Adaptive Identification and Recovery of Jointly Sparse Vectors

Roy Amel and Arie Feuer, *Life Fellow, IEEE*

Abstract—In this paper we present a novel approach to the solution of a sequence of SMV problems with a joint support. This type of problem arises in a number of applications such as multiband signal reconstruction and source localization. The approach we present is adaptive in that it solves it as a sequence of weighted SMV problems rather than collecting the measurement vectors and solving an MMV problem. The weights are adaptively updated from one instance to the next. This approach avoids delays and large memory requirements (at the cost of increased computational load) with the added capability of tracking changes in joint signal supports.

Index Terms—Sparse, multiple measurement vectors (MMV), adaptive, multiband, signal recovery.

I. INTRODUCTION

WE consider in this paper the following problem. A sequence of measurement vectors, $\mathbf{y}(t) \in \mathbb{R}^m$ is being generated and is known to satisfy the model

$$\mathbf{y}(t) = A\mathbf{x}_0(t) \qquad (1)$$

where $\mathbf{x}_0(t) \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$ and $n > m$. We wish to recover the vector sequence $\mathbf{x}_0(t)$ from the available measurements. Since this is obviously an under-determined problem, it does not have a unique solution in general. However, it becomes feasible when we add the prior of sparsity to the vectors $\mathbf{x}_0(t)$—only a small number of their entries are different from zero, *the same entries at each time* t. This type of problem arises in a number of applications, such as sub-Nyquist sampling of multiband signals (see e.g., [2] and [3]) and source localization (see e.g., [4]). To make this more formal, we introduce the notion of *support* of the vector $\mathbf{x}$, $S_{\mathbf{x}} = \{i \,|\, x_i \neq 0\}$ and assume that there exists a set $S \subset \{1, 2, \ldots, n\}$ such that $S_{\mathbf{x}_0(t)} \subseteq S$ where we refer to S as the joint support. We refer the reader to Table I where we summarize some of the notation we use in this paper.

One possible approach would be to solve at each time instance the following optimization problem:

$$\mathbf{x}_0(t) = \arg\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_0 \text{ such that } A\mathbf{x} = \mathbf{y}(t) \qquad (P_0)$$

TABLE I
NOTATION

$\|\mathbf{x}\|$	A vector of absolute values of the entries of $\mathbf{x}$
$\|\mathbf{x}\|_1$	l_1 norm.
$\|\mathbf{x}\|_2$	l_2 norm.
$\|\mathbf{x}\|_0$	l_0 pseudo-norm
$S_{\mathbf{x}}$	Support of $\mathbf{x}$
$|S|$	Cardinality of the finite set S.
x_i	i^{th} entry of $\mathbf{x}$.
$sign(x)$	1 if $x > 0$; -1 if $x < 0$; 0 elsewhere.
$Rank(Y)$	The rank of the matrix Y

where

$$\|\mathbf{x}\|_0 = \lim_{p \to 0} \sum_{i=1}^{n} |x_i|^p = |S_{\mathbf{x}}| \qquad (2)$$

is just the count of the nonzero entries of $\mathbf{x}$. The problem formulated in (P_0) is commonly referred to in the literature as the Sparse Representation (SR) problem for a Single Measurement Vector (SMV) and has received much attention over the past few years (see e.g., [6] or [7] and the many references there). The two main issues are clearly (i) under what conditions does (P_0) have a unique solution, and (ii) how can it be found.

We quote one result for (i) and for that we need the following:

Definition 1 ([7]): The *spark* of a given matrix A, denoted as $Spark(A)$, is the smallest number of columns of A which are linearly dependent.

Then:

Claim 2 ([7] and [8]): If a linear system $A\mathbf{x} = \mathbf{y}$ has a solution $\mathbf{x}_0$ satisfying $\|\mathbf{x}_0\|_0 < Spark(A)/2$, then this solution is necessarily unique for (P_0), namely, the sparsest possible.

Regarding (ii), a direct solution of (P_0) is known to be NP-Hard [7]. However, this problem was well studied, see e.g., [6], [9], [7]. Generally, there exist two main families of algorithms which solve (P_0) under additional assumptions. The first family consists of greedy algorithms (GA), [6] and [7], such as the Orthogonal Matching Pursuit (OMP). The main idea behind the GA is reducing complexity by finding a series of locally optimal single-term updates. The second family comprises the convex relaxation techniques, for example: FOcal Under-determined System Solver (FOCUSS) [10] and Basis Pursuit (BP) [6], [9], [7], [11]–[13]. In BP the objective function is relaxed to a convex form, the l_1 norm, which is known to be tractable and with polynomial-time complexity [14]. Stated formally:

$$\mathbf{x}_1 = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ such that } A\mathbf{x} = \mathbf{y} \qquad (P_1)$$

Going back to our problem and the sequential SMV solution we considered, we observe that we have not utilized the *common*

joint sparsity prior. In an attempt to utilize this prior we consider an alternative solution to our problem. Collect the data as long as it is being generated, say L measurement vectors and formulate the following optimization problem:

$$X = \arg \min_{X \in R^{n \times L}} \left\| \left(\sum_{t=1}^{L} |\mathbf{x}(t)| \right) \right\|_0 \text{ such that}$$
$$AX = Y \qquad \qquad (P_0^L)$$

where: $Y = [\mathbf{y}(1), \mathbf{y}(2), \ldots, \mathbf{y}(L)] \in \mathbb{R}^{m \times L}, X = [\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(L)] \in \mathbb{R}^{n \times L}, A \in \mathbb{R}^{m \times n}$. This problem has been addressed in the literature and is commonly referred to as the Multiple Measurement Vector (MMV) problem (see e.g., [15] or [16]). We quote a uniqueness result for this problem:

Claim 3 ([16]): If the set of equations $AX = Y$ has a solution X_0 satisfying $\|X_0\|_0 < (Spark(A) - 1 + Rank(Y))/2$, X_0 is the unique solution to the (P_0^L) problem.

Comparing the conditions in Claims 2 and 3 we readily observe the benefit of utilizing the joint sparsity prior in a considerably less restrictive condition for solution uniqueness in the MMV case. As (P_0^L) is also an NP-hard problem there are a number of alternative algorithms in the literature for solving it. They too can be grouped into greedy algorithms, such as OMP-MMV (see e.g., [16]) and relaxation algorithms such as the M-FOCUSS (see e.g., [15]).

However, the MMV approach to our problem suffers from a major drawback—it results in a (potentially large) time delay of L samples before the recovered signals are available. In addition, it would involve computations with large dimensional matrices and require a significant memory size to store the data. Using adaptive signal processing terminology, we refer to these solutions as *offline* solutions. Motivated by a similar dilemma in the adaptive signal processing literature (see e.g., [17]) we present here, what we believe to be a novel *adaptive* approach which is sequential in nature but utilizes the solutions up to time t when solving for time $t+1$. As such, there is no time delay, the data storage requirements are minimal and we have the added capability, demonstrated in our simulation, of tracking changes in signal support. We should also add that because of its sequential nature, L could be infinitely large in which case it is sometimes referred to as IMV (see e.g., [2]).

Another interesting consideration for the MMV problem is the question of how does one define the joint support. Typically, once a non-zero value appears in any entry at any time, the index of this entry enters the joint support. The question of frequency of occurrence is never raised. Intuition tells us though that, when this happens rarely, the right thing would be to ignore it. Our adaptive algorithms basically do this—only if a particular entry is consistently nonzero, it will affect the result.

The paper is structured as follows: Chapter 2 describes the setup of the problem. Chapter 3 details an algorithm based on BP and discusses some of its properties. Chapter 4 describes an algorithm based on OMP. In Chapter 5 we present some simulation results, where the presented algorithms are also applied to the multiband signal reconstruction problem and Chapter 6 concludes the paper.

II. GENERAL SETUP DESCRIPTION

Let $\mathbf{x}_0(t)$, $t = 1, \ldots, L$ be a stochastic vector process with the following properties:
1) There exists a set of indices $S \subset \{1, \ldots, n\}$ such that $|S| = k \leq m < n$ (possibly, $m \ll n$) and $x_{0,i}(t) = 0$ for all $i \notin S$ and all t.
2) The sequences $\{x_{0,i}(t)\}_{i \in S}$ are all stationary processes with pdfs $f_i(x)$.

Note that we do not assume the $f_i(x)$ to be continuous. Namely, we allow for the possibility that $\Pr(x_{0,i}(t) = 0) > 0$ for $i \in S$ at some t (namely, $|S(t)| < |S|$).

The measurement vector $\mathbf{y}(t)$ is generated via

$$\mathbf{y}(t) = A\mathbf{x}_0(t) \quad \text{for } t = 1, \ldots, L \qquad (3)$$

We wish to recover the vectors $\{\mathbf{x}_0(t)\}_{t=1}^{L}$ given the matrix (dictionary) A and the measurements $\{\mathbf{y}(t)\}_{t=1}^{L}$.

III. ADAPTIVE WEIGHTED BASIS PURSUIT (AW-BP) ALGORITHM

As discussed earlier, our approach is to process the data sequentially, to generate an estimate of $\mathbf{x}_0(t)$ at each time t. However, we wish to utilize the information acquired up to time t when we repeat the process at time $t + 1$. Motivated by [18], where the concept of 're-weighting' is introduced but in a totally different way, we prove the following simple result:

Claim 4: Let $\mathbf{x}_0$ be the unique solution to

$$\mathbf{x}_0 = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_0 \text{ such that } A\mathbf{x} = \mathbf{y} \qquad (4)$$

with the support set S, and let

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_1 \text{ such that } A\mathbf{x} = \mathbf{y} \qquad (5)$$
$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \|W\mathbf{x}\|_1 \text{ such that } A\mathbf{x} = \mathbf{y} \qquad (6)$$

where W is a diagonal matrix such that

$$\min_{i \notin S}\{W_{ii}\} > \max_{j \in S}\{W_{jj}\} \geq 0 \qquad (7)$$

Then, assuming $\Pr\{\hat{\mathbf{x}} = \mathbf{x}_0\} > 0$ and $\Pr\{\tilde{\mathbf{x}} = \mathbf{x}_0\} < 1$ we have

$$\Pr\{\hat{\mathbf{x}} = \mathbf{x}_0\} > \Pr\{\tilde{\mathbf{x}} = \mathbf{x}_0\} \qquad (8)$$

(For the proof see Appendix)

So, our idea is to forward, from time t to time $t + 1$, a weight matrix which carries with it a "soft" form of the support information. Specifically, we present the Adaptive Weighted Basis Pursuit (AW-BP) algorithm.

A. Algorithm Description and Performance

1) Initialize: Choose

$$\tilde{W}(0) = I^{n \times n}$$
$$1 < \rho < \infty$$
$$\ell \in \mathbb{Z}^+ \tag{9}$$

2) At $t > 0$:
   a) Find

$$\tilde{\mathbf{x}}(t) = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_1 \text{ s.t. } A\mathbf{x} = \mathbf{y}(t) \tag{10}$$

   b) Update $\tilde{W}(t)$ by

$$\tilde{W}_{ii}(t) = \begin{cases} \max\left\{\frac{1}{\rho}\tilde{W}_{ii}(t-1), \rho^{-\ell}\right\} & \text{if } \tilde{x}_i(t) \neq 0 \\ \min\{\rho\tilde{W}_{ii}(t-1), \rho^\ell\} & \text{if } \tilde{x}_i(t) = 0 \end{cases} \tag{11}$$

   c) Calculate $W_{ii}(t)$ by

$$W_{ii}(t) = \begin{cases} 0 & \text{if } \tilde{W}_{ii}(t) < 1 \\ 1 & \text{if } \tilde{W}_{ii}(t) \geq 1 \end{cases} \tag{12}$$

   d) Find

$$\hat{\mathbf{x}}(t) = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \|W(t)\mathbf{x}\|_1 \text{ s.t. } A\mathbf{x} = \mathbf{y}(t) \tag{13}$$

3) Use $\hat{\mathbf{x}}(t)$ as the estimate of $\mathbf{x}_0(t)$.

In hardware implementation, $\tilde{W}(t)$ in Step 2c can be replaced by $\tilde{W}(t-1)$ so that Steps 2a and 2d, which are the time consuming steps, can be carried out in parallel at the cost of one time interval delay in the adaptive learning curve. This is clearly observed in our experiment results in Fig. 2.

We wish to point out that the matrices $\tilde{W}(t)$ are the vehicles through which we utilize the support information we gained at time t to time $t+1$ in a non-greedy way. This is clearly done by the re-enforcement process described in (11). We have experimented extensively in replacing the hard limiter in (12) with other types of limiters, or even taking $W(t) = \tilde{W}(t)$, but found the above choice to work best.

Let us next present some performance results for the AW-BP algorithm. Let $q_i = \Pr\{\tilde{x}_i(t) \neq 0\}$, then we have from (11), for all $i \in \{1, \ldots, n\}$ [see (14) at the bottom of the page].This means that $\tilde{W}_{ii}(t)$ are stationary, finite state Markov chains with the same states, $\{\rho^{r-\ell-1}\}_{r=1}^{2\ell+1}$, but different transition probabilities, q_i. We prove the following properties for these chains:

Claim 5: Consider the Markov chains as defined in (11) with initial distribution vector $\mathbf{e}_{\ell+1}^T$ (the $(\ell+1)$th row of the $(2\ell+1)$ dimensional identity matrix). Then for each chain we have:

1) The transition matrix is given by

$$P_i = \begin{bmatrix} q_i & 1-q_i & 0 & \cdots & 0 & 0 & 0 \\ q_i & 0 & 1-q_i & \cdots & 0 & 0 & 0 \\ 0 & q_i & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1-q_i & 0 \\ 0 & 0 & 0 & \cdots & q_i & 0 & 1-q_i \\ 0 & 0 & 0 & \cdots & 0 & q_i & 1-q_i \end{bmatrix} \tag{15}$$

   with eigenvalues

$$\lambda_{i,1} = 1$$
$$\lambda_{i,r+1} = 2\cos\left(\frac{r\pi}{2(\ell+1)}\right)\sqrt{q_i(1-q_i)},$$
$$r = 1, 2, \ldots, 2\ell \tag{16}$$

2) The stationary distribution vector as defined by the equation, $\boldsymbol{\pi}_i^T = [\pi_{i,1}, \pi_{i,2}, \ldots, \pi_{i,2\ell+1}] = \boldsymbol{\pi}_i^T P_i$, is

$$\pi_{i,r} = \gamma_i^{r-1}\frac{1-\gamma_i}{1-\gamma_i^{2\ell+1}} \tag{17}$$

   where $\gamma_i = \frac{1-q_i}{q_i}$.

3) For $p_i(t)^T = [p_{i,1}(t), \ldots, p_{i,2\ell+1}(t)] = \mathbf{e}_{\ell+1}^T P_i^t$, the state probability distribution vector at time t, there exist constants $C_{i,r} > 0$ so that

$$|p_{i,r}(t) - \pi_{ir}|$$
$$\leq C_{i,r}\left(2\cos\left(\frac{\pi}{2(\ell+1)}\right)\sqrt{q_i(1-q_i)}\right)^t$$
$$\text{for } r = 1, 2, \ldots, 2\ell+1 \tag{18}$$

(For the proof see Appendix).

Claim 6: Let the process $\mathbf{y}(t)$ be generated as in (3) and $\tilde{\mathbf{x}}(t)$ be defined by (10),

$$q_i \equiv \Pr\{(\tilde{x}_i(t) \neq 0\} \tag{19}$$

and

$$r_{S,i}(\ell, t) = \Pr\{W_{ii}(t) = 0 \mid i \in S\}$$
$$r_{S^c,i}(\ell, t) = \Pr\{W_{ii}(t) = 1 \mid i \notin S\} \tag{20}$$

for the matrix sequence $W(t)$ generated by the AW-BP. Then, for all $i \in S$

$$r_{S,i}(\ell) = \lim_{t \to \infty} r_{S,i}(\ell, t)$$
$$= \frac{1-\gamma_i^\ell}{1-\gamma_i^{2\ell+1}} \tag{21}$$

and there exist constants $C_i > 0$ such that

$$|r_{S,i}(\ell, t) - r_{S,i}(\ell)|$$
$$\leq C_i\left(2\cos\left(\frac{\pi}{2(\ell+1)}\right)\sqrt{q_i(1-q_i)}\right)^t \tag{22}$$

$$\tilde{W}_{ii}(t) = \begin{cases} \max\left\{\frac{1}{\rho}\tilde{W}_{ii}(t-1), \rho^{-\ell}\right\} & \text{with probability } 0 < q_i < 1 \\ \min\{\rho\tilde{W}_{ii}(t-1), \rho^\ell\} & \text{with probability } 1 - q_i. \end{cases} \tag{14}$$

where $\gamma_i = \frac{1-q_i}{q_i}$, and for all $i \notin S$

$$r_{S^c,i}(\ell) = \lim_{t \to \infty} r_{S^c,i}(\ell,t)$$
$$= \gamma_i^\ell \frac{1 - \gamma_i^\ell + 1}{1 - \gamma_i^{2\ell+1}} \tag{23}$$

and there exist constants $D_i > 0$ such that

$$|r_{S^c,i}(\ell,t) - r_{S^c,i}(\ell)|$$
$$\leq D_i \left(2 \cos \left(\frac{\pi}{2(\ell+1)} \right) \sqrt{q_i(1-q_i)} \right)^t \tag{24}$$

where, again, $\gamma_i = \frac{1-q_i}{q_i}$.

Proof: From the definition of $\tilde{W}_{ii}(t)$ and (12) we have

$$\Pr\{W_{ii}(t) = 0\} = \Pr\{\tilde{W}_{ii}(t) < 1\} \tag{25}$$
$$= \sum_{r=1}^{\ell} p_{i,r}(t) \tag{26}$$

and by Claim 5(2)

$$r_{S,i}(\ell) = \sum_{r=1}^{\ell} \pi_{i,r} \tag{27}$$
$$= \frac{1 - \gamma_i^\ell}{1 - \gamma_i^{2\ell+1}} \tag{28}$$

Then, (22) follows directly from Claim 5(3) with $C_i = \sum_{r=1}^{\ell} C_{i,r}$.

Similarly, for $i \notin S$ we have

$$\Pr\{W_{ii}(t) = 1\} = \Pr\{\tilde{W}_{ii}(t) \geq 1\}$$
$$= \sum_{r=\ell+1}^{2\ell+1} p_r$$

and by Claim 5(2)

$$r_{S^c,i}(\ell) = \sum_{r=\ell+1}^{2\ell+1} \pi_{i,r} \tag{29}$$
$$= \gamma_i^\ell \frac{1 - \gamma_i^{\ell+1}}{1 - \gamma_i^{2\ell+1}} \tag{30}$$

and (24) follows directly from Claim 5(3) with $D_i = \sum_{r=\ell+1}^{2\ell+1} C_{i,r}$ which completes the proof. ∎

Claim 7: Let us assume that

$$q_i > \frac{1}{2} \text{ for all } i \in S \tag{31}$$

and

$$q_i < \frac{1}{2} \text{ for all } i \notin S \tag{32}$$

Then, for all $i \in S, 0 \leq \gamma_i < 1$, $r_{S,i}(\ell)$ as given in (21) is monotonically increasing function of $\ell \geq 0$ and $\lim_{\ell \to \infty} r_{S,i}(\ell) = 1$, and for all $i \notin S$, $\gamma_i > 1$, $r_{S^c,i}(\ell)$ as given in (23) is monotonically increasing function of $\ell \geq 0$ and $\lim_{\ell \to \infty} r_{S^c,i}(\ell) = 1$.

Proof: The proof follows directly from (31), (32), (21) and (23) applied to the derivatives of $r_{S,i}(y)$ and $r_{S^c,i}(y)$, $y \in \mathbb{R}$. ∎

An immediate conclusion from Claim 7 and (12) is that after some finite time, for sufficiently large ℓ, we get almost surely that $W_{ii}(t) = 0$ for all $i \in S$ and $W_{ii}(t) = 1$ for all $i \notin S$ which then by (13) leads to $\hat{x}(t) = \mathbf{x}_0(t)$ almost surely.

B. Discussion

We wish to briefly discuss here the assumptions made in Claim 7. First note that these assumptions do not imply that $\Pr\{\tilde{\mathbf{x}}(t) = \mathbf{x}_0(t)\} \geq \frac{1}{2}$. Next, in order to test these assumptions we have conducted extensive simulations the results of which we present here. Let $n = 200$ and $m = 100$, then we generated a dictionary $A \in \mathbb{R}^{m \times n}$ with entries $a_{u,j} \sim N(0,1)$ (i.e., Gaussian with zero mean and variance one) and normalized its columns (atoms) to have norm 1. Keeping A constant, for $k \in \{1, \ldots, n\}$ we chose randomly the set $S_\mathbf{x}$ so that $|S_\mathbf{x}| = k$ and generated $\mathbf{x}_0 \in \mathbb{R}^n$ so that $x_{i,0} = 0$ for $i \notin S_\mathbf{x}$ and $x_{i,0} \sim N(0,1)$ for $i \in S_\mathbf{x}$. Given A and $\mathbf{x}_0$ we calculated $\tilde{\mathbf{x}}$ according to

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_1 \text{ s.t. } A\mathbf{x} = A\mathbf{x}_0$$

For each cardinality k we repeated this 1000 times and counted the occurrences $\{\tilde{x}_i \neq 0 \text{ for } i \in S_\mathbf{x}\}$, denoted T_i. We used $0.001T_i$ as estimates of q_i. Fig. 1 shows how $\min_{i \in S} \{0.001T_i\}$ and $\max_{i \notin S} \{0.001T_i\}$ changed with the cardinality. We clearly observe that the assumptions we made in Claim 7 hold for all cardinalities up to $k = 90$.

Remark 8: The introduction and choice of the integer ℓ represents the trade off between convergence and tracking properties of our algorithms. The larger ℓ is the better the convergence but the tracking is slower. This is clearly demonstrated in Fig. 4. Our analysis indicates that the choice of ρ does not affect the algorithm performance as long as $\rho > 1$, at least theoretically.

IV. ADAPTIVE WEIGHTED-ORTHOGONAL MATCHING PURSUIT (AW-OMP) ALGORITHM

Another implementation of the adaptive concept we introduced is the Adaptive Weighted Orthogonal Matching Pursuit (AW-OMP).

Algorithm Description

1) Initialize: Choose

$$\widetilde{W}(0) = I^{n \times n}$$
$$1 < \rho < \infty$$
$$\ell \in \mathbb{Z}^+ \tag{33}$$

2) At $t > 0$:

a) Find, using OMP

$$\tilde{\mathbf{x}}(t) = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{y}(t) \tag{34}$$

b) Update $\widetilde{W}(t)$ by

$$\widetilde{W}_{ii}(t) = \widetilde{W}_{ii}(t) \tag{35}$$
$$= \begin{cases} \max \left\{ \frac{1}{\rho} \widetilde{W}_{ii}(t-1), \rho^{-\ell} \right\} & \text{if } \tilde{x}_i(t) \neq 0 \\ \min \left\{ \rho \widetilde{W}_{ii}(t-1), \rho^\ell \right\} & \text{if } \tilde{x}_i(t) = 0 \end{cases}$$

c) Calculate $W_{ii}(t)$ by

$$W_{ii}(t) = \begin{cases} 0 & \text{if } \widetilde{W}_{ii}(t) < 1 \\ 1 & \text{if } \widetilde{W}_{ii}(t) \geq 1 \end{cases} \tag{36}$$

d) Find, using OMP

$$\hat{\mathbf{x}}(t) = \arg \min_{\mathbf{x} \in \mathbb{R}^n} \|W(t)\mathbf{x}\|_0 \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{y}(t) \tag{37}$$

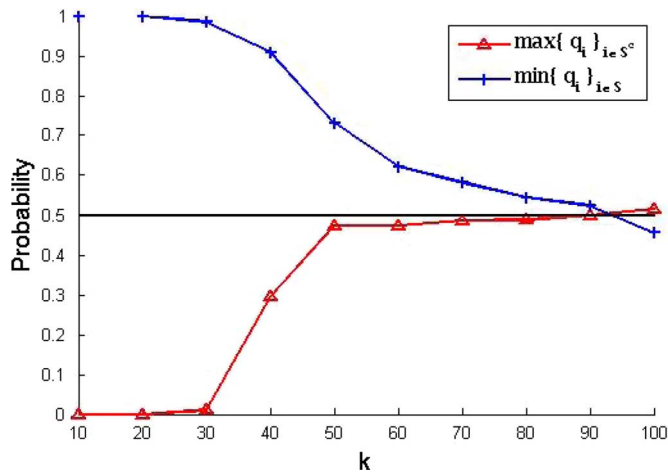3) Use $\hat{\mathbf{x}}(t)$ as the estimate of $\mathbf{x}_0(t)$.

Fig. 1. $\min\{q_i : i \in S\}$ and $\max\{q_i : i \notin S\}$ as functions of the support cardinality, k.

Fig. 2.   Success probability for true recovery as a function of time (also shown is the effect of using $W(t-1)$ instead of $W(t)$).

As the reader can readily note, this is quite similar to the AW-BP we introduced in the previous section. The difference is in the Steps (2b) and (2d) where BP is replaced by OMP. The OMP is described in many references, e.g., [6], [7]. In Step (2d) a straightforward modification of the standard OMP is required.

Remark 9: Under assumptions similar to those made in Claims 5, 6 and 7, similar properties can be proved for the AW-OMP.

Remark 10: Another modification we have experimented with is a compromise between the adaptive approach we have described so far and the offline approach most solutions to the MMV problem use. Instead of carrying out the iterations we described at every time step as the data is measured, one can accumulate a block of data vectors of small size $M < L$ and solve MMV versions of the BP and OMP at each iteration and carry the weights to the next block (so, one gets Block versions of the AW-BP and AW-OMP). The results of our experiments will be presented in the sequel.

V. SIMULATION RESULTS

To test the algorithms presented above we have conducted extensive simulation of two types of experiments. In the first we have used data generated according to the setup in Section 2 and tested our algorithms on this data. In the second, we applied our algorithms to data generated by sampling a multiband signal as described in [2].

A. Simulated Data

We start our experiments by generating data according to the setup in Section II where we chose $n = 200$, $m = 100$, pdfs $f_i \sim N(0,1)\ \forall\, i \in S$ and dictionary $A^{100 \times 200}$ with entries $N(0,1)$ and normalized columns. With each support we have generated 100 measurement vectors on which we applied our algorithm

1) AW-BP: We start by applying the AW-BP algorithm. To execute the steps of (10) and (13) we used CVX (see [26]). In Fig. 2 we show the convergence results for two support sizes, 60 and 80. Each experiment consisted of 100 runs and at each time we counted the relative number of perfect support estimations. We observe that for a support of cardinality 60, after 25
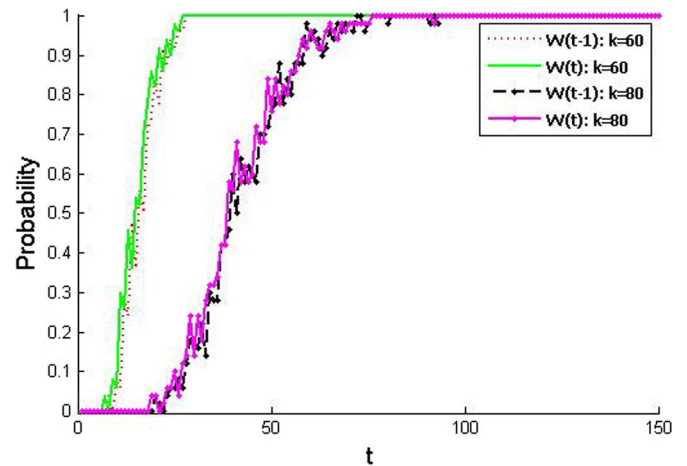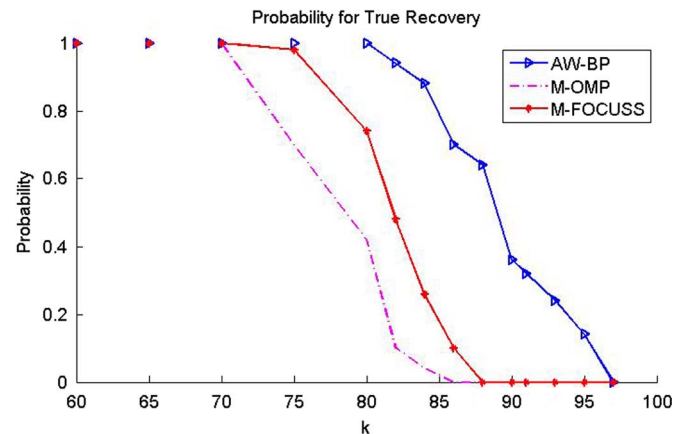

Fig. 3.   Success probability for true recovery as a function of the support cardinality.

time steps we get perfect estimation with probability 1 which we view as convergence. For support of cardinality 80 convergence occurs after 96 steps.

In Fig. 3 we present a comparison between the M-FOCUSS ([15]), M-OMP ([16]) and the AW-BP algorithms. We performed the simulation with 150 measurement vectors and the performance is measured in terms of probability for true recovery of $\mathbf{x}(150)$, as a function of the support cardinality. The M-OMP and M-FOCUSS algorithms use all vectors at once whereas the AW-BP processes one vector at a time. The M-FOCUSS was implemented using $p = 0.8$ as in [15], while for its termination we used $\delta = 10^{-6}$ as compared to $\delta = 0.01$ used in [15]. The parameters of the AW-BP are: $\rho = 1.05$, $l = \infty$. We have also experimented with the choice of ρ. Taking different values had very little effect on algorithm performance (see Remark 8). However, increasing it, at some point we started to encounter numerical instability. The AW-BP, after convergence, seems to outperform the other two algorithms (which work off-line). This may seem inconceivable, but one should keep in mind that these are three distinct algorithms neither of which solves directly (P_0^L).

In Fig. 4 we present the support change tracking capability of the AW-BP algorithm. For support cardinality $k = 60$, we have
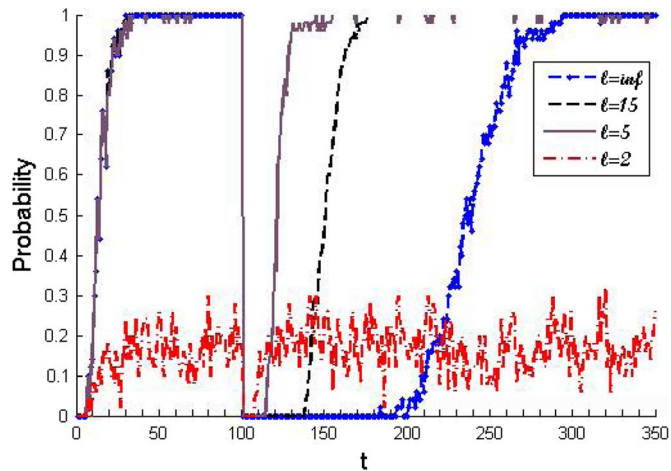
Fig. 4. Demonstration of tracking capability for different values of ℓ-success probability for true recovery as function of time for k = 60.
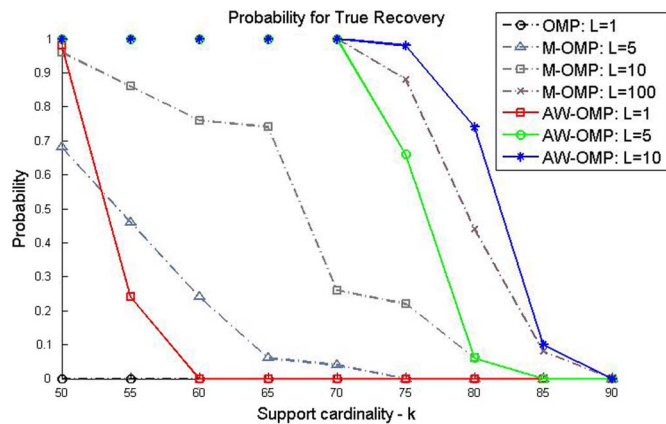

Fig. 5. Performance comparison between the M-OMP and the AW-OMP algorithms for sub-block sizes M = 1, 5, 10 and 100.

TABLE II
COMPUTATION COMPARISON

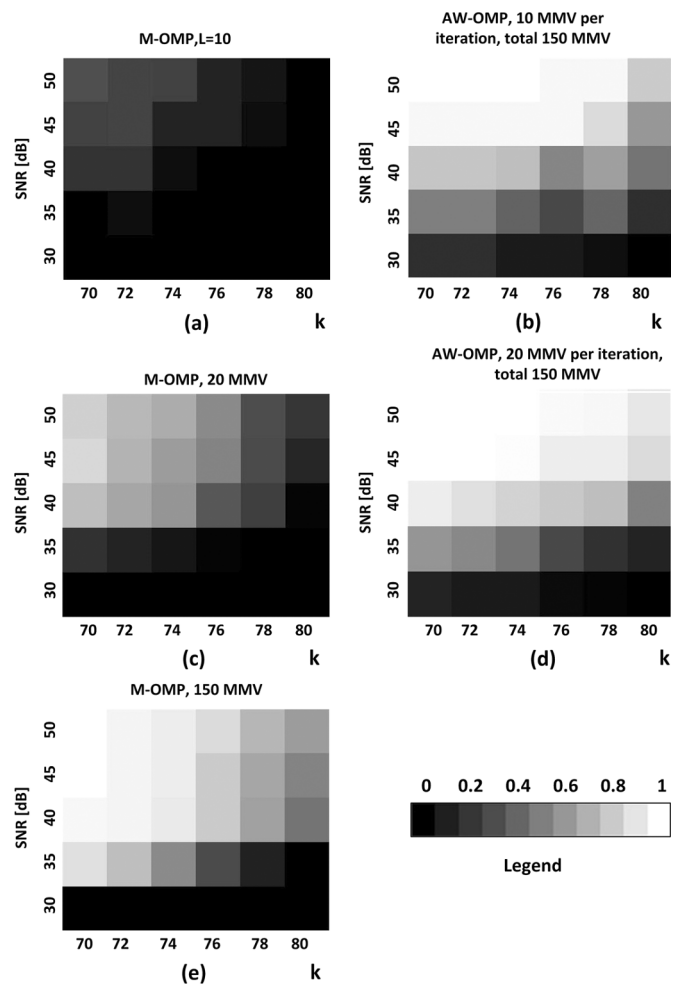|	AW-OMP	M-OMP
Multiplications	$4.35E + 08$	$1.53E + 08$
Summations	$4.23E + 08$	$1.49E + 08$
Divisions	$1.52E + 06$	$5.00E + 03$
Comparators	$3.83E + 06$	$1.14E + 04$
Total	$8.63E + 08$	$3.02E + 08$

Fig. 6. The figures represent probability for true identification as function of SNR and support cardinality, k. In (a) and (b) we compare the results for choosing in both M-OMP and AW-OMP $M = 10$. In (c) and (d) the comparison the comparison is for $M = 20$. In (e) we present the results for M-OMP for $L = 150$.

changed the support at time $t = 100$ and we show the algorithm behavior for different values of ℓ. We clearly observe that for larger ℓ the tracking is slower.

2) **AW-OMP***:* Next we applied the AW-OMP on the same data. Here we have experimented with the block version of the AW-OMP (see Remark 10) and show the results in Fig. 5 where we varied the block sizes as well. The M-OMP was applied on each block independently using different number of measurement vectors. It is noticeable that the AW-OMP achieves better results than the M-OMP for the same block size. Interestingly enough, the AW-OMP with block size 10, by $t = 100$, outperforms the M-OMP which is applied to the whole data set ($L = 100$).

To get an idea of the computational aspects of the approach we consider the set-up with $n = 200$, $m = 100$ and $k = 50$ and applied the AW-OMP up to $t = 100$. We counted the number operations carried out in the process. Then we applied the M-OMP on the collected data set. The comparison is presented in Table II. As anticipated, our approach is computationally more demanding. We should point out however, that this load is not uniform in time. As convergence occurs in our algorithms, the computational load goes down significantly as Step 2d is reduced to a solution of a set of linear equations.

We have repeated the above experiments with noisy measurements and show the results in Fig. 6. Each square in each figure corresponds to a certain SNR (in dB) and support cardinality and its grey level reflects the corresponding probability of success (perfect recovery of support), with white being probability 1 and black 0. Figs. 6(a) and (b) compare the performance of M-OMP and AW-OMP (after convergence) using measurement vectors sub-blocks of size $M = 10$. Figs. 6(c) and (d) compare the performance of M-OMP and AW-OMP (after convergence) using $M = 20$. In Fig. 6(e) we show the results of applying M-OMP on the whole data (150 measurement vectors). In all the results presented we observe the performance advantage of the adaptive algorithms even when the the whole data set is used (off line
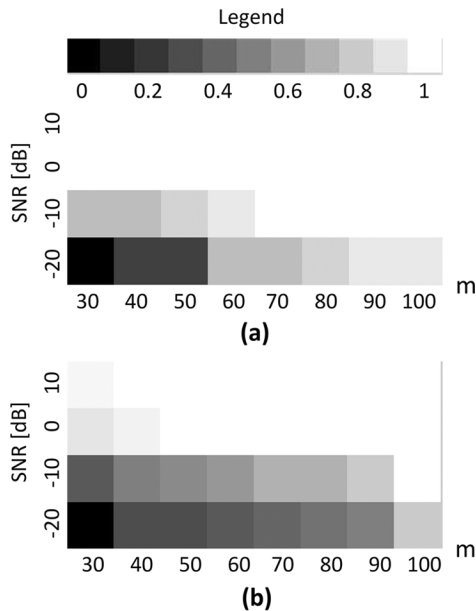
Fig. 7. Probability for true recovery of Sub-Nyquist system. (a) Using block AW-OMP, L = 10 and 100 measurement vectors total. **The results are after convergence**. (b) Using the reconstruction as described in [2].

processing) and this goes beyond the obvious advantage of the online processing (smaller or even no delays and less demand on memory).

B. Reconstruction of a Multiband Signal

There are at least two examples in the literature of applications which are cast in the form of the model in (1). One is the source localization problem in radars with sensor arrays (see e.g., [4]) and the second is the reconstruction of under-sampled multi-band signals (see e.g., [2]). We chose to apply our adaptive approach to the latter and used the example presented in the work of Mishali and Eldar [2]. They consider a signal which consists of N disjoint frequency bands at unknown locations, each of width bounded by B. The measurement vectors are generated by an ingenious sampling of that signal, considerably below Nyquist rate, as described in detail in [2]. They show that the measured data vectors satisfy the frequency domain equation $\hat{\mathbf{y}}(f) = A\hat{\mathbf{x}}(f)$ where $\hat{\mathbf{y}}(f)$ is the DFT of the measurements vectors $\mathbf{y}[t]$, A is a known matrix and the vector $\hat{\mathbf{x}}(f)$ is $2N$ sparse for every f. To reconstruct the multiband signal, the support of $\hat{\mathbf{x}}(f)$ is found first, which then leads to calculating $\hat{\mathbf{x}}(f)$ and the signal reconstruction. Both the sampling and reconstruction described in [2] were thoroughly investigated by Mishali and Eldar who also proceeded to successfully implement it in hardware.

We propose here an alternative approach to the reconstruction by looking at the time domain version of the above equation, $\mathbf{y}[t] = A\mathbf{x}[t]$, where $\mathbf{y}[t]$ are the measurement vectors, A as above and $\mathbf{x}[t]$ is $2N$ sparse (same support as $\hat{\mathbf{x}}(f)$) for all t. With this equation we apply our AW-OMP to the same signal as in [2] ((37) and Option A in their Table I).

In Fig. 7(a) we show the performance using AW-OMP (*after convergence*) and, in Fig. 7(b) the results using the algorithm described in [2]. The rsults seem quite comparable.

VI. CONCLUSION

The set up considered in this paper is of, possibly a large number, of measurement vectors being generated sequentially from a set of jointly sparse signals with a common dictionary. One approach would be to collect all the measurement vectors and solve an MMV problem with one of the available algorithms. This approach has the advantage of low computational requirements and takes advantage of all the joint sparsity benefits (see Claim 3). However, it requires a large memory, introduces a (possibly large) delay in the signal reconstruction and it does not have the capability of tracking a changing joint sparsity. The other possibility is solving independently an SMV problem each time a new measurement vector becomes available. This approach, while having minimal memory requirements and no delay, does not make use of the joint sparsity prior. In this paper we introduced a novel approach which has the benefits of the two extremes above. It can be viewed as an adaptive method of solution. It is based on a sequential solution of (weighted) SMV (or weighted blocked MMV) problems. The resulting signal supports are then carried over, in a *non-greedy way* via a weighting matrix, from one instance to the next.

Different applications of this concept can be realized by using different methods of solving the sequence of SMV problems. In our paper we have chosen two representative examples to demonstrate our approach. One is using the Basis Pursuit (BP), resulting in what we termed AW-BP and the other, the Orthogonal Matching Pursuit (OMP) resulting in AW-OMP. However the method can be applied through many other existing algorithms such as: IRLS ([7]), LARS ([19]), FOCUSS ([7], [10]), M-FOCUSS ([15]) etc. Clearly, the resulting properties will depend on the actual underlying algorithm chosen.

All our experiments with the AW-BP and AW-OMP demonstrated a clear advantage over their sequential independent counterparts (both SMV and blocked MMV). While this is not surprising, we also observed a performance similar (or even better) to MMV *when the whole data set is considered*. Comparing the performance of the AW-BP to the AW-OMP we observed that the first has a significant advantage in its recovery rate for growing support cardinality, while the latter is a faster algorithm with significantly lower computational load.

APPENDIX
PROOF OF CLAIM 4

Proof: Let

$$\tilde{\mathcal{L}} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_1 \leq \|\mathbf{x}_0\|_1\}$$
$$\hat{\mathcal{L}} = \{\mathbf{x} \in \mathbb{R}^n : \|W\mathbf{x}\|_1 \leq \|W\mathbf{x}_0\|_1\} \qquad (38)$$

Clearly $\mathbf{x}_0 \in \tilde{\mathcal{L}} \cap \hat{\mathcal{L}}$ and for any $\mathbf{x} \in \hat{\mathcal{L}}$ we have

$$\sum_{i \notin S} W_{ii}|x_i| \leq \sum_{j \in S} W_{jj}(|x_{0,j}| - |x_j|) \qquad (39)$$

Then, by (7) and (39) we get

$$\min_{i \notin S} \{W_{ii}\} \sum_{i \notin S} |x_i| \leq \sum_{i \notin S} W_{ii} |x_i|$$

$$\leq \sum_{j \in S} W_{jj}(|x_{0,j}| - |x_j|)$$

$$\leq \max_{j \in S} \{W_{jj}\} \sum_{j \in S}(|x_{0,j}| - |x_j|)$$

$$\leq \min_{i \notin S} \{W_{ii}\} \sum_{j \in S}(|x_{0,j}| - |x_j|)$$

So $\|\mathbf{x}\|_1 \leq \|\mathbf{x}_0\|_1 \Rightarrow \mathbf{x} \in \hat{\mathcal{L}} \Rightarrow \hat{\mathcal{L}} \subseteq \tilde{\mathcal{L}}$.

Let $\mathbf{x}_1 \in \mathbb{R}^n$ be such that $\sum_{j \in S} |x_{1,j}| = 0$ and $\|\mathbf{x}_0\|_1 = \|\mathbf{x}_1\|_1$, then $\mathbf{x}_1 \in \tilde{\mathcal{L}}$ and $\sum_{j \in S} |x_{0,j}| = \sum_{i \notin S} |x_{1,i}|$. On the other hand

$$\sum_{i \notin S} W_{ii} |x_{1,i}| \geq \min_{i \notin S} \{W_{ii}\} \sum_{i \notin S} |x_{1,i}|$$

$$> \max_{j \in S} \{W_{jj}\} \sum_{j \in S} |x_{0,j}|$$

$$\geq \sum_{j \in S} W_{jj} |x_{0,j}|$$

which means that $\mathbf{x}_1 \notin \hat{\mathcal{L}}$ and we can conclude that $\hat{\mathcal{L}} \subset \tilde{\mathcal{L}}$. This means that there are only three possibilities

$$\mathcal{P}_1 \equiv \{\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{y}\} \cap \hat{\mathcal{L}} \setminus \{\mathbf{x}_0\} \neq \varnothing\}$$

$$\mathcal{P}_2 \equiv \{\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{y}\} \cap \hat{\mathcal{L}} \setminus \{\mathbf{x}_0\} = \varnothing \text{ and}$$
$$\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{y}\} \cap \tilde{\mathcal{L}} \setminus \{\mathbf{x}_0\} \neq \varnothing\}$$

and

$$\mathcal{P}_3 \equiv \{\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{y}\} \cap \hat{\mathcal{L}} \setminus \{\mathbf{x}_0\} = \varnothing \text{ and}$$
$$\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{y}\} \cap \tilde{\mathcal{L}} \setminus \{\mathbf{x}_0\} = \varnothing\}$$

We observe that $\mathcal{P}_1 \Rightarrow \{\tilde{\mathbf{x}} \neq \mathbf{x}_0 \text{ and } \hat{\mathbf{x}} \neq \mathbf{x}_0\}$, $\mathcal{P}_2 \Rightarrow \{\tilde{\mathbf{x}} \neq \mathbf{x}_0 \text{ and } \hat{\mathbf{x}} = \mathbf{x}_0\}$ and $\mathcal{P}_3 \Rightarrow \{\tilde{\mathbf{x}} = \mathbf{x}_0 \text{ and } \hat{\mathbf{x}} = \mathbf{x}_0\}$. By the claim assumptions and since $\hat{\mathcal{L}} \subset \tilde{\mathcal{L}}$, we note that $\Pr\{\mathcal{P}_i\} > 0$ so we readily conclude that

$$\Pr\{\hat{\mathbf{x}} = \mathbf{x}_0\} = \Pr\{\mathcal{P}_2\} + \Pr\{\mathcal{P}_3\}$$
$$> \Pr\{\mathcal{P}_3\}$$
$$= \Pr\{\tilde{\mathbf{x}} = \mathbf{x}_0\}$$

which completes the proof. ∎

PROOF OF CLAIM 5

Proof:
1) We readily observe from (14) that the Markov chain has $2\ell + 1$ states, $x_r = \rho^{r-\ell-1}$, and the matrix P_i in (15) is a direct consequence of (14). To find the eigenvalues of P_i we are interested in the roots of

$$F_i(\lambda) = \det(\lambda I_{2\ell+1} - P_i) \tag{40}$$

Let us denote

$$E_{i,R}(\lambda)$$

$$= \det \left(\lambda I_R - \begin{bmatrix} 0 & (1-q_i) & \cdots & 0 & 0 \\ q_i & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & (1-q_i) \\ 0 & 0 & \cdots & q_i & 0 \end{bmatrix} \right) \tag{41}$$

and

$$G_{i,R}(\lambda)$$

$$= \det \left(\lambda I_R - \begin{bmatrix} q_i & (1-q_i) & \cdots & 0 & 0 \\ q_i & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & (1-q_i) \\ 0 & 0 & \cdots & q_i & 0 \end{bmatrix} \right) \tag{42}$$

Then, using determinant properties and expanding $F_i(\lambda)$ by the last row and $E_{i,R}(\lambda)$ and $G_{i,R}(\lambda)$ by the first rows of the respective matrices, we get the following relationships:

$$E_{i,R}(\lambda)$$
$$= \lambda E_{i,R-1}(\lambda) - q_i(1-q_i)E_{i,R-2}(\lambda) \tag{43}$$
$$G_{i,r}(\lambda)$$
$$= (\lambda - q_i)E_{i,r-1}(\lambda) - q_i(1-q_i)E_{i,r-2}(\lambda) \tag{44}$$
$$F_i(\lambda)$$
$$= (\lambda - (1-q_i))G_{u,2\ell}(\lambda)$$
$$- q_i(1-q_i)G_{u,2\ell-1}(\lambda) \tag{45}$$

By straight forward substitutions we get from (43)–(45)

$$F_i(\lambda) = (\lambda - 1)E_{i,2\ell}(\lambda) \tag{46}$$

As $E_{i,2\ell}(\lambda)$ is the characteristic polynomial of a tridiagonal Toeplitz matrix it is known (see e.g., [20]) to have the roots:

$$2\sqrt{q_i(1-q_i)} \cos\left(\frac{r\pi}{2\ell+1}\right), \quad r = 1, 2, \ldots, 2\ell \tag{47}$$

and combined with (46), (16) follows.
2) It can be shown (see e.g., [21]) that the Markov chain with the given P results in a stable Markov chain and will converge to a steady state distribution. The proof of (17) is by straight forward substitution.
3) Let $1 = \lambda_{i,1} > |\lambda_{i,2}| \geq \ldots \geq |\lambda_{i,2\ell+1}|$ be the eigenvalues of P_i and $\{\mathbf{u}_{i,r}^T\}_{r=1}^{2\ell+1}$ the corresponding left eigenvectors which are linearly independent (as the eigenvalues are distinct). Note that $\mathbf{u}_{i,1}^T = \boldsymbol{\pi}_i^T$, the stationary distribution vector. Then we can write

$$P_i = V_i \Lambda_i U_i \tag{48}$$

where

$$U_i = \begin{bmatrix} \mathbf{u}_{i,1}^T \\ \mathbf{u}_{i,2}^T \\ \vdots \\ \mathbf{u}_{i,2\ell+1}^T \end{bmatrix}$$

$$V_i = U_i^{-1} = \begin{bmatrix} \mathbf{v}_{i,1} & \mathbf{v}_{i,2} & \cdots & \mathbf{v}_{i,2\ell+1} \end{bmatrix}$$

$$\Lambda_i = \text{diag}\{\lambda_{i,r}\} \tag{49}$$

where we note that $\mathbf{v}_{i,1} = \mathbf{1}$ (a vector of ones). Namely, we can also rewrite

$$P_i = \sum_{j=1}^{2\ell+1} \lambda_{i,j} \mathbf{v}_{i,j} \mathbf{u}_{i,j}^T$$

$$= \mathbf{1}\boldsymbol{\pi}_i^T + \sum_{j=2}^{2\ell+1} \lambda_{i,j} \mathbf{v}_{i,j} \mathbf{u}_{i,j}^T$$

and

$$P_i^t = \mathbf{1}\boldsymbol{\pi}_i^T + \sum_{j=2}^{2\ell+1} \lambda_{i,j}^t \mathbf{v}_{i,j} \mathbf{u}_{i,j}^T \tag{50}$$

Then we get

$$\mathbf{p}_i(t)^T = \mathbf{e}_\ell^T P_i^t$$

$$= \boldsymbol{\pi}_i^T + \sum_{j=2}^{2\ell+1} \lambda_{i,j}^t \mathbf{e}_{\ell+1}^T \mathbf{v}_{i,j} \mathbf{u}_{i,j}^T$$

or

$$|p_{i,r}(t) - \pi_{i,r}| \le C_{i,r} \lambda_{i,2}^t \text{ for } r = 1, 2, \ldots, 2\ell+1 \tag{51}$$

where $C_{i,r} = |\sum_{j=2}^{2\ell+1} \mathbf{e}_{\ell+1}^T \mathbf{v}_{i,j} (\mathbf{u}_{i,j}^T)_r|$ which completes the proof of the claim. ∎

REFERENCES

[1] M. Mishali and Y. C. Eldar, "Blind multiband signal reconstruction: Compressed Sensing for analog signals," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.

[2] M. Mishali and Y. C. Eldar, "From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals," *IEEE J. Sel. Topics in Signal Process., Special Issue on Compressed Sens.*, vol. 4, no. 2, pp. 375–391, Apr. 2010.

[3] R. Venkataramani and Y. Bresler, "Perfect reconstruction formulas and bounds on aliasing error in sub-Nyquist nonuniform sampling of multiband signals," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2173–2183, Sep. 2000.

[4] D. Malioutov, M. Cetin, and A. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[5] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 1289–1306, 2006.

[6] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, p. 34, 2009.

[7] M. Elad, *Sparse and Redundant Representations. From Theory to Applications in Signal and Image Processing*, 1st ed. New York, NY, USA: Springer, 2010.

[8] M. Elad, "Sparse representations are most likely to be the sparsest possible," *EURASIP J. Appl. Signal Process.*, pt. Paper No. 96247, 2006.

[9] E. J. Candes and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[10] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, Mar. 1997.

[11] Y. Kopsinis, K. Slavakins, and S. Theodoridis, "Online sparse system identification and signal reconstruction using projection onto weighted 1_l balls," *IEEE Trans. Signal Process.*, vol. 59, pp. 936–952, Mar. 2011.

[12] D. L. Donoho, "For most large underdetermined systems of linear equations, the minimal '1-norm near-solution approximates the sparsest near-solution," *Commun. Pure Appl. Math.*, vol. 59, no. 7, pp. 907–934, Jul. 2006.

[13] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159, 2001.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

[15] S. F. Cotter, B. D. Rao, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, pp. 2477–2488, Jul. 2005.

[16] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple-measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, pp. 4634–4643, Dec. 2006.

[17] A. H. Sayed, *Fundamentals of Adaptive Filtering*. Hoboken, NJ, USA: Wiley, 2003.

[18] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

[19] B. Efron, T. Hastie, I. M. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–499, 2004.

[20] G. D. Smith, *Numerical Solutions of Partial Differential Equations*. New York, NY, USA: Clarendon, 1978.

[21] C. M. Grinstead and J. L. Snell, *Introduction to Probability*, 2nd ed. New York, NY, USA: Amer. Math. Soc., 2003.

[22] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4692–4702, Oct. 2008.

[23] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, pp. 489–509, Feb. 2006.

[24] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[25] R. Amel and A. Feuer, "Adaptive algorithm for online identification and recovering of jointly sparse signals," in *Proc. Sparse 11, 4th Workshop on Signal Process. Adapt. Sparse Structure Represent.*, Jun. 2011, p. 113.

[26] M. C. Grant, CVX [Online]. Available: http://cvxr.com/

[27] J. A. Tropp, "Algorithms for simultaneous approximation. Part II: Convex relaxation," *Signal Process. Special Issue*, vol. 86, pp. 598–602, Apr. 2006.

Roy Amel received the B.Sc. degree in electrical engineering in 2009 and the M.Sc. degree in electrical engineering in 2013 both from the Technion—Israel Institute of Technology, Haifa.

Since 2012, he has been a System Engineer of the physical layer (include all the blocks from the Antenna to the bit) of Intel WIFI products.

Arie Feuer (M'76–SM'93–F'04–LF'14) received the B.Sc. and M.Sc. degrees from the Technion—Israel Institute of Technology in 1967 and 1973, respectively, and the Ph.D. degree from Yale University in 1978. In 2009, he received an honorary doctorate from the University of Newcastle.

He has been with the Department of Electrical Engineering, Technion—Israel Institute of Technology, since 1983 were he is currently a Professor Emeritus. From 1967 to 1970, he was in industry working on automation design and between 1978 and 1983 with Bell Labs in Holmdel, NJ. In the last 22 years, he has been regularly visiting the Electrical Engineering and Computer Science Department at the University of Newcastle. His current research interests include: Medical imaging—in particular, ultrasound and CT; resolution enhancement of digital images and videos; 3D video and multi-view data; Sampling and combined representations of signals and images.

Dr. Feuer served as the president of the Israel Association of Automatic Control between 1994 and 2002, and was a member of the IFAC Council during 2002–2005.