DELAY-DOPPLER ESTIMATION VIA STRUCTURED LOW-RANK MATRIX RECOVERY

Sunav Choudhary, Sajjad Beygi and Urbashi Mitra
Ming Hsieh Department of Electrical Engineering,
University of Southern California, Los Angeles, CA-90089
{sunavcho,beygihar,ubli}@usc.edu

ABSTRACT

The estimation of a narrowband time-varying channel under finite block length and transmission bandwidth is investigated. A novel method is proposed for estimation in the delay-Doppler domain by exploiting structural constraints on low-rank matrix recovery. The proposed algorithm uses Gauss-Seidel iterations on the low-rank parameterization under noisy training signal measurements. Theoretical global identifiability results for the channel leakage (due to finite block length and transmission bandwidth) are stated and the necessity of considering Doppler shift induced structure is demonstrated. Justification is provided for the choice of simulation parameters and initialization strategies to achieve good convergence rates and some ill-posed scenarios are also described. It is further shown that simple sparsity-based algorithms like basis pursuit/nuclear norm minimization do not perform well on the said constraint set for measurement operators arising out of training sequences.

***Index Terms*—** Time-varying channel, delay-Doppler, low-rank matrix recovery, alternating minimization, sparse approximation, non-convex optimization

1. INTRODUCTION

Wireless communications have enabled new systems such as intelligent traffic safety (vehicle-to-vehicle communication) [1–4], robotic networks [5], underwater surveillance systems [6] *etc.*. To establish high data rate reliable wireless communication between a transmitter and a receiver, accurate channel state information is needed at the receiver. Training-based methods, which probe the channel in time and frequency with known signals and reconstruct the channel response from the output signals, are most commonly used to accomplish this task (see [7] and references therein). There are many well-known approaches for the training-based channel estimation approach. For example, least-squares (LS) [7, 8], Wiener filters [8, 9], compressed sensing (CS) methods based on (element-wise) sparsity structure of dominant paths in the channel [4, 7, 10], and hybrid sparse and diffuse (HSD) estimators [11, 12]. The conventional LS and Wiener-filtering estimators do not take advantage of the inherent structure of the channel. The other drawback of Wiener-filtering is that the knowledge of the scattering function is required [8]; however, the scattering function is not typically known at the receiver. Often, a flat spectrum in the delay-Doppler domain is assumed, which introduces performance degradation due to the mismatch with respect to the true scattering function [8]. Compressed sensing (CS) methods [4, 7, 10] take advantage of the inherit sparsity structure of dominant components in the channel; however due to finite block length and finite transmission bandwidth the sparsity of the channel decreases in practical communication systems. This effect is known as channel *leakage* [4, 10]. It has been shown that leakage [4, 10] and basis mismatch [13] significantly degrade the performance of CS methods.

This research has been funded in part by the following grants and organizations: ONR N00014-09-1-0700, ONR N00014-15-1-2550, AFOSR FA9550-12-1-0215, DOT CA-26-7084-00, NSF CNS-1213128, NSF CCF-1410009, NSF CCF-1117896 and NSF CPS-1446901.

Contributions: We show that channel matrix in the time-delay (or Doppler-delay) domain representation is a low rank matrix with rank equal to the number of dominant paths. Using the low rank structure, we develop an alternating minimization based approach to reconstruct the channel matrix at the receiver using measurements from the training sequence. Our approach optimizes a weighted mean squared error cost function and works directly in the low-rank parametrized space. We show that the global optimum can be recovered in the absence of noise, even though the underlying problem is non-convex under the given parametrization. We explain our selection of weights to speed up convergence as compared to an unweighted MMSE estimate and highlight why a minimum nuclear norm initialization of the algorithm *fails*. Performance of the algorithm is demonstrated by numerical experiments in parameter regimes where the inverse problem is well-conditioned, and it is also shown that basis pursuit fails with gross errors.

The remainder of the paper is organized as follows. Section 2 derives the communication system model, shows that the model has a low rank property, and introduces the weighted MMSE estimation problem. Section 3 describes our alternating minimization algorithm and states its theoretical properties. Section 4 is devoted to numerical results. Section 5 concludes the paper.

Notations: We denote a scalar by x, a column vector by $\boldsymbol{x}$, and its i-th element with $x[i]$. Similarly, we denote a matrix by $\mathbf{X}$ and its (i, j)-th element by $X[i, j]$. The transpose of $\mathbf{X}$ is given by $\mathbf{X}^T$ and its conjugate transpose by $\mathbf{X}^H$. A diagonal matrix with elements $\mathbf{x}$ is written as diag$\{\mathbf{x}\}$ and the identity matrix as $\mathbf{I}$. The set of real numbers by $\mathbb{R}$, and the set of complex numbers by $\mathbb{C}$. The element-wise (Schur) product is denoted by $\odot$.The MATLAB® indexing rules are used to denote parts of a vector/matrix.

2. SYSTEM MODEL AND LOW RANK STRUCTURE

We assume that the pilot sequence $s[n]$ is transmitted over a narrowband linear time-varying channel

$$h(t, \tau) = \sum_{i=1}^{R} a_i \delta(\tau - \tau_i) e^{j2\pi\nu_i t}. \qquad (1)$$

Then, based on the derivation in [4] Section III, the discrete-time representation of the system can be written as

$$y[n] = \sum_{m=-\infty}^{+\infty} h\,[n, m]\, s[n - m] + z[n], \qquad (2)$$

where

$$h[n, m] = \iint_{-\infty}^{+\infty} h\,(t + nT_s, \tau)\, p(t - \tau + mT_s) q(-t)\, dt d\tau. \qquad (3)$$

Here $p(t)$ and $q(t)$ are respectively the transmitted pulse shape and anti-aliasing filter, and T_s is the sampling period. Note that $h[n,m]$ is causal with maximum delay $M-1$, *i.e.* $h[n,m]=0$ for $m \geq M$ and $m<0$. For simplicity, we begin by focusing on a single path channel and extend to a multi-path channel scenario in Section 3. Defining $\Pi(t)=p(t)*q(t)$, the contribution to the received signal from a single dominant path $h_i(t,\tau)=a_i\delta(t-\tau_i)e^{j2\pi\nu_i t}$ is of the form

$$y_i[n]=\sum_{m=0}^{M-1}\sum_{k=-K}^{K}\boldsymbol{H}_i\left[k,m\right]e^{j\frac{2\pi nk}{2K+1}}s[n-m]+z[n], \tag{4}$$

for $n=0,1,...,N_r-1$, where $2K+1\geq N_r$ denotes the total number sample measurements,

$$\boldsymbol{H}_i[k,m]=\sum_{n=0}^{N_r-1}\frac{h_i[n,m]}{2K+1}e^{-j\frac{2\pi nk}{2K+1}},\text{ for }|k|\leq K, \tag{5}$$

is the *discrete delay-Doppler, spreading function* of the channel, and

$$\begin{aligned}h_i[n,m]&=a_ie^{j2\pi\nu_i((n-m)T_s+\tau_i)}\Pi(mT_s-\tau_i)\\&=\boldsymbol{h}_f(n)\boldsymbol{g}(m),\end{aligned} \tag{6}$$

with $\boldsymbol{g}(m)=a_ie^{j2\pi\nu_i(\tau_i-mT_s)}\Pi(mT_s-\tau_i)$ and $\boldsymbol{h}_f(n)=e^{j2\pi n\nu_iT_s}$. We note that the leakage in the delay-Doppler plane is due to the non-zero support of $\boldsymbol{g}\in\mathbb{C}^M$ and the $(2K+1)$-point Discrete Fourier Transform (DFT) of $\boldsymbol{h}_f\in\mathbb{C}^{N_r}$. The leakage with respect to Doppler decreases with the observation length N_r, and the leakage with respect to delay decreases with the bandwidth of the transmitted signal.

Let the pilot sequence $s[n]$ be of length N_r+M-1 over $n=-(M-1),\dots,N_r-1$ and let us collect the N_r received samples in a column vector

$$\boldsymbol{y}=(y[0],\dots,y[N_r-1])^{\mathrm{T}}. \tag{7}$$

Using (6) and assuming one dominant path, the observation model in (2) can be rewritten as the vector equation $\boldsymbol{y}=\mathcal{A}\left(\boldsymbol{h}_f\boldsymbol{g}^{\mathrm{T}}\right)+\boldsymbol{z}$, where $\mathcal{A}\colon\mathbb{C}^{N_r\times M}\to\mathbb{C}^{N_r}$ is a linear operator. In particular, $\mathcal{A}$ is completely described by the set of matrices $\boldsymbol{A}_n\in\mathbb{C}^{N_r\times M}$, $n=0,1,\dots,N_r-1$ that are uniquely defined so as to satisfy $\boldsymbol{y}[n]=\mathrm{Tr}\left(\boldsymbol{A}_n\boldsymbol{h}_f\boldsymbol{g}^{\mathrm{T}}\right)+\boldsymbol{z}[n]$ and are given by

$$\boldsymbol{A}_n(k,:)=\begin{cases}\boldsymbol{0}^{\mathrm{T}}, & k\neq n,\\ \boldsymbol{s}(n:-1:n-M+1)^{\mathrm{T}}, & k=n.\end{cases} \tag{8}$$

The weighted minimum-mean-squared-error (MMSE) estimate of $(\boldsymbol{g},\boldsymbol{h}_f)$ with weight matrix $\boldsymbol{W}$, can be expressed as the solution to the optimization problem

$$\begin{aligned}&\underset{\boldsymbol{g},\boldsymbol{h}_f}{\text{minimize}} && \left\|\boldsymbol{W}\left(\boldsymbol{y}-\mathcal{A}\left(\boldsymbol{h}_f\boldsymbol{g}^{\mathrm{T}}\right)\right)\right\|_2^2\\&\text{subject to} && \boldsymbol{h}_f\in\mathcal{D}_{\boldsymbol{h}},\quad \boldsymbol{g}\in\mathcal{D}_{\boldsymbol{g}},\end{aligned} \tag{P$_1$}$$

where $\mathcal{D}_{\boldsymbol{h}}$ and $\mathcal{D}_{\boldsymbol{g}}$ respectively denote the structural restrictions on $\boldsymbol{h}_f$ and $\boldsymbol{g}$. Clearly, $\mathcal{D}_{\boldsymbol{h}}$ is completely parameterized by f since $\boldsymbol{h}_f\in\mathbb{C}^{N_r}$ is an intrinsically one dimensional non-vanishing and non-linear function of $f=\nu_iT_s$. It is also apparent that $\mathcal{D}_{\boldsymbol{g}}$ has a complicated analytical dependence on f, and to simplify analysis we relax the structural restriction $\boldsymbol{g}\in\mathcal{D}_{\boldsymbol{g}}$ to $\boldsymbol{g}\in\mathbb{C}^M$. The price we pay is that we shall need more observations ($\boldsymbol{y}$ would need to be longer) to guarantee that the weighted MMSE estimate of $(\boldsymbol{g},\boldsymbol{h}_f)$ will recover the ground truth, in the absence of noise. The *relaxed* version of Problem (P$_1$) is

$$\begin{aligned}&\underset{\boldsymbol{g},f}{\text{minimize}} && \left\|\boldsymbol{W}\left(\boldsymbol{y}-\mathcal{A}\left(\boldsymbol{h}_f\boldsymbol{g}^{\mathrm{T}}\right)\right)\right\|_2^2\\&\text{subject to} && f\in(-0.5,0.5],\quad \boldsymbol{g}\in\mathbb{C}^M\end{aligned} \tag{P$_2$}$$

Algorithm 1 AMALR

Input: $\boldsymbol{w}\geq\boldsymbol{0}$, $\boldsymbol{y}$, $\boldsymbol{s}$.
Output: $\boldsymbol{G}^{\mathrm{lo}}$ and $\boldsymbol{f}^{\mathrm{lo}}\in(-0.5,0.5]^R$
Steps:

1. Initialize $\boldsymbol{f}^{(0)}=\boldsymbol{0}\in(-0.5,0.5]^R$
2. At iteration $k\geq1$ do
   (a) $\boldsymbol{G}^{(k)}\leftarrow\underset{\boldsymbol{G}\in\mathbb{C}^{M\times R}}{\arg\operatorname{loc}\min}\ J_{\boldsymbol{w}}\left(\boldsymbol{G},\boldsymbol{f}^{(k-1)}\right)$
   (b) $\boldsymbol{f}^{(k)}\leftarrow\underset{\boldsymbol{f}\in(-0.5,0.5]^R}{\arg\operatorname{loc}\min}\ J_{\boldsymbol{w}}\left(\boldsymbol{G}^{(k)},\boldsymbol{f}\right)$
3. Repeat until the value of the objective function has converged, *i.e.* until $J_{\boldsymbol{w}}\left(\boldsymbol{G}^{(k)},\boldsymbol{f}^{(k)}\right)-J_{\boldsymbol{w}}\left(\boldsymbol{G}^{(k+1)},\boldsymbol{f}^{(k+1)}\right)<\varepsilon$.

---

where we have used periodicity of $\boldsymbol{h}_f$ *w.r.t.* f with a period of one. In the sequel, we shall only be interested in non-negative diagonal weight matrices $\boldsymbol{W}=\mathrm{diag}(\boldsymbol{w})\in\mathbb{R}^{N_r\times N_r}$ for some non-negative weight vector $\boldsymbol{w}\in\mathbb{R}^{N_r}$. For subsequent reference, we define

$$\begin{aligned}J_{\boldsymbol{w}}(\boldsymbol{g},f)&\triangleq\left\|\boldsymbol{W}\left(\boldsymbol{y}-\mathcal{A}\left(\boldsymbol{h}_f\boldsymbol{g}^{\mathrm{T}}\right)\right)\right\|_2^2\\&=\sum_{n=0}^{N_r-1}\boldsymbol{w}^2[n]\left|\boldsymbol{y}[n]-\mathrm{Tr}\left(\boldsymbol{A}_n\boldsymbol{h}_f\boldsymbol{g}^{\mathrm{T}}\right)\right|^2\end{aligned} \tag{9}$$

as our *objective function*.

If the number of dominant paths is $R>1$ then our objective function simply becomes

$$\begin{aligned}J_{\boldsymbol{w}}(\boldsymbol{G},\boldsymbol{f})&=\sum_{n=0}^{N_r-1}\boldsymbol{w}^2[n]\left|\boldsymbol{y}[n]-\sum_{i=1}^{R}\mathrm{Tr}\left(\boldsymbol{A}_n\boldsymbol{h}_{\boldsymbol{f}[i]}\boldsymbol{G}[:,i]^{\mathrm{T}}\right)\right|^2\\&=\sum_{n=0}^{N_r-1}\boldsymbol{w}^2[n]\left|\boldsymbol{y}[n]-\mathrm{Tr}\left(\boldsymbol{A}_n\boldsymbol{H}_{\boldsymbol{f}}\boldsymbol{G}^{\mathrm{T}}\right)\right|^2\end{aligned} \tag{10}$$

where $\boldsymbol{H}_{\boldsymbol{f}}\in\mathbb{C}^{M\times R}$ is such that $\boldsymbol{H}_{\boldsymbol{f}}[:,i]=\boldsymbol{h}_{\boldsymbol{f}[i]}$, the contribution to $h[:,:]$ from the i^{th} dominant path is $h_i[:,:]=\boldsymbol{h}_{\boldsymbol{f}[i]}\boldsymbol{G}[:,i]^{\mathrm{T}}$, and Problem (P$_2$) is transformed into

$$\begin{aligned}&\underset{\boldsymbol{G},\boldsymbol{f}}{\text{minimize}} && J_{\boldsymbol{w}}(\boldsymbol{G},\boldsymbol{f})\\&\text{subject to} && \boldsymbol{f}\in(-0.5,0.5]^R,\quad \boldsymbol{G}\in\mathbb{C}^{M\times R}.\end{aligned} \tag{P$_3$}$$

$\boldsymbol{H}_{\boldsymbol{f}}\in\mathbb{C}^{M\times R}$ such that $\boldsymbol{H}_{\boldsymbol{f}}[:,i]=\boldsymbol{h}_{\boldsymbol{f}[i]}$.

3. CHANNEL RECOVERY ALGORITHM

Let $(\boldsymbol{G}^{\mathrm{opt}},\boldsymbol{f}^{\mathrm{opt}})$ be a global optimum of Problem (P$_3$). It is easy to see that Problem (P$_3$) is a non-convex optimization problem *w.r.t.* the parameterization in $(\boldsymbol{G},\boldsymbol{f})$. Observing that $\boldsymbol{H}_{\boldsymbol{f}}\boldsymbol{G}^{\mathrm{T}}$ is a rank R matrix, Problem (P$_3$) can also be regarded as a low-rank matrix recovery problem [14], but with added structural constraints on the matrix factors $\boldsymbol{G}$ and $\boldsymbol{H}_{\boldsymbol{f}}$ (typically R is much smaller than M or N_r, thus making $\boldsymbol{H}_{\boldsymbol{f}}\boldsymbol{G}^{\mathrm{T}}$ a low-rank matrix). We shall adopt an alternating minimization based approach, as described in Algorithm 1, to solve for a local optimum to Problem (P$_3$). Our approach utilizes the low-rank structure of $\boldsymbol{H}_{\boldsymbol{f}}\boldsymbol{G}^{\mathrm{T}}$ directly without a further relaxation to a nuclear norm minimization problem [15] (see Section 4.2 for an explanation of why nuclear norm minimization would fail). The '$\arg\operatorname{loc}\min$' operator in step 2 of the algorithm returns a local minimizer (as opposed to the '$\arg\min$' operator that returns a global minimizer) and hence can be implemented using simple gradient descent.

Remark 1. *It is easy to see that Algorithm 1 terminates, since the objective function value $J\left(\boldsymbol{G}^{(k)},\boldsymbol{f}^{(k)}\right)$ is lower bounded by zero and decreases with each iteration (thereby, improvement in $J\left(\boldsymbol{G}^{(k)},\boldsymbol{f}^{(k)}\right)$ will fall below the stopping resolution $\varepsilon>0$ after* *a finite number of steps leading to termination of the algorithm). Indeed, we have for every $k \geq 1$,*

$$J\left(\boldsymbol{G}^{(k)}, \boldsymbol{f}^{(k)}\right) \geq J\left(\boldsymbol{G}^{(k+1)}, \boldsymbol{f}^{(k)}\right) \geq J\left(\boldsymbol{G}^{(k+1)}, \boldsymbol{f}^{(k+1)}\right) \tag{11}$$

where the first inequality is due to step (2a) *and the second inequality is due to step* (2b) *of the algorithm.*

We state a result (Theorem 1) pertaining to the output of Algorithm 1 when the elements of the pilot sequence $\boldsymbol{s}$ are drawn *i.i.d.* from a uniform distribution over the 4-QAM constellation $\left\{\pm\frac{1}{\sqrt{2}} \pm \frac{j}{\sqrt{2}}\right\}$. We have chosen the 4-QAM constellation for the ensuing simplicity of analytical computation; the results easily extend to other distributions that are symmetric about the origin and have finite fourth moments. Theorem 1 states that under no noise and sufficiently small stopping resolution, the output $\left(\boldsymbol{G}^{\text{lo}}, \boldsymbol{f}^{\text{lo}}\right)$ of Algorithm 1 is the global optimum $(\boldsymbol{G}^{\text{opt}}, \boldsymbol{f}^{\text{opt}})$ of Problem (P$_3$) whenever a *unique* global optimum exists. The proof of Theorem 1 is deferred to a future publication owing to shortage of space.

Theorem 1. *Let elements of $\boldsymbol{s} \in \mathbb{C}^{N_r+M-1}$ be drawn i.i.d. uniformly from the 4-QAM constellation $\left\{\pm\frac{1}{\sqrt{2}} \pm \frac{j}{\sqrt{2}}\right\}$ and define $\boldsymbol{\Delta} \triangleq \boldsymbol{G}\boldsymbol{H}_{\boldsymbol{f}}^{\mathrm{T}} - \boldsymbol{G}^{opt}\boldsymbol{H}_{\boldsymbol{f}^{opt}}^{\mathrm{T}} \in \mathbb{C}^{M\times N_r}$. Given a generic weight vector $\boldsymbol{w} \in \mathbb{R}^{N_r}$ in the absence of noise, $\boldsymbol{\Delta} \neq \boldsymbol{0}$ implies $\frac{\partial}{\partial \boldsymbol{G}} J_{\boldsymbol{w}}(\boldsymbol{G}, \boldsymbol{f}) \neq \boldsymbol{0}$, if the global optimum $(\boldsymbol{G}^{opt}, \boldsymbol{f}^{opt})$ is uniquely identifiable.*

Remark 2. *The unique identifiability of $(\boldsymbol{G}^{opt}, \boldsymbol{f}^{opt})$ as the solution to Problem* (P$_3$) *is necessary for any recovery guarantee, as the output of Algorithm 1. To see this, note that if there is a second global optimum $(\boldsymbol{G}_*, \boldsymbol{f}_*)$ to Problem* (P$_3$) *then there is no way to determine which of the two solutions is the correct one. In this case, we have $\boldsymbol{\Delta}_* = \boldsymbol{G}_*\boldsymbol{H}_{\boldsymbol{f}_*}^{\mathrm{T}} - \boldsymbol{G}^{opt}\boldsymbol{H}_{\boldsymbol{f}^{opt}}^{\mathrm{T}} \neq \boldsymbol{0}$, but global optimality of $(\boldsymbol{G}_*, \boldsymbol{f}_*)$ implies $\frac{\partial}{\partial \boldsymbol{G}} J_{\boldsymbol{w}}(\boldsymbol{G}_*, \boldsymbol{f}_*) = \boldsymbol{0}$ thus contradicting the conclusion of Theorem 1 if the unique identifiability clause was removed from the statement of the theorem.*

4. DISCUSSION AND SIMULATIONS

In this section we examine the trade-offs in selecting the parameters involved in the simulation based testing of Algorithm 1 and present relevant numerical results. We also comment on the convergence rate, failure of nuclear norm minimization and parameter regimes where Problem (P$_3$) is highly ill-conditioned.

4.1. Selecting Weights

Although Theorem 1 is not very sensitive to the particular choice of the weight vector $\boldsymbol{w}$, proper selection of $\boldsymbol{w} \in \mathbb{R}^{N_r}$ has a huge impact on the practical performance of Algorithm 1, *e.g.* Problem (P$_3$) represents ordinary MMSE estimation for $\boldsymbol{w} = \boldsymbol{1}$ but the observed convergence rate is extremely slow. This behavior stems from the non-linear role of $\boldsymbol{f}$ in the cost function $J_{\boldsymbol{w}}(\boldsymbol{G}, \boldsymbol{f})$ that introduces a bias in the partial derivative *w.r.t.* $\boldsymbol{G}$. To improve the convergence rate, we set the weight vector as

$$\boldsymbol{w}[n] = \begin{cases} 1, & n = 0, \\ \frac{1}{\sqrt{n}}, & 1 \leq n \leq N_r. \end{cases} \tag{12}$$

4.2. Failure of Nuclear Norm Minimization

Since Problem (P$_3$) is a low-rank matrix recovery problem [14], a plausible initialization strategy could be as follows. We solve the nuclear norm regularized (to promote a low-rank solution) least squares problem

$$\begin{aligned} \underset{\boldsymbol{X}}{\text{minimize}} \quad & \|\boldsymbol{W}(\boldsymbol{y} - \mathcal{A}(\boldsymbol{X}))\|_2^2 + \lambda\|\boldsymbol{X}\|_* \\ \text{subject to} \quad & \boldsymbol{X} \in \mathbb{C}^{N_r \times M} \end{aligned} \tag{P$_4$}$$

to obtain a global optimum $\boldsymbol{X}_*$ for some parameter $\lambda > 0$. Thereafter, we find the best rank R approximation to $\boldsymbol{X}_*$ using the singular value decomposition to obtain matrices $\boldsymbol{U} \in \mathbb{C}^{N_r\times R}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{R\times R}$ and $\boldsymbol{V} \in \mathbb{C}^{M\times R}$ respectively representing the left singular vectors, the singular values and the right singular vectors. Finally, we initialize Algorithm 1 by setting $\boldsymbol{G} = \boldsymbol{V}^*\boldsymbol{\Sigma}$ and/or $\boldsymbol{H}_{\boldsymbol{f}} = \boldsymbol{U}$.

The above initialization strategy does not work for Problem (P$_3$) because of the specific structure of the observation operator $\mathcal{A}$. According to (8), the n^{th} observation $\boldsymbol{y}[n]$ has contributions only from the n^{th} row of the unknown matrix $\boldsymbol{H}_{\boldsymbol{f}}\boldsymbol{G}^{\mathrm{T}}$, *i.e.* rows of $\boldsymbol{H}_{\boldsymbol{f}}\boldsymbol{G}^{\mathrm{T}}$ are not mixed by $\mathcal{A}$ to form $\boldsymbol{y}$. Since the nuclear norm minimization strategy ignores all structural information in $\boldsymbol{H}_{\boldsymbol{f}}$, the absence of mixing of rows of $\boldsymbol{H}_{\boldsymbol{f}}\boldsymbol{G}^{\mathrm{T}}$ in the output makes this a bad initialization.

4.3. Practical Convergence Rate and Stopping Criterion

Although Theorem 1 says that Problem (P$_3$) does not have any local optima if it has a unique global optimum $(\boldsymbol{G}^{\text{opt}}, \boldsymbol{f}^{\text{opt}})$, Algorithm 1 might not output $(\boldsymbol{G}^{\text{opt}}, \boldsymbol{f}^{\text{opt}})$ as the answer if the improvement in the value of the objective function $J_{\boldsymbol{w}}(\boldsymbol{G}, \boldsymbol{f})$ across iterations falls below the stopping resolution $\varepsilon > 0$ resulting in premature termination. In our experiments, we have observed that in such cases the algorithm's output $\left(\boldsymbol{G}^{\text{lo}}, \boldsymbol{f}^{\text{lo}}\right)$ is nowhere near the right answer $(\boldsymbol{G}^{\text{opt}}, \boldsymbol{f}^{\text{opt}})$, resulting in gross errors. Furthermore, it can be theoretically shown that the convergence rate is quite slow when the current iterate $\left(\boldsymbol{G}^{(k)}, \boldsymbol{f}^{(k)}\right)$ is far away from the right answer, thus making initialization strategies very important.

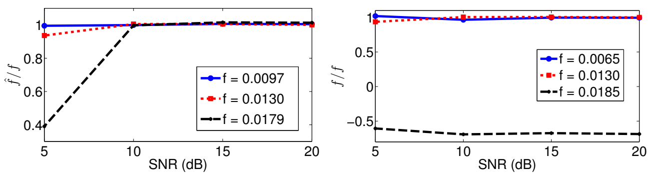
Problem (P$_3$) may become ill-conditioned if columns of $\boldsymbol{H}_{\boldsymbol{f}}$ have high correlation, *e.g.* if $R = 2$ and $\boldsymbol{f}[1]$ is close enough to $\boldsymbol{f}[2]$, then we get

$$\begin{aligned} \boldsymbol{H}_{\boldsymbol{f}}\boldsymbol{G}^{\mathrm{T}} &= \boldsymbol{H}_{\boldsymbol{f}}[:,1]\boldsymbol{G}[:,1]^{\mathrm{T}} + \boldsymbol{H}_{\boldsymbol{f}}[:,2]\boldsymbol{G}[:,2]^{\mathrm{T}} \\ &\approx \boldsymbol{H}_{\boldsymbol{f}}[:,1]\boldsymbol{G}[:,1]^{\mathrm{T}} + \boldsymbol{H}_{\boldsymbol{f}}[:,1]\boldsymbol{G}[:,2]^{\mathrm{T}} \\ &= \boldsymbol{H}_{\boldsymbol{f}}[:,1](\boldsymbol{G}[:,1] + \boldsymbol{G}[:,2])^{\mathrm{T}} \end{aligned} \tag{13}$$

implying that the rank one matrix $\boldsymbol{H}_{\boldsymbol{f}}[:,1](\boldsymbol{G}[:,1] + \boldsymbol{G}[:,2])^{\mathrm{T}}$ is nearly indistinguishable from the right answer $\boldsymbol{H}_{\boldsymbol{f}}\boldsymbol{G}^{\mathrm{T}}$ of rank two. *This is a fundamental limitation of the system model/problem setup and not related to any particular algorithm*, *i.e.* all algorithms would suffer in performance for $\boldsymbol{H}_{\boldsymbol{f}}$ with highly correlated columns. This limitation can be overcome by either redesigning the system architecture or changing the parameters of the sampling process. The condition number of the matrix $\boldsymbol{H}_{\boldsymbol{f}}^{\mathrm{H}}\boldsymbol{H}_{\boldsymbol{f}}$ provides a convenient way to quantify the overall ill-conditioning in $\boldsymbol{H}_{\boldsymbol{f}}$ for a given set of frequencies $\boldsymbol{f}$. To provide meaningful simulation results, we only consider instances of $\boldsymbol{f}$ for which the condition number of $\boldsymbol{H}_{\boldsymbol{f}}^{\mathrm{H}}\boldsymbol{H}_{\boldsymbol{f}}$ is less than 10. For this bound on the condition number, numerical computations reveal that

1. for $R = 2$, a frequency separation of at least $0.35/N_r$ is necessary, and
2. for $R = 4$, a frequency separation of at least $0.7/N_r$ is necessary.

Initialization plays an important role in determining the convergence rate and so we mention the following observations. With $R = 1$, a stopping resolution of $\varepsilon = 10^{-5}$ and $\boldsymbol{w}$ as in (12), the algorithm terminates prematurely whenever f^{opt} is farther than $1/N_r$ from the initialized frequency $f^{(0)}$, for a wide range of delay spread choices M and no observation noise. With the same $\boldsymbol{w}$ and ε but for $R = 4$, the all zero initialization $\boldsymbol{f}^{(0)} = \boldsymbol{0}$ gives good results



(a) Relative accuracy $\hat{f}/f$ of estimated frequency $\hat{f}$ versus SNR at a constant frequency factor of $\eta = 0.7$ with different oversampling factors $\rho \in \{1.1, 1.5, 2\}$.

(b) Relative accuracy $\hat{f}/f$ of estimated frequency $\hat{f}$ versus SNR at a constant oversampling factor of $\rho = 1.5$ with different frequency factors $\eta \in \{0.35, 0.7, 1\}$.

Fig. 1: Estimation accuracy results averaged over 10 realizations of the observation noise vector. Closer the value of $\hat{f}/f$ to 1, higher is the relative accuracy of the estimate.

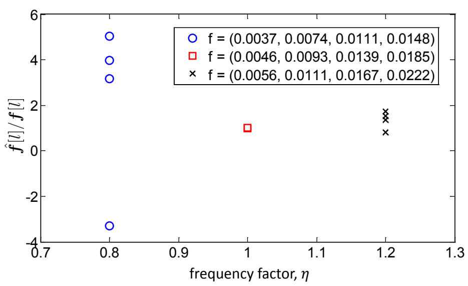


Fig. 2: Scatter plot of relative accuracy $\left\{\hat{\boldsymbol{f}}[l]/\boldsymbol{f}[l] \;\middle|\; 1 \leq l \leq 4\right\}$ of all components in the estimated frequency vector $\hat{\boldsymbol{f}}$ versus frequency factor $\eta \in \{0.8, 1, 1.2\}$ at a constant oversampling factor of $\rho = 1.5$ and 20dB SNR.Results for different η are coded by markers of different color/shape. For a given η, stronger the clustering of the markers around 1 along the y-axis, higher is the overall relative accuracy of $\hat{\boldsymbol{f}}$.

when $\boldsymbol{f}^{\text{opt}}$ has frequency separation between successive indices in the interval $[0.9/N_r, 1.1/N_r]$, *i.e.* very close to $1/N_r$, in the absence of noise and for a wide range of delay spreads. For $R \geq 5$ dominant components, a good initialization strategy is an interesting open question.

4.4. Numerical Results: One Dominant Component

We have $R = 1$ and we set the delay spread to $M = 35$. The number of free variables in the model clearly equals $(M+1) \cdot R = M + 1$. We set the number of observations as $N_r = \lfloor \rho \cdot (M+1) \cdot R \rfloor = \lfloor \rho \cdot (M+1) \rfloor$ where ρ represents the over sampling factor. We vary ρ in the range $[1, 2]$. We vary the Doppler shift f as a fraction of $1/N_r$, *i.e.* we test the algorithm against a Doppler shift of $f = \eta/N_r$ where the frequency factor η is in the range $[-1, 1]$. We assume that the observations are corrupted by additive circularly symmetric complex white Gaussian noise with signal-to-noise-ratio (SNR) between 5dB and 20dB.

Figure 1a shows the accuracy of frequency estimation $\hat{f}/f$ versus SNR for different oversampling factors ρ while keeping the frequency factor η as constant (note that changing ρ changes N_r and hence changes the numerical value of the Doppler shift f even though η is constant). Not surprisingly, larger oversampling leads to better estimation performance at low SNR. In terms of actual numerical values, we see accurate estimation above 10dB SNR even for a high Doppler frequency of $f = \frac{\eta}{\lfloor \rho \cdot (M+1) \rfloor} = 0.0179$ at $(\eta, \rho) = (0.7, 1.1)$.

Figure 1b shows the accuracy of frequency estimation $\hat{f}/f$ versus SNR for different frequency factors η while keeping the oversampling factor ρ as constant. For $\eta < 1$, the relative accuracy $\hat{f}/f$ is fairly good even at SNRs as low as 5dB. However, for $\eta = 1$, estimation fails completely at all SNRs suggesting that our algorithm may have terminated prematurely due to a slow convergence rate (see Section 4.3).

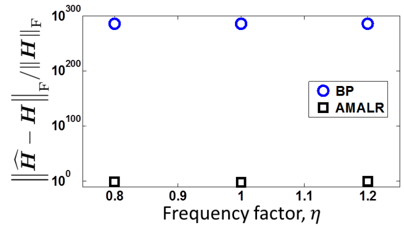


Fig. 3: Plot of normalized MSE performance of BP and AMALR versus frequency factor $\eta \in \{0.8, 1, 1.2\}$ at a constant oversampling factor of $\rho = 1.5$ and 20dB SNR. BP fails completely, due to non-utilization of structural properties. AMALR gives less than 2% normalized MSE.

4.5. Numerical Results: Multiple Dominant Components

We have $R = 4$ and we set the delay spread to $M = 35$. Borrowing the terminology from Section 4.4, we set the oversampling factor to $\rho = 1.5$ and SNR as 20dB. We set the Doppler frequency vector as $\boldsymbol{f} = \left(\frac{\eta}{N_r}, \frac{2\eta}{N_r}, \frac{3\eta}{N_r}, \frac{4\eta}{N_r}\right)$ and vary η in $[0.8, 1.2]$ where the lower bound on η is influenced by well-conditioning requirements and the upper bound is dictated by convergence rate requirements (see Section 4.3). Figure 2 shows a scatter plot of the estimation accuracy $\hat{\boldsymbol{f}}[l]/\boldsymbol{f}[l]$ of each frequency component $l \in \{1, 2, 3, 4\}$ versus the frequency factor η while keeping the oversampling factor ρ and SNR as constants. The relative accuracy of estimation $\left\{\hat{\boldsymbol{f}}[l]/\boldsymbol{f}[l] \;\middle|\; 1 \leq l \leq 4\right\}$ is very good for $\eta = 1$, is fairly good at $\eta = 1.2$, but breaks down completely for $\eta = 0.8$ which might be attributed to a combination of moderately high condition number and premature termination. Figure 3 shows the performance of basis pursuit (BP) relative to that of Algorithm 1 (AMALR) on the same set of frequencies, compared *w.r.t.* normalized MSE for the effective channel matrix $\boldsymbol{H} = \boldsymbol{H}_{\boldsymbol{f}} \boldsymbol{G}^{\mathrm{T}}$. It is clear that BP fails completely (even though (4) suggests that $\boldsymbol{H}$ is sparse in Fourier domain), and we attribute it to non-utilization of the specific low-rank factored structure of the channel matrix and the non-mixing nature of the observation operator as described in Section 4.2.

5. CONCLUSIONS

In this paper, we have investigated the estimation of a narrowband time-varying channel under finite block length and transmission bandwidth. A novel low-rank matrix recovery based formulation with structural constraints was proposed to estimate the channel in the delay-Doppler domain, utilizing separability in the Doppler and delay directions. An alternating minimization algorithm was proposed with a weighted MMSE cost function for the estimation step using noisy training signal measurements. Identifiably results for the channel leakage in both delay and Doppler directions for the channel were developed. Justification for the selection of weights and simulation parameters was given and performance was verified by simulations. Investigation of other initialization strategies is left for future research.

References

[1] D. W. Matolak, “V2V communication channels: State of knowledge, new results, and what’s next,” in *Communication Technologies for Vehicles*. Springer, 2013, pp. 1–21.

[2] S. Beygi, E. G. Ström, and U. Mitra, “Structured sparse approximation via generalized regularizers: with application to V2V channel estimation,” in *IEEE Global Telecommunications Conference (GLOBECOM)*, 2014.

[3] S. Beygi, E. G. Strom, and U. Mitra, “Geometry-based stochastic modeling and estimation of vehicle to vehicle channels,” in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 4289–4293.

[4] S. Beygi, U. Mitra, and E. G. Ström, “Nested Sparse Approximation: Structured Estimation of V2V Channels Using Geometry-Based Stochastic Channel Model,” *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4940–4955, Sep. 2015. [Online]. Available: http://arxiv.org/abs/1412.2999

[5] G. Hollinger, S. Choudhary, P. Qarabaqi, C. Murphy, U. Mitra, G. Sukhatme, M. Stojanovic, H. Singh, and F. Hover, “Communication protocols for underwater data collection using a robotic sensor network,” in *2011 IEEE GLOBECOM Workshops (GC Wkshps)*, Houston, USA, Dec. 2011, pp. 1308–1313.

[6] S. Beygi and U. Mitra, “Multi-scale multi-lag channel estimation using low rank structure of received signal,” in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 4299–4303.

[7] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, “Compressed channel sensing: A new approach to estimating sparse multipath channels,” *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.

[8] A. F. Molisch, *Wireless Communications*, 2nd ed. Wiley, Dec. 2010.

[9] S. Jnawali, S. Beygi, and H.-R. Bahrami, “RF impairments compensation and channel estimation in MIMO-OFDM systems,” in *2011 IEEE Vehicular Technology Conference (VTC Fall)*, 2011, pp. 1–5.

[10] G. Taubock, F. Hlawatsch, D. Eiwen, and H. Rauhut, “Compressive estimation of doubly selective channels in multicarrier systems: Leakage effects and sparsity-enhancing processing,” *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 255–271, 2010.

[11] N. Michelusi, U. Mitra, A. Molisch, and M. Zorzi, “UWB Sparse/Diffuse Channels, Part I: Channel Models and Bayesian Estimators,” *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5307–5319, Oct. 2012.

[12] ——, “UWB Sparse/Diffuse Channels, Part II: Estimator Analysis and Practical Channels,” *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5320–5333, Oct. 2012.

[13] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, “Sensitivity to basis mismatch in compressed sensing,” *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, 2011.

[14] D. Gross, “Recovering Low-Rank Matrices From Few Coefficients in Any Basis,” *IEEE Trans. Inf. Theory*, vol. 57, no. 3, pp. 1548–1566, 2011.

[15] E. J. Candès and B. Recht, “Exact matrix completion via convex optimization,” *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.