OPTIMAL SAMPLING STRUCTURE FOR ASYNCHRONOUS MULTI-ACCESS CHANNELS

Xiao Li[†], *Andrea Rueetschi*[†], *Anna Scaglione*[†], *Yonina C. Eldar*[*]

[†] Dept. of Electrical and Computer Engineering, University of California, Davis, CA, USA
[*] Dept. of Electrical Engineering, Technion - Israel Institute of Technology, Haifa, Israel

ABSTRACT

A sensitive receiver operation in multi-access reception is to detect the presence of training signals and identify sources from which the signals are sent, with a certain physical delay and center frequency. This is typically a sequential search, where the receiver tests the presence of specific signals and then acquires synchronization parameters (delays, Dopplers), for each component. In this paper, we develop an optimal compressive multichannel sampling (CMS) architecture, the output samples of which are fed to the proposed Sparsity Regularized (SR) Generalized Likelihood Ratio Test (GLRT). It is shown that SR-GLRT using the optimal sampling scheme exhibits better performance than conventional compressed sensing structure, and furthermore effectively scales down the storage requirement and complexity with greater flexibility than conventional architectures.

***Index Terms*—** Multichannel sampling, compressed sensing.

1. INTRODUCTION

One of the most critical tasks of many receivers in a multi-access channel, referred to as *link acquisition*, is that of detecting the presence of any signals, and identifying the *link parameters* (e.g., delays, Doppler) of an *unknown subset* $\mathcal{I}$ out of I possible sources. Similar to [2, 3], the active sources transmit distinct preambles $\phi_i(t)$, $i \in \mathcal{I}$ to a receiver, whose task is to process the received signal $x(t)$ and test for the presence of any specific component active in the link. This is in general a complicated hypothesis testing problem, where existing acquisition algorithms usually acquire a sufficient statistic by directly sampling (DS) $x(t)$ at (or above) the Nyquist rate for estimation and detection. Another prevalent choice is to facilitate the search of both the *active set* $\mathcal{I}$ and the *link parameters* by comparing the magnitudes of the match filtered (MF) outputs of the signal $x(t)$.

Recently, there have been advances in exploiting *sparsity* to solve signal detection, active user identification and parameter estimation, to replace the DS or MF approaches. For instance, [2–4] assume the presence of signals and deal with identification of the users and/or estimation of signal parameters by creating a dictionary from the known templates $\phi_i(t)$ of the signal and viewing the signal $x(t)$ as a sparse linear combination of the elements inside the dictionary. On the other hand, [5–8] propose detection schemes using generic compressed measurements from a discrete model without resolving the system parameters. In these papers, Doppler and delays are not explicitly considered, and also the acquisition structure, from the analog signal to the digital measurements, is not explicitly optimized.

In this paper, we propose an acquisition algorithm using a Sparsity Regularized (SR) GLRT with an optimal sampling structure, and further assess its performance against traditional designs in compressed sensing as well as MF. Our contribution compared to [2, 3, 5–8] is in the design of the sampling structure that maximizes the weighted average Kullback-Leibler (KL) distance of the hypotheses in the test.

2. SPARSE MODEL-BASED LINK ACQUISITION

During the training phase a certain user from the active set $i \in \mathcal{I}$ transmits a specific preamble $\phi_i(t)$ to the destination, and the observation at the receiver can be written as

$$x(t) = \sum_{i\in\mathcal{I}}\sum_{r=1}^{R} h_{i,r}\phi_i(t-\tau_{i,r})e^{\mathrm{i}\omega_{i,r}t} + v(t), \tag{1}$$

where $0 \le \tau_{i,r} \le \tau_{\max}$ is an unknown delay of the ith user in the rth multipath bounded by the delay spread $\tau_{\max}$, $|\omega_{i,r}| \le \omega_{\max}$ is the Doppler frequency bounded by the Doppler spread $\omega_{\max}$, and $h_{i,r}$ is the unknown channel fade. We further assume that the maximum multipath order R is known and the noise component $v(t)$ is Gaussian with autocorrelation $\mathbb{E}\{v(t)v^*(s)\} = \sigma^2\delta(t-s)$.

Here, we approximate $\tau_{i,r} \approx q_{i,r}\Delta\tau$ and $\omega_{i,r} \approx k_{i,r}\Delta\omega$ for some integers $q_{i,r}$ and $k_{i,r}$ with a certain resolution $\Delta\tau = \tau_{\max}/Q$, $\Delta\omega = \omega_{\max}/K$. We further introduce the triple-indexed coefficient

$$\alpha_{i,k,q} = \sum_{j\in\mathcal{I}} h_{j,r}\delta[i-j]\delta[k-k_{j,r}]\delta[q-q_{j,r}] \tag{2}$$

to indicate whether the ith user is transmitting at a certain delay $\tau = q\Delta\tau$ with a certain carrier offset $\omega = k\Delta\omega$. Denoting $\phi_{i,k,q}(t) \triangleq \phi_i(t-q\Delta\tau)e^{\mathrm{i}k\Delta\omega t}$ as the MF template with $\mathcal{Q} = \{0, 1, \cdots, Q-1\}$ and $\mathcal{K} = \{-K, \cdots, K\}$, the signal is approximately expressed as

$$x(t) = \sum_{i=1}^{I}\sum_{k\in\mathcal{K}}\sum_{q\in\mathcal{Q}} \alpha_{i,k,q}\phi_{i,k,q}(t) + v(t). \tag{3}$$

From (2), we can see that the link parameters are embedded in the sparse coefficients $\alpha_{i,k,q}$'s, including delays $\tau_{i,r} = q_{i,r}\Delta\tau$, Doppler frequencies $\omega_{i,r} = k_{i,r}\Delta\omega$ and the active users $i \in \mathcal{I}$. We define the active links, or the delay-Doppler pairs, for the ith user as

$$\mathcal{A}_i \triangleq \left\{(k,q) : |\alpha_{i,k,q}|^2 > 0, k \in \mathcal{K}, q \in \mathcal{Q}\right\}. \tag{4}$$

If a user is not transmitting $i \notin \mathcal{I}$, then the set is empty $\mathcal{A}_i = \varnothing$ and no components are present $|\mathcal{A}_i| = 0$. With $\mathcal{A}_i$ defined above, the goal of the *link acquisition* is to identify i) the existence of any signal $\alpha_{i,k,q} \neq 0$ in the observation; ii) the set of active users indicated by the indices i with a non-trivial link parameter set $\mathcal{A}_i \neq \varnothing$; iii) the delay-Doppler pairs for active users $\mathcal{A}_i \subseteq \mathcal{K} \times \mathcal{Q}$, $i \in \mathcal{I}$.

3. COMPRESSIVE ACQUISITION

The first stage of our scheme, whose task is to collect compressive measurements like in [2–8], can be viewed as the general front-end typical of sub-Nyquist or Finite Rate of Innovation (FRI) sampling

[1,9], consisting of a P-channel filter-bank that computes

$$c_p[n] \triangleq \langle x(t), \psi_p(t-nD)\rangle, \quad p=1,\cdots,P, \tag{5}$$

where D is the shift of observation window. We call this architecture the Compressive Multichannel Sampling (CMS) scheme and the streaming samples $\{c_p[n]\}_{n\in\mathbb{Z}}^{p=1,\cdots,P}$ are used for link acquisition. However, due to the space limitation, we consider only one window of samples $\{c_p\}_{p=1}^P$ in this paper for acquisition. For a more comprehensive discussions on the CMS structure, readers are referred to the the extension of this work [10], where the sequential acquisition algorithm along with a comprehensive analysis of the complexity vis-à-vis the state-of-the-art is elaborated. In this paper, we simply compare the acquisition performance using the block of samples $\{c_p\}_{p=1}^P$ against those obtained by conventional random projections in compressed sensing and those by simple matched filtering.

To facilitate notations in derivations, we refer to a triple index (i,k,q) as the $((i-1)|\mathcal{K}||\mathcal{Q}|+(k-1)|\mathcal{Q}|+q)$th element and the vector $\boldsymbol{\alpha}$ with $(\boldsymbol{\alpha})_{(i,k,q)}=\alpha_{i,k,q}$. In this work, we consider the use of sampling kernels $\psi_p(t)$ that are linear combinations of the matched filters. The following theorem specifies the samples $\mathbf{c}\triangleq[c_1,\cdots,c_P]^T$ in relation to $\boldsymbol{\alpha}$.

Theorem 1. *Using $\psi_p(t)=\sum_{i=1}^I\sum_{k\in\mathcal{K}}\sum_{q\in\mathcal{Q}} b_{p,(i,k,q)}\phi_{i,k,q}(t)$, $p=1,\cdots,P$ as sampling kernels, the samples $\mathbf{c}$ are expressed as*

$$\mathbf{c}=\mathbf{B}\mathbf{M}\boldsymbol{\alpha}+\mathbf{v}. \tag{6}$$

1. *$\mathbf{B}$ is a $P\times I|\mathcal{K}||\mathcal{Q}|$ matrix defined as $[\mathbf{B}]_{p,(i,k,q)}\triangleq b_{p,(i,k,q)}$;*
2. *$\mathbf{M}$ is a matrix with the $[(i',k',q'),(i,k,q)]$th entry given by*

$$R_{\phi_{i',k',q'}\phi_{i,k,q}}=e^{-jk\Delta\omega q'\Delta\tau}R^{(k-k')}_{\phi_{i'},\phi_i}\left[(q'-q)\Delta\tau\right],$$

where $R^{(k-k')}_{\phi_{i'},\phi_i}(\Delta t)$ is the ambiguity function

$$R^{(k-k')}_{\phi_{i'},\phi_i}(\Delta t)=\int\phi^*_{i'}(t)\phi_i(t-\Delta t)e^{\mathrm{i}(k-k')\Delta\omega t}\mathrm{d}t. \tag{7}$$

3. *The noise covariance is $\mathbf{R}_{vv}=\mathbb{E}(\mathbf{v}\mathbf{v}^H)=\sigma^2\mathbf{B}\mathbf{M}\mathbf{B}^H$.*

Proof. The proof is omitted because it is tedious but straightforward. □

The description in this paper considers an analog front-end that aims at reducing the samples processed per observation window. In practical settings, this filterbank (similar to that used in [9]) can be implemented in an analog circuit only for some specific $\phi_i(t)$. In other cases, an equivalent digital filterbank operating at the Nyquist rate is the only viable implementation. We note, however, that even if the purpose of the front-end is not a reduced rate analog-to-digital conversion, this stage is still useful as a *data compressor* to create a more manageable record for the following estimation/detection stage. This notion generalizes and improves what was used or implied in other SR-GLRT works [2,3,8]. The freedom in choosing $\mathbf{B}$ also allows us to optimize link acquisition performance, which is discussed in details in Section 5. Now we define the full support set

$$\mathcal{A}\triangleq\{(i,k,q):\ (k,q)\in\mathcal{A}_i, i\in\mathcal{I}\}. \tag{8}$$

To reflect the link vector as a function of the link parameter set $\mathcal{A}$, we re-write $\boldsymbol{\alpha}_{\mathcal{A}}\triangleq\boldsymbol{\alpha}$ and refer to this vector as the *true link vector*.

4. SPARSITY REGULARIZED GLRT (SR-GLRT)

The SR-GLRT scheme tackles the link acquisition problem exploiting the compressive observation model given in Theorem 1. The goal of link acquisition is to discriminate the true pattern $\mathcal{A}$ against all possible patterns $\mathcal{S}\neq\mathcal{A}$ as a hypothesis testing problem

$$\mathcal{H}_{\mathcal{S}}:\ \ \mathbf{c}=\mathbf{B}\mathbf{M}\boldsymbol{\alpha}_{\mathcal{S}}+\mathbf{v} \tag{9}$$

over all possible $\mathcal{S}$'s, where $\mathcal{S}=\varnothing$ indicates the null hypothesis.

The link acquisition is thus to detect the set $\mathcal{S}$ for all possible $\mathcal{H}_{\mathcal{S}}$'s with $\boldsymbol{\alpha}_{\mathcal{S}}$ and the noise level σ^2 being nuisance. It is well known that the traditional GLRT is to find the set $\mathcal{S}$ maximizing

$$P\left(\mathcal{H}_{\mathcal{S}}|\boldsymbol{\alpha}_{\mathcal{S}},\sigma^2\right)=\frac{1}{\pi^P\sigma^{2P}|\mathbf{R}_{\psi\psi}|}e^{-\frac{1}{\sigma^2}\|\mathbf{c}-\mathbf{B}\mathbf{M}\boldsymbol{\alpha}_{\mathcal{S}}\|^2_{\mathbf{R}^{-1}_{\psi\psi}}}$$

in the presence of unknown $\boldsymbol{\alpha}_{\mathcal{S}}$ (i.e., amplitudes and $\mathcal{S}$) and σ^2. Notorious problems are that if the size of the support $|\mathcal{A}|$ is unknown, the GLRT tends to over-parametrize, while the problem becomes combinatorial and NP-hard if $|\mathcal{A}|$ is known but large. In [2,3], the authors proposed a sparse formulation

$$\widehat{\boldsymbol{\alpha}}\triangleq\arg\min_{\boldsymbol{\alpha}}\ \|\mathbf{c}-\mathbf{B}\mathbf{M}\boldsymbol{\alpha}\|^2_{\mathbf{R}^{-1}_{\psi\psi}}+\lambda\cdot f(\boldsymbol{\alpha}), \tag{10}$$

where λ is a regularization parameter and $f(\boldsymbol{\alpha})$ is the sparsity constraint. If the $f(\boldsymbol{\alpha})=\|\cdot\|_0$, the problem is usually solved via greedy methods such as orthogonal matching pursuit (OMP) etc. A relaxation of (10) is by setting $f(\boldsymbol{\alpha})=\|\cdot\|_1$. The relaxation is analogous to the Least Absolute Shrinkage and Selection Operator (LASSO) in statistics, which can be solved by convex optimization. The set of active users $\widehat{\mathcal{I}}$ and parameters $\widehat{\mathcal{A}}_i$'s can be extracted from the estimated link vector $\widehat{\boldsymbol{\alpha}}$ to give the soft estimate of the link vector $\widehat{\boldsymbol{\alpha}}_{\widehat{\mathcal{A}}}$ with $\widehat{\mathcal{A}}=\left\{(i,k,q):(k,q)\in\widehat{\mathcal{A}}_i, i\in\widehat{\mathcal{I}}\right\}$ and the noise variance $\widehat{\sigma}^2_{\widehat{\mathcal{A}}}=\|\mathbf{c}-\mathbf{B}\mathbf{M}\widehat{\boldsymbol{\alpha}}_{\widehat{\mathcal{A}}}\|^2_{\mathbf{R}^{-1}_{\psi\psi}}/P$ [11]. Finally, the likelihood for hypothesis $\mathcal{H}_{\widehat{\mathcal{A}}}$ is $P\left(\mathcal{H}_{\widehat{\mathcal{A}}}|\widehat{\boldsymbol{\alpha}}_{\widehat{\mathcal{A}}},\widehat{\sigma}^2_{\widehat{\mathcal{A}}}\right)\propto 1/\left\|\mathbf{c}-\mathbf{B}\mathbf{M}\widehat{\boldsymbol{\alpha}}_{\widehat{\mathcal{A}}}\right\|^{2P}_{\mathbf{R}^{-1}_{\psi\psi}}$.

The hypothesis $\mathcal{H}_{\widehat{\mathcal{A}}}$ obtained above is inferred to be the true hypothesis given $\mathcal{S}=\widehat{\mathcal{A}}\neq\varnothing$. Thus, signals are detected by comparing the likelihoods between $\mathcal{H}_{\widehat{\mathcal{A}}}$ and $\mathcal{H}_{\varnothing}$, which is the foremost task of link acquisition. The likelihood ratio test is performed as

$$\eta\triangleq\frac{P\left(\mathcal{H}_{\widehat{\mathcal{A}}}|\widehat{\boldsymbol{\alpha}}_{\widehat{\mathcal{A}}},\widehat{\sigma}^2_{\widehat{\mathcal{A}}}\right)}{P\left(\mathcal{H}_{\varnothing}|\widehat{\sigma}^2_{\varnothing}\right)}=\frac{\|\mathbf{c}\|^{2P}_{\mathbf{R}^{-1}_{\psi\psi}}}{\left\|\mathbf{c}-\mathbf{B}\mathbf{M}\widehat{\boldsymbol{\alpha}}_{\widehat{\mathcal{A}}}\right\|^{2P}_{\mathbf{R}^{-1}_{\psi\psi}}}>\eta_0. \tag{11}$$

The signal is declared present $\mathcal{H}_{\widehat{\mathcal{A}}}$ when $\eta>\eta_0\geq 1$ so that the receiver knows that certain signal components have been captured in the observation. Then, from the link vector $\widehat{\boldsymbol{\alpha}}_{\widehat{\mathcal{A}}}$, we can extract the delay-Doppler pairs $\widehat{\tau}_{i,r}=q\Delta\tau$, $\widehat{\omega}_{i,r}=k\Delta\omega$ with $(k,q)\in\widehat{\mathcal{A}}_i$.

5. OPTIMAL SAMPLING KERNEL

In this section, we design the sampling kernels $\{\psi_p(t)\}_{p=1}^P$ by optimizing $b_{p,(i,k,q)}$'s to maximize the weighted average Kullback-Leibler (KL) distances between any $\mathcal{H}_{\mathcal{S}}$'s in (9). The motivation comes from Stein's lemma, stating that the ability to identify distinct supports $\mathcal{S}$ and $\mathcal{S}'$ depends on the pair-wise KL distance between $\mathcal{H}_{\mathcal{S}}$ and $\mathcal{H}_{\mathcal{S}'}$. Introducing $\mathcal{G}(\mathbf{B})\triangleq\mathbf{M}^H\mathbf{B}^H\left(\mathbf{B}\mathbf{M}\mathbf{B}^H\right)^{-1}\mathbf{B}\mathbf{M}$, the *pair-wise KL* distance is given from [12] as

$$\mathbb{D}\left(\mathcal{H}_{\mathcal{S}}\|\mathcal{H}_{\mathcal{S}'}\right)=\frac{\left(\boldsymbol{\alpha}_{\mathcal{S}}-\boldsymbol{\alpha}_{\mathcal{S}'}\right)^{H}\mathcal{G}\left(\mathbf{B}\right)\left(\boldsymbol{\alpha}_{\mathcal{S}}-\boldsymbol{\alpha}_{\mathcal{S}'}\right)}{\sigma^{2}}. \tag{12}$$

It is possible that specific choices of P and the Gram matrix $\mathbf{M}$ lead to indistinguishable sparsity patterns $\mathbb{D}\left(\mathcal{H}_{\mathcal{S}}\|\mathcal{H}_{\mathcal{S}'}\right)=0$. A non-zero pair-wise KL distance for any $\mathcal{S}\neq\mathcal{S}'$ requires $\mathrm{spark}\left(\mathcal{G}\left(\mathbf{B}\right)\right)\geq|\mathcal{S}|+|\mathcal{S}'|$, where $\mathrm{spark}\left(\cdot\right)$ is the Kruskal rank of a matrix. In other words, the design of $\mathbf{B}$ cannot cure intrinsic problems caused by the Gram matrix $\mathbf{M}$ or the choice of P.

Thus, instead of looking at the pair-wise KL distance, we propose to use the average KL distance as the metric to optimize the sampling kernels[1]. To define the average KL distance, we associate each distinct pair of supports $\mathcal{S}$ and $\mathcal{S}'$ with the weight $\gamma_{\mathcal{S},\mathcal{S}'}$. Furthermore, we associate each nuisance amplitude $\boldsymbol{\alpha}_{\mathcal{S}}$ a continuous weighting function $P(\boldsymbol{\alpha}_{\mathcal{S}})$ for any $\mathcal{S}$. Then the weighted average of all *pair-wise KL* distances is

$$\overline{\mathbb{D}}=\sum_{\mathcal{S}}\sum_{\mathcal{S}'\neq\mathcal{S}}\gamma_{\mathcal{S},\mathcal{S}'}\iint P(\boldsymbol{\alpha}_{\mathcal{S}})P(\boldsymbol{\alpha}_{\mathcal{S}'})\mathbb{D}\left(\mathcal{H}_{\mathcal{S}}\|\mathcal{H}_{\mathcal{S}'}\right)\mathrm{d}\boldsymbol{\alpha}_{\mathcal{S}}\mathrm{d}\boldsymbol{\alpha}_{\mathcal{S}'}.$$

If we use uniform weights $\gamma_{\mathcal{S},\mathcal{S}'}$ for all $\mathcal{S},\mathcal{S}'$, and weighting functions $P(\boldsymbol{\alpha}_{\mathcal{S}})=\prod_{(i,k,q)\in\mathcal{S}}P(\alpha_{i,k,q})$ over the amplitudes with $\int\boldsymbol{\alpha}_{\mathcal{S}}P(\boldsymbol{\alpha}_{\mathcal{S}})\mathrm{d}\boldsymbol{\alpha}_{\mathcal{S}}=\mathbf{0}$ and $\int|\alpha_{i,k,q}|^{2}P(\alpha_{i,k,q})\mathrm{d}\alpha_{i,k,q}=\text{constant}$, the average KL distance $\overline{\mathbb{D}}$ with proper normalization is equal to

$$\overline{\mathbb{D}}=\frac{1}{\sigma^{2}}\mathrm{Tr}\left[\mathbf{M}^{H}\mathbf{B}^{H}\left(\mathbf{B}\mathbf{M}\mathbf{B}^{H}\right)^{-1}\mathbf{B}\mathbf{M}\right]. \tag{13}$$

Finally, given P and $\mathbf{M}$, we propose an optimal $\mathbf{B}$ that maximizes $\overline{\mathbb{D}}$ if there is a unique solution; when there are multiple solutions that yield identical $\overline{\mathbb{D}}$, we further choose $\mathbf{B}$ that gives the least occurrence of events $\mathbb{D}\left(\mathcal{H}_{\mathcal{S}}\|\mathcal{H}_{\mathcal{S}'}\right)=0$ from the feasible set. We use the results in the following lemma for our optimization.

Lemma 1. **(Ratio Trace Maximization [13])** *Given two $L\times L$ positive semi-definite matrices $\mathbf{S}$ and $\mathbf{G}$, and an arbitrary $L\times P$ full column rank matrix $\mathbf{W}$, the ratio trace problem is formulated as*

$$\mathbf{W}^{\mathrm{opt}}=\arg\max_{\mathbf{W}}\mathrm{Tr}\left[\left(\mathbf{W}^{H}\mathbf{S}\mathbf{W}\right)^{-1}\mathbf{W}^{H}\mathbf{G}\mathbf{W}\right]. \tag{14}$$

The optimal $\mathbf{W}^{\mathrm{opt}}=[\mathbf{w}_1^{\mathrm{opt}},\cdots,\mathbf{w}_P^{\mathrm{opt}}]$ is given by the generalized eigenvectors $\mathbf{w}_p^{\mathrm{opt}}$, $p=1,\cdots,P$ corresponding to P largest generalized eigenvalues of the pair $(\mathbf{S},\mathbf{G})$ with $P\leq\mathrm{rank}(\mathbf{S})$.

Theorem 2. *Let $\mathbf{M}=\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^{H}$, i.e. $\boldsymbol{\Sigma}$ is the eigenvalue matrix in descending order and $\mathbf{U}$ is the eigenvector matrix of $\mathbf{M}$. Denote the following set of P principal eigenvectors with $P\leq\mathrm{rank}(\mathbf{M})$*

$$\mathcal{U}=\Big\{\mathbf{U}_P=[\mathbf{u}_1,\cdots,\mathbf{u}_P]: \tag{15}$$

$$\mathbf{M}=\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^{H},\mathbf{U}=\left[\mathbf{u}_1,\cdots,\mathbf{u}_{I|\mathcal{K}||\mathcal{Q}|}\right]\Big\} \tag{16}$$

and let $\boldsymbol{\Xi}_P$ be an arbitrary non-singular $P\times P$ matrix. When the principal eigen-vectors are unique $|\mathcal{U}|=1$, the matrix $\mathbf{B}=\boldsymbol{\Xi}_P\mathbf{U}_P^{H}$ is chosen uniquely to maximize the average KL distance $\overline{\mathbb{D}}$. When the eigenvectors are not unique $|\mathcal{U}|>1$, we choose $\widehat{\mathbf{U}}_P=\max_{\mathbf{U}_P\in\mathcal{U}}\mathrm{spark}\left(\mathbf{U}_P^{H}\right)$ to maximize the average KL distance and minimize the occurrence of events $\mathbb{D}\left(\mathcal{H}_{\mathcal{S}}\|\mathcal{H}_{\mathcal{S}'}\right)=0$.

[1]Note that the computation of the weights is done offline, and does not add complexity to the online processing.

Proof. See Appendix A. □

An extreme example of $|\mathcal{U}|>1$ is when $\{\phi_{i,k,q}(t)\}_{i=1,\cdots,I}^{k\in\mathcal{K},q\in\mathcal{Q}}$ form an orthogonal basis such that $\mathbf{M}=\mathbf{I}$. In this case, Theorem 2 is similar to the criterion in compressed sensing that aims to find a matrix with as large $s\mathrm{spark}(\mathbf{U}_P^{H})$ as possible that guarantees the recovery of any $s/2$-sparse vectors (assuming s is even).

6. NUMERICAL RESULTS

In this section, we compare the SR-GLRT performances with the CMS architecture using our optimal design versus other random projection schemes in compressed sensing, and benchmark against an $I|\mathcal{K}||\mathcal{Q}|$-channel[2] exhaustive MF bank $\{\phi_{i,k,q}(t)\}_{i\in\mathcal{I},k\in\mathcal{K},q\in\mathcal{Q}}$ followed by a peak picking stage. Specifically, the simulation uses a P-channel CMS structure with $P=100$ to acquire the samples $\mathbf{c}$ and solve the SR-GLRT using Orthogonal Matching Pursuit (OMP) for the sparse recovery. We simulate a multi-access channel with $I=10$ users, $|\mathcal{I}|=4$ out of which are active. We use a length-255 preamble $\phi_i(t)=\sum_{m=0}^{254}s_i[m]g(t-mT)$ with quasi-orthogonal BPSK symbols $s_i[m]$ of unit power and $g(t)=\mathrm{sinc}(t/T)$.

We simulate underspread channel conditions with delay spread $\tau_{\max}=4T$ and Doppler spread $\omega_{\max}=(2\pi/T)\times5\cdot10^{-3}$. This frequency offset is comparable to a 5 kHz shift when $T=1\,\mu s$. We model $\tau_{i,r}$ as uniform random variables $\tau_{i,r}\sim\mathcal{U}(0,\tau_{\max})$, and similarly $\omega_{i,r}\sim\mathcal{U}(-\omega_{\max},\omega_{\max})$, whereas the fading coefficients $R=2$ are complex Gaussian $h_{i,r}\sim\mathcal{CN}\big(0,1/(|\mathcal{I}|R)\big)$ with $\mathbb{E}(h_{i,r}h_{i',r'}^{*})=1/(|\mathcal{I}|R)\delta[i-i']\delta[r-r']$. We set the acquisition resolution to be $\Delta\tau=T/2$ and $\Delta\omega=\omega_{\max}/5$, and, therefore $\mathcal{Q}=\{0,\cdots,7\}$ and $\mathcal{K}=\{-5,\cdots,5\}$.

The Receiver Operating Characteristics (ROC) curve and the source identification rate $P(\widehat{\mathcal{I}}=\mathcal{I})$ in Fig. 1 show that the optimal sampling kernel, denoted by CMS-KL, exhibits a better performance than random designs of $\mathbf{B}$ using matrices whose entries are Gaussian (CMS-G), Bernoulli (CMS-B), or randomly selected rows of a DFT matrix (CMS-F). It can also be observed from Fig. 2 that the root mean squared error (RMSE) of the delay and Doppler estimates are improved. Furthermore, the SR-GLRT using the optimal design effectively scales down the number of measurements needed for this task than MF (i.e., $P=100\ll I|\mathcal{K}||\mathcal{Q}|=880$) and provides superior performance to the conventional MF approach at reasonable SNR, because it better decorrelates the measurements and denoise the signal rather than simply picking the peak.

7. CONCLUSIONS

In this paper we evaluated the performance of the SR-GLRT using the proposed optimal CMS structure, and compared it with random compressed sensing designs and the MF approach. It is shown that the optimal structure outperforms other random designs of conventional compressed sensing matrices, and furthermore, the SR-GLRT effectively reduces the requirement for computations and storage by handling less data while improving upon the MF approach.

A. PROOF OF THEOREM 2

By analogy with Lemma 1, we have $\mathbf{S}=\mathbf{M}$ and $\mathbf{G}=\mathbf{M}\mathbf{M}^{H}$ in (13). Let $\mathbf{B}\triangleq\begin{bmatrix}\mathbf{b}_1 & \cdots & \mathbf{b}_P\end{bmatrix}^{H}$, where $\mathbf{b}_p$ is a length-$I|\mathcal{K}||\mathcal{Q}|$

[2]In practice, matched filters across the delay $q\in\mathcal{Q}$ are implemented in time domain so the $|\mathcal{Q}|$-channels are not strictly necessary, but this does not change the fact that SR-GLRT greatly reduces the required samples.

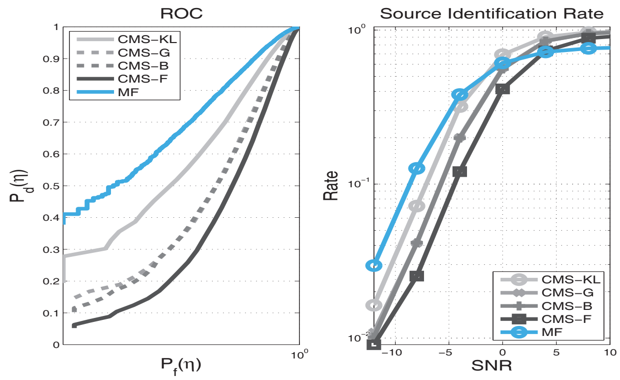

Fig. 1. ROC curve at SNR=−8 dB and successful user identification rate $P(\widehat{\mathcal{I}} = \mathcal{I})$ of the 100-channel CMS receiver, against different random designs of $\mathbf{B}$ and the MF using $I|\mathcal{K}||\mathcal{Q}| = 880$ templates.

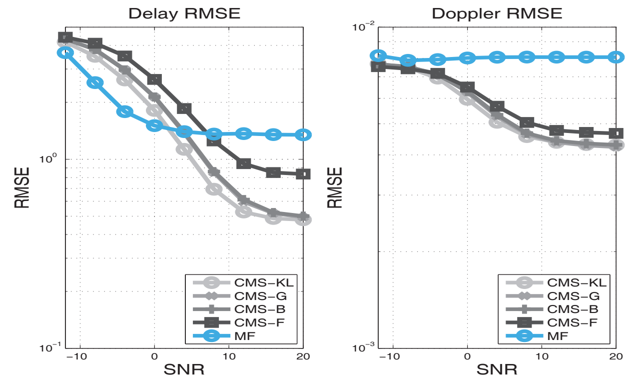

Fig. 2. Delay and Doppler estimation RMSE of the 100-channel CMS receiver, against different random designs of $\mathbf{B}$ and the MF using $I|\mathcal{K}||\mathcal{Q}| = 880$ templates.

column vector such that $\mathbf{b}_p = \mathbf{w}_p$. In this setting, according to Lemma 1, the optimal $\mathbf{b}_p$ is chosen as the generalized eigenvector of the matrix pair $(\mathbf{S}, \mathbf{G})$ such that $\mathbf{M}\mathbf{b}_p = \lambda_p \mathbf{M}\mathbf{M}^H \mathbf{b}_p$, with $p = 1, \cdots, P$ and $P \leq \mathrm{rank}(\mathbf{M})$. Using $\mathbf{M} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^H$ and the property $\mathbf{U}^H\mathbf{U} = \mathbf{I}$, we have

$$\boldsymbol{\Sigma}\mathbf{U}^H\mathbf{b}_p = \lambda_p \boldsymbol{\Sigma}\boldsymbol{\Sigma}^H\mathbf{U}^H\mathbf{b}_p, \quad p = 1, \cdots, P. \tag{17}$$

If $\mathbf{b}_p = \mathbf{u}_p$, where $\mathbf{u}_p$ is the pth column in $\mathbf{U}$, then the above relationship holds for all p because $\mathbf{u}_i^H\mathbf{u}_j = \delta[i-j]$. This gives

$$\text{L.H.S. : } \sigma_p \mathbf{U}^H\mathbf{u}_p = \sigma_p \mathbf{e}_p, \;\; \text{R.H.S. : } \lambda_p \boldsymbol{\Sigma}\boldsymbol{\Sigma}^H\mathbf{U}^H\mathbf{u}_p = \lambda_p \sigma_p^2 \mathbf{e}_p,$$

leading to a generalized eigenvalue of $\lambda_p = 1/\sigma_p$, where $\sigma_p > 0$ is the pth eigenvalue in $\boldsymbol{\Sigma}$ and $\mathbf{e}_p$ is the canonical basis with 1 in the pth entry and 0 otherwise. Denote by $\boldsymbol{\Sigma}_P$ and $\mathbf{U}_P$ the principal eigenvalue and eigenvector matrices. Then the optimal $\mathbf{B}$ is chosen as $\mathbf{B} = \boldsymbol{\Xi}_P\mathbf{U}_P^H$, where $\boldsymbol{\Xi}_P$ is an arbitrary non-singular $P \times P$ matrix. According to (13), this choice gives

$$\overline{\mathbb{D}} = \mathrm{Tr}\left(\boldsymbol{\Sigma}_P^H \underbrace{\boldsymbol{\Xi}_P^H\boldsymbol{\Xi}_P^{-H}}_{=\mathbf{I}} \boldsymbol{\Sigma}_P^{-1} \underbrace{\boldsymbol{\Xi}_P^{-1}\boldsymbol{\Xi}_P}_{=\mathbf{I}} \boldsymbol{\Sigma}_P\right) = \sum_{p=1}^{P} \sigma_p,$$

which is independent of $\boldsymbol{\Xi}_P$. If the principal eigenvectors $\mathbf{U}_P$ are unique, the above $\mathbf{B}$ uniquely maximizes the average KL distance $\overline{\mathbb{D}}$. This choice of $\mathbf{B}$ in general spreads out the individual KL distance, while the occurence of the events $\mathbb{D}\left(\mathcal{H}_{\mathcal{S}} \| \mathcal{H}_{\mathcal{S}'}\right) = 0$ is analyzed below. The same reasoning also applies when $\mathbf{U}_P$ is not unique.

Now we examine the event $\mathbb{D}\left(\mathcal{H}_{\mathcal{S}} \| \mathcal{H}_{\mathcal{S}'}\right) = 0$. Let $\boldsymbol{\beta}_{\mathcal{S}\cup\mathcal{S}'} = (\boldsymbol{\alpha}_{\mathcal{S}} - \boldsymbol{\alpha}_{\mathcal{S}'})$ be a sparse vector with $|\mathcal{S}|, |\mathcal{S}'| \leq s$, and $s \leq |\mathcal{I}|R$. Substituting the matrix $\mathbf{B} = \boldsymbol{\Xi}_P\mathbf{U}_P^H$ back to (12) and simplifying the expression, the individual KL distance is

$$\mathbb{D}\left(\mathcal{H}_{\mathcal{S}} \| \mathcal{H}_{\mathcal{S}'}\right) = \frac{1}{\sigma^2}\boldsymbol{\beta}_{\mathcal{S}\cup\mathcal{S}'}^H \mathbf{U}_P \boldsymbol{\Sigma}_P \mathbf{U}_P^H \boldsymbol{\beta}_{\mathcal{S}\cup\mathcal{S}'}, \forall \mathcal{S} \neq \mathcal{S}'$$

with $|\mathcal{S}|, |\mathcal{S}'| \leq s$. Note that $\boldsymbol{\beta}_{\mathcal{S}\cup\mathcal{S}'}$ is a $2s$-sparse vector and $\mathbb{D}\left(\mathcal{H}_{\mathcal{S}} \| \mathcal{H}_{\mathcal{S}'}\right)$ is bounded away from zero if any $2s$-sparse vectors do not fall into the null space of the matrix $\mathbf{U}_P^H$. In order to minimize the occurrence of the event $\mathbb{D}\left(\mathcal{H}_{\mathcal{S}} \| \mathcal{H}_{\mathcal{S}'}\right) = 0$, it is equivalent to maximize the Kruskal rank spark $\left(\mathbf{U}_P^H\right)$ such that $\mathbf{B}$ can recover any s-sparse vector $\boldsymbol{\alpha}_{\mathcal{S}}$ with s being maximized in this process.

B. REFERENCES

[1] K. Gedalyahu and Y. C. Eldar, "Time-Delay Estimation From Low-Rate Samples: A Union of Subspaces Approach," *IEEE Trans. Sig. Process.*, vol. 58, no. 6, pp. 3017-3031, Jun. 2010.

[2] A. Fletcher, S. Rangan and V. Goyal, "On-off Random Access Channels: A Compressed Sensing Framework," submitted, arXiv:0903.1022.

[3] L. Applebaum, W. U. Bajwa, M. F. Duarte and R. Calderbank, "Asynchronous Code-Division Random Access Using Convex Optimization", submitted, arXiv:1101.1477v1, Jun. 2011.

[4] H. Zhu and G. B. Giannakis, "Exploiting Sparse User Activity in Multiuser Detection," to appear in *IEEE Trans. on Commun.*

[5] M. Duarte, M. Davenport, M. Wakin and R. Baraniuk, "Sparse Signal Detection from Incoherent Projections," *IEEE ICASSP*, Toulouse, France, May. 2006.

[6] M. Davenport, M. Wakin and R. Baraniuk, "Detection and Estimation with Compressive Measurements," *Rice ECE Dept. Technical Report TREE 0610*, Nov. 2006.

[7] J. Haupt and R. Nowak, "Compressive Sampling for Signal Detection," *IEEE ICASSP*, Apr. 2007.

[8] Z. Wang, G. R. Arce and B. M. Sadler, "Subspace Compressive Detection for Sparse Signals," *IEEE ICASSP*, Apr. 2008.

[9] M. Vetterli, P. Marziliano and T. Blu, "Sampling Signals with Finite Rate of Innovation," *IEEE Trans. on Sig. Process.*, vol. 50, no. 6, pp. 1417-1428, Jun. 2002.

[10] X. Li, A. Rueetschi, A. Scaglione and Y. C. Eldar, "Compressive Link Acquisition in Multiuser Communications", submitted to *IEEE Trans. on Sig. Process.*.

[11] S. M. Kay, "Fundamentals of Statistical Signal Processing", Upper Saddle River, NJ: Prentice-Hall, 1998.

[12] D. Bajović, B. Sinopoli and J. Xavier, "Robust Linear Dimensionality Reduction for Hypothesis Testing with Application to Sensor Selection", *47-th Allerton Conference on Communication, Control, and Computing*, Illinois, USA, Oct. 2009.

[13] R. O. Duda, P. E. Hart and D. G. Stork, "Pattern Classification", John Wiley & Sons, second edition, 1999.