$ active bands. On the right, three mixtures of aliased spectra in each channel.

multiplication $x(t) \cdot p_i(t)$ after filtering and mixture extracting is given as [5, equation (23)]:

$$y_i(f + qf_p) = \sum_{l=-L_0}^{L_0} c_{i,(l+q)} X(f - lf_p), \quad f \in \left[-\frac{f_p}{2}, \frac{f_p}{2}\right], \quad (1)$$

where $-Q \leq q \leq Q$ is the mixture index, $X(f - lf_p)$ is the l th band with bandwidth f_p and c_{il} is the l th Fourier coefficient corresponding to the periodic sequence of channel i . The total number of bands $L = 2L_0 + 1$ is determined such that the sum contains all nonzero shifted bands over $[-f_s/2, f_s/2]$, i.e. $L_0 = \left\lceil \frac{f_{NYQ} + f_s}{2f_p} \right\rceil - 1$ where $\lceil \cdot \rceil$ is the ceiling operator.

Furthermore, for each value of q , equation (1) reads:

$$\mathbf{y}_q(f) = \mathbf{A}[q]\mathbf{x}(f), \quad f \in \left[-\frac{f_p}{2}, \frac{f_p}{2}\right] \quad (2)$$

where $\mathbf{y}_q(f)$ and $\mathbf{x}(f)$ are continuous vectors of length I and L respectively and $\mathbf{A}[q] \in \mathbb{C}^{I \times L}$ gathers the Fourier coefficients such that $\mathbf{A}[q]_{il} = c_{i,(l+q)}$. If we go further, we can express the Fourier coefficients $c_{i,(l+q)}$ of a periodic sequence $\{s_{i0}, \dots, s_{i(M-1)}\}$ as:

$$c_{i,(l+q)} = d_{(l+q)} \sum_{k=0}^{M-1} s_{ik} e^{-j\frac{2\pi}{M}(l+q)k}. \quad (3)$$

The definition of $d_{(l+q)}$ is given in [5, equation (16)]:

$$d_{(l+q)} = \begin{cases} \frac{1}{M} & l + q = 0 \\ \frac{1 - e^{-j2\pi(l+q)/M}}{2j\pi(l+q)} & l + q \neq 0. \end{cases} \quad (4)$$

According to (3), $\mathbf{A}[q]$ can be written as:

$$\mathbf{A}[q] = \mathbf{S}\mathbf{F}[q], \quad -Q \leq q \leq Q, \quad (5)$$

where $\mathbf{S} \in \mathbb{C}^{I \times M}$ is the sequence matrix gathering one period of each sequence, i.e. $\mathbf{S}_{im} = s_{im}$. Moreover, $\mathbf{F}[q] \in \mathbb{C}^{M \times L}$ gathers the elements: $\mathbf{F}[q]_{ml} = d_{l+q} e^{-j\frac{2\pi}{M}(l+q)m}$. This matrix can be interpreted as the Discrete Fourier Matrix (DFT) which is frequency shifted by $e^{-j\frac{2\pi}{M}qk}$ and scaled by d_{l+q} .

Finally, by vertical concatenation of $\mathbf{y}_q(f)$ for the different values of q , we obtain the system $\mathbf{Y}(f) = \mathbf{A}(\mathbf{S})\mathbf{x}(f)$:

$$\underbrace{\begin{bmatrix} \mathbf{y}_{-Q}(f) \\ \vdots \\ \mathbf{y}_{+Q}(f) \end{bmatrix}}_{\mathbf{Y}(f)} = \underbrace{\begin{bmatrix} \mathbf{S}\mathbf{F}[-Q] \\ \vdots \\ \mathbf{S}\mathbf{F}[+Q] \end{bmatrix}}_{\mathbf{A}(\mathbf{S})} \mathbf{x}(f), \quad (6)$$

where $\mathbf{A}(\mathbf{S})$, denoted as the dictionary, is of size $(I \cdot Q') \times L$.

III. SEQUENCE OPTIMIZATION

The selection of the sequences in \mathbf{S} is crucial to the performance of the signal recovery stage. In this section, we focus on optimizing the selection of these sequences in order to obtain a dictionary $\mathbf{A}(\mathbf{S})$ which is as incoherent as possible.

The mutual coherence is a common measure of the correlation of a dictionary $\mathbf{A}(\mathbf{S})$, it is defined as the maximum absolute and normalized inner product between any two distinct columns of $\mathbf{A}(\mathbf{S})$:

$$\begin{aligned} \mu(\mathbf{A}(\mathbf{S})) &= \max_{l \neq m} \frac{|\mathbf{A}(\mathbf{S})_l^H \mathbf{A}(\mathbf{S})_m|}{\|\mathbf{A}(\mathbf{S})_l\|_2 \|\mathbf{A}(\mathbf{S})_m\|_2} \\ &= \left\| \tilde{\mathbf{A}}(\mathbf{S})^H \tilde{\mathbf{A}}(\mathbf{S}) - \mathbf{I}_L \right\|_{\infty}, \end{aligned} \quad (7)$$

where $\tilde{\mathbf{A}}(\mathbf{S})$ is the normalized version of $\mathbf{A}(\mathbf{S})$ (i.e. $\forall l, \tilde{\mathbf{A}}(\mathbf{S})_l = \mathbf{A}(\mathbf{S})_l \|\mathbf{A}(\mathbf{S})_l\|^{-1}$), \mathbf{I}_L is the identity matrix of dimension L and $\|\cdot\|_{\infty}$ is the infinity norm that returns the maximum value of its entry. The mutual coherence of a dictionary of size $(I \cdot Q') \times L$ is bounded by $\sqrt{\frac{L - (I + Q')}{(I + Q')(L - 1)}} \leq \mu(\mathbf{A}(\mathbf{S})) \leq 1$ where the lower bound is known as the Welch bound [11].

In the literature, many methods have been proposed to minimize the mutual coherence. However, these methods consider simpler models. A first family of methods considers the optimization of the dictionary $\mathbf{A}(\mathbf{S})$ regardless of the sequence matrix \mathbf{S} which is the only matrix that we can optimize for the MWC. A second family of methods consider the optimization of \mathbf{S} when the relationship between the dictionary and \mathbf{S} is linear, this is the special case when $Q' = 1$ ($\mathbf{A}(\mathbf{S}) = \mathbf{S}\mathbf{F}[0]$). In the sequel, we present methods that aim at minimizing the

mutual coherence by optimizing \mathbf{S} . Some of these methods are inspired by methods of both families cited above.

A. Quasi-Newton method

The first method that we propose is to tackle the minimization of the mutual coherence directly using a quasi-Newton method, more precisely the BFGS algorithm [12], to resolve the following non-linear optimization problem:

$$\min_{\mathbf{S}} \mu(\mathbf{A}(\mathbf{S})). \quad (8)$$

The quasi-Newton is an iterative descent method that starts with a given initialization and gradually minimize the cost function. Since the problem is non-convex, this method can only guarantee to find a local minimum solution that depends on the initialization, which we propose to be random in this paper.

B. Gradient method

The authors in [9] propose a gradient descent that minimizes the Frobenius norm instead of the infinity norm method. Thus, the optimization problem is rendered simpler at the price of not minimizing the mutual coherence directly but effectively.

This method can be directly applied in the special case when $Q' = 1$. Hereafter, we extend it to the case $Q' \neq 1$ which corresponds to solve the following optimization problem:

$$\min_{\mathbf{S}} \|\mathbf{A}(\mathbf{S})^H \mathbf{A}(\mathbf{S})\|_F^2 = \min_{\mathbf{S}} \left\| \sum_{q=-Q}^Q \mathbf{F}[q]^H \mathbf{S}^H \mathbf{S} \mathbf{F}[q] - \mathbf{I}_L \right\|_F^2. \quad (9)$$

The update of \mathbf{S} at each iteration (u) of the gradient descent algorithm is obtained by calculating the derivative of the cost function in (9) with respect to \mathbf{S} , yielding the following update step:

$$\mathbf{S}^{(u+1)} = \mathbf{S}^{(u)} - \eta \left(\sum_{q=-Q}^Q \mathbf{S}^{(u)} \mathbf{F}[q] \left(\mathbf{F}[q]^H (\mathbf{S}^{(u)})^H \mathbf{S}^{(u)} \mathbf{F}[q] - \mathbf{I}_L \right) \mathbf{F}[q]^H \right) \quad (10)$$

where η is the step-size and the algorithm is initialized by a random $\mathbf{S}^{(0)}$.

C. c-sidco method

The third method that we adapt is called c-sidco and is given in [10]. It can be applied to minimize the mutual coherence of the dictionary $\mathbf{A}(\mathbf{S})$ directly (without optimizing the sequence matrix \mathbf{S}). The results in [10] show that this method yields matrices with relatively small mutual coherence, close to the Welch bound. Therefore it appears to be beneficial to adapt this method in order to optimize \mathbf{S} . The main difficulty here is that the optimization problem should take into account the specific structure of $\mathbf{A}(\mathbf{S})$ where only \mathbf{S} can be optimized, which is not the case for the c-sidco method as proposed in [10]. Thus, we propose an iterative alternating scheme of two steps which is summarized in Algorithm 1. The algorithm alternates between the minimization of the mutual coherence

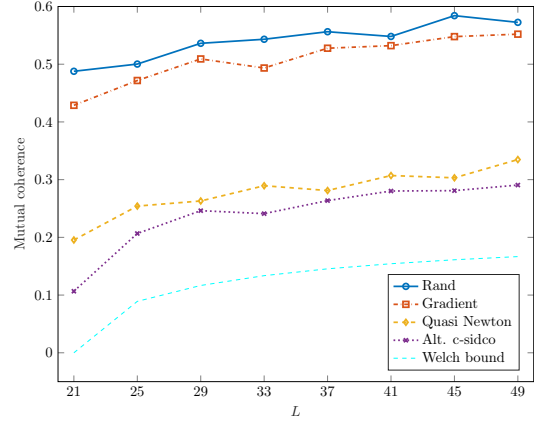


Fig. 2: The mutual coherence for different sequence optimization strategy with respect to the dictionary column number.

Algorithm 1: Alternating c-sidco

Output: $\hat{\mathbf{S}}$

Initialization: random $\hat{\mathbf{S}}$

```

1 for  $k = 1 \rightarrow N_{iter}$  do
2   for  $l \in \text{RANDOMIZED} \{1, \dots, L\}$  do
3      $T_l = 1 - \|\mathbf{A}(\hat{\mathbf{S}})_l^H \mathbf{A}(\hat{\mathbf{S}})\|_\infty$ 
4      $\hat{\mathbf{h}}_l = \underset{\mathbf{h}_l}{\operatorname{argmin}} \max_{m; m \neq l} |\mathbf{A}(\hat{\mathbf{S}})_m^H \mathbf{h}_l|$ 
5     s.t.  $\|\mathbf{h}_l - \mathbf{A}(\hat{\mathbf{S}})_l\|_2^2 \leq T_l$ 
6      $\hat{\mathbf{H}}_l = \hat{\mathbf{h}}_l \|\hat{\mathbf{h}}_l\|_2^{-1}$ 
7   end
8    $\hat{\mathbf{S}} = \underset{\mathbf{S}}{\operatorname{argmin}} \left\| \hat{\mathbf{H}}^T - \mathbf{S} \begin{bmatrix} \mathbf{F}[-Q] & \dots & \mathbf{F}[+Q] \end{bmatrix} \right\|_F^2$ 
9 end
```

of a matrix by applying one iteration of c-sidco (lines 2–6) and the sequence estimation in the least-squares sense (line 7). Let us now detail the proposed two-step algorithm.

The columns of a matrix $\hat{\mathbf{H}}$ are optimized one at a time and in a randomized order using c-sidco strategy. That is for a column l , a vector $\mathbf{h}_l \in \mathbb{C}^I$ is estimated such that the mutual coherence is minimized between \mathbf{h}_l and the columns of $\mathbf{A}(\hat{\mathbf{S}})$ in the positions different than l ($\{1, \dots, L\} \setminus \{l\}$). This leads to the following convex optimization problem appearing in line 4:

$$\min_{\mathbf{h}_l} \max_{m; m \neq l} |\mathbf{h}_m^H \mathbf{A}(\hat{\mathbf{S}})_l| \quad \text{s.t.} \quad \|\mathbf{h}_l - \mathbf{A}(\hat{\mathbf{S}})_l\|_2^2 \leq T_l. \quad (11)$$

This problem can be solved using an interior point method [12]. The constraint parameters T_l are chosen as $T_l = 1 - \max_{m, m \neq l} |\mathbf{A}(\hat{\mathbf{S}})_m^H \mathbf{A}(\hat{\mathbf{S}})_l|$ to guarantee a decreasing mutual coherence. Finally, the normalized version of $\hat{\mathbf{h}}_l$ is stored in the l th column of $\hat{\mathbf{H}}$.

In the second step, the sequence matrix \mathbf{S} is estimated in the least squares sense using the previously estimated $\hat{\mathbf{H}}$. However, this is not straightforward since the relationship between the dictionary and \mathbf{S} is not linear. Fortunately, this is possible by reshaping $\hat{\mathbf{H}}$ such that it is decomposed into Q' blocks of size $I \times L$ which are horizontally concatenated.

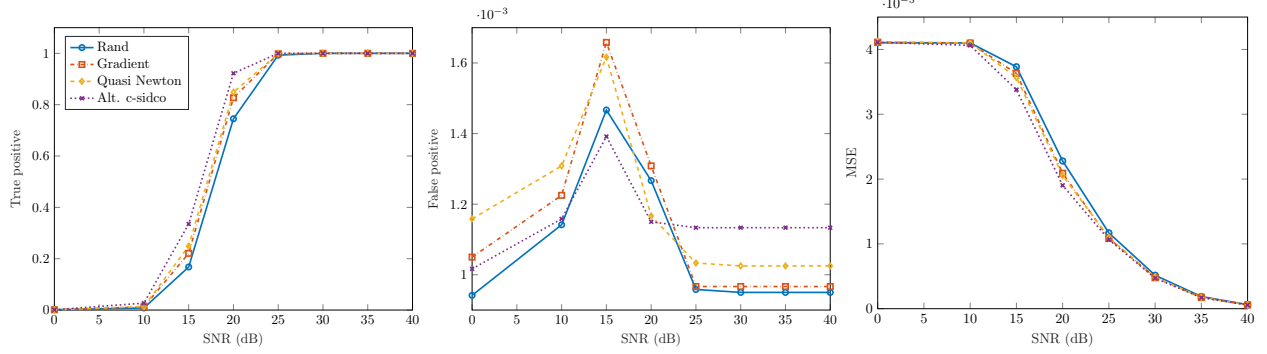


Fig. 3: Signal recovery performance.

Let us denote the reshaped dictionary of $\widehat{\mathbf{H}}$ as $\widehat{\mathbf{H}}^r$, it can be written as:

$$\widehat{\mathbf{H}}^r = [\widehat{\mathbf{H}}[-Q] \quad \dots \quad \widehat{\mathbf{H}}[+Q]] = \mathbf{S} [\mathbf{F}[-Q] \quad \dots \quad \mathbf{F}[+Q]]. \quad (12)$$

Hence, $\widehat{\mathbf{S}}$ can be found by resolving the above linear system in the least squares sense.

IV. RESULTS

We evaluate the proposed methods on synthetic wideband radar signals. The overall signal is composed of J components (e.g. chirp signals, phase codes) and can be expressed as $x(t) = \sum_{j=1}^J as(\tau_j, f_j) + n(t)$ where $as(\tau_j, f_j)$ is the j component with amplitude a , time delay τ_j and carrier frequency f_j . $n(t)$ is an additive Gaussian white noise with unit variance. In our experiments we set the Nyquist rate f_{NYQ} to $5GHz$, the length of a sequence period is M ($f_p = f_{NYQ}/M$). We set the physical number of channels to $I = 7$, while the sampling frequency f_s is set to $f_s = 3f_p$ ($Q' = 3$ and $Q = 1$), yielding 21 mixtures (3 mixtures per channel).

For the signal recovery stage, we use the Simultaneous OMP (S-OMP) algorithm [13] to approximate $x(f)$ in (6). This is a greedy algorithm that considers the Multiple Measurement Vector (MMV) system in which the sparse solutions share common support (the support contains the indices of the active bands l). However, the model given in (6) is an Infinite Measurement Vector (IMV) since f is continuous. Therefore, to transform the IMV system to an MMV we suppose that the frequencies are discrete and have the same dimension as the discrete temporal outputs $y[n]$ (a discrete Fourier transform is applied on the temporal outputs). Furthermore, we include a radar oriented stopping rule to S-OMP in order to estimate the number of active bands J , this stopping rule ensures a small fixed false alarm probability $P_{fa} = 10^{-3}$.

A. Mutual coherence minimization

To compare the performance of the methods in terms of minimizing the mutual coherence, we vary the periodic function frequency f_p such that the band number L corresponding to the dictionary column number varies between 21 and 45. For each value of L we generate 25 uniformly random dictionaries which are used to initialize the proposed methods. The results are presented in Figure 2 where the 25 values of

mutual coherence are averaged for each L . The results show that the gradient descent method enhances slightly the mutual coherence level. The quasi-Newton and c-sidco methods yield much better results, and the c-sidco method is the closest to the Welch bound.

B. Recovery performance

We study now the impact of sequence optimization on the performance of the recovery stage with respect to the SNR defined as $\text{SNR} = 10 \log |a|^2$ (supposing unit noise variance and unit energy normalized components $s(\tau_j, f_j)$). We consider four dictionaries respectively obtained with uniformly random generated sequences and gradient, quasi-Newton and c-sidco methods optimized sequences. The mutual coherence of these dictionaries are respectively 0.54, 0.47, 0.26 and 0.22. We fix the periodic function frequency $f_p = 151MHz$ yielding sequences of period length $M = 33$ and $L = 37$ bands and the SNR varies between 0 and 40 dB. For each SNR we generate 2000 signals $x(t)$ with $J = 3$ active bands. The methods are then compared using three criteria: the True Positive (TP), False Positive (FP) and the Mean Squared Error (MSE). The TP and FP are defined as:

$$TP = |\widehat{\mathcal{S}} \cap \mathcal{S}^*|/|\mathcal{S}^*|, \quad FP = |\widehat{\mathcal{S}} \cap \overline{\mathcal{S}^*}|/|\overline{\mathcal{S}^*}|, \quad (13)$$

where $\widehat{\mathcal{S}}$, \mathcal{S}^* and $\overline{\mathcal{S}^*}$ are respectively the estimated, ground-truth and the ground-truth complement supports. The MSE is calculated as the sum of the squared differences between the estimated bands and the ground truth bands normalized by the Frobenius norm of the ground truth bands.

The averaged plots for the 2000 signals are displayed in Figure 3. These results show that as the mutual coherence decreases the detection quality, indicated by the support cardinal and TP scores, increases. The estimation quality is thus enhanced for the methods with lower mutual coherence. Moreover, the FP of all the methods is close to the fixed P_{fa} .

V. CONCLUSION

In this paper, the problem of sequence optimization for the MWC is considered in the case where multiple mixtures are extracted in each channel. We present a novel problem formulation and propose three methods, inspired by the literature, which are extended to deal with this complex case. Results show that the methods were able to minimize the mutual coherence, allowing thus to enhance recovery performance.

REFERENCES

- [1] J. H. Ender, "On compressive sensing applied to radar," *Signal Processing*, vol. 90, no. 5, pp. 1402–1414, 2010.
- [2] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [3] J. L. Paredes, G. R. Arce, and Z. Wang, "Ultra-wideband compressed sensing: Channel estimation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 3, pp. 383–395, 2007.
- [4] Farokh Marvasti, *Nonuniform sampling: theory and practice*, Springer Science & Business Media, 2012.
- [5] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-nyquist sampling of sparse wideband analog signals," *IEEE Journal of selected topics in signal processing*, vol. 4, no. 2, pp. 375–391, 2010.
- [6] D. L. Donoho et al., "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [7] M. Elad, *Sparse and redundant representations: from theory to applications in signal and image processing*, Springer Science & Business Media, 2010.
- [8] M. Elad, "Optimized projections for compressed sensing," *IEEE Transactions on Signal Processing*, vol. 55, no. 12, pp. 5695–5702, 2007.
- [9] V. Abolghasemi, S. Ferdowsi, B. Makkiabadi, and S. Sanei, "On optimization of the measurement matrix for compressive sensing," in *European Signal Processing Conference, EUSIPCO*. IEEE, 2010, pp. 427–431.
- [10] C. Rusu, N. González-Prelcic, and R. W. Heath Jr, "Algorithms for the construction of incoherent frames under various design constraints," *Signal Processing*, vol. 152, pp. 363–372, 2018.
- [11] L. Welch, "Lower bounds on the maximum cross correlation of signals," *IEEE Transactions on Information theory*, vol. 20, no. 3, pp. 397–399, 1974.
- [12] J. Nocedal and S. Wright, *Numerical optimization*, Springer, 1999.
- [13] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part I: Greedy pursuit," *Signal processing*, vol. 86, no. 3, pp. 572–588, 2006.