An integrated sparsity and model-based probabilistic framework for estimating the spatial variations of communication channels

Mehrzad Malmirchegini, Yasamin Mostofi *
Department of Electrical and Computer Engineering, University of New Mexico, United States

ARTICLE INFO

Article history:
Received 16 December 2010
Received in revised form 30 August 2011
Accepted 5 September 2011
Available online 23 September 2011

Keywords:
Estimating channel spatial variations
Sparsity-based estimation
Probabilistic channel modeling and estimation

ABSTRACT

In this paper, we consider estimating the spatial variations of a wireless channel, based on a small number of measurements. We propose an integrated sparsity and model-based channel prediction framework. Our approach properly takes advantage of both channel compressibility in the frequency domain and channel probabilistic characterization in the spatial domain. We test our framework using outdoor and indoor channel measurements. The results confirm the superior performance of the proposed integrated approach.

1. Introduction

In the past few years, the sensor network revolution has created the possibility of exploring and controlling the environment in ways not possible before [1,2]. The vision of a multi-agent robotic network cooperatively learning and adapting in harsh unknown environments to achieve a common goal is closer than ever. A mobile cooperative network needs to maintain its connectivity in order to accomplish its task. In order to achieve this, each robot should consider the impact of its motion decisions on its link qualities, when planning its trajectory. This requires each robot to assess the quality of the communication links in the locations that it has not yet visited. As a result, proper prediction of the communication signal strength in a given area, based on only a few measurements, becomes considerably important. As a robot moves around, it can learn the signal strength at positions along its motion trajectory. However, there is simply not enough time to directly measure the channel at every location in the area. Therefore, the spatial variations of a channel should be estimated based on a considerably incomplete data set, which is a challenging task. Recently, there have been significant breakthroughs in the area of non-uniform sampling theory. The new theory of compressive sampling (also known by other terms such as compressed sensing, compressive sensing or sparse sensing) shows that under certain conditions, it is possible to reconstruct a signal from a considerably incomplete set of observations, i.e., with a number of measurements much less than that predicted by the Nyquist–Shannon theorem [3,4]. This has opened new and fundamentally different possibilities, in terms of estimation and processing, in several different fields such as communications [5,6], signal processing [7–9], and sensor/mobile networks [10,11], just to name a few.

In this paper, we consider the spatial variations of a narrowband channel to a fixed station. We are interested in mapping the spatial variations of the channel over a given field of operation, based on a small number of channel measurements, as is
relevant to the operation of robotic networks. In [12,13], we proposed utilizing compressive sensing, based on the sparsity of the communication channel in the frequency domain, in order to map the channel with a small number of measurements. In [14], we proposed another framework for the estimation of channel spatial variations, based on sparse measurements. More specifically, we showed how the well-established probabilistic modeling of the channel can be utilized to estimate the channel at unvisited locations, based on a small number of measurements. In this paper, we build on our past work and further characterize the underlying tradeoffs between these two approaches. Motivated by our analysis, we then propose an integrated framework that can keep the strengths of both approaches, in order to estimate the spatial variations of a wireless channel, based on a very small number of measurements. We use several channel measurements, from both indoor and outdoor environments, in order to validate the performance of the proposed framework in terms of its prediction quality.

The paper is organized as follows. Section 2 presents the probabilistic modeling of a wireless channel and formulates the problem. In Section 3, we first summarize how channel sparsity in the frequency domain can be utilized for estimating the spatial variations of a wireless channel, based on a small number of measurements. We then analyze the impact of the underlying channel parameters on the sparsity of the channel in the frequency domain. In Section 4, we discuss our probabilistic model-based channel estimation framework. In Section 5, we show the underlying tradeoffs between our sparsity-based and model-based approaches. Section 6 then proposes an integrated framework that combines the strengths of both and shows its superior performance, using real channel measurements. We conclude in Section 7.

2. Problem setup

In the wireless communication literature, it is well established that a communication channel between two nodes can be modeled as a multi-scale dynamical system with three major dynamics: multipath fading (small-scale fading), shadowing and path loss [15]. Fig. 1 shows an example of the received signal strength across a route in our basement. The three main dynamics of the received signal power are marked on the figure. Multiple replicas of the transmitted signal can arrive at the receiver due to the reflection from the surrounding objects, resulting in a fast variation of the received signal strength called multipath fading. By spatially averaging the received signal locally and over distances that the channel can still be considered stationary, a slower dynamic emerges, which is called shadowing. Shadowing is the result of the transmitted signal being possibly blocked by a number of obstacles before reaching the receiver. Finally, by averaging over the variations of shadowing, a distance-dependent trend is seen, which is referred to as path loss.

Let $\mathcal{J}$ denote a 2D workspace. Let $\Upsilon(q)$ denote the received signal strength (power), in the transmission from a fixed transmitter at $q_b \in \mathcal{J}$, to a mobile node at $q \in \mathcal{J}$. We refer to the fixed transmitter as the base station in this paper. Table 1 contains a list of key variables used in this paper. Consider the case where the channel to the base station is narrowband [15]. Furthermore, assume that the workspace is not changing with time, i.e. the environmental features that impact the wireless transmission in the workspace are time-invariant. Then, we have the following for the baseband equivalent transmission: $\Upsilon(q) = g(q)P_T + \rho$, where $P_T$ and $g(q)$ denote the transmitted power and channel gain (square of the amplitude of the baseband equivalent channel), at position $q$, respectively and $\rho$ represents the power of the receiver thermal noise [15]. Under the assumption that the receiver noise is negligible (as compared to the signal power), $\Upsilon(q)$ becomes proportional to $g(q)$ and can be modeled as a multi-scale dynamical system with three major dynamics: multipath fading, shadowing and path loss.\footnote{Note that the assumptions of a stationary environment, as well as negligible thermal noise, are only for the purpose of mathematical derivations. In Sections 5 and 6, we demonstrate our results with several real channel measurements. These measurements naturally experience receiver thermal noise as well as other noise sources (such as quantization of the wireless card). Furthermore, small environmental features can change during our experiments. For instance, people or cars can pass by. Our framework, however, does not attempt to model and predict such changes and treats them like a disturbance.}

We can then characterize $\Upsilon(q)$ by a 2D non-stationary random field as follows [15]: $\Upsilon(q) = \Upsilon_{PL}(q)\Upsilon_{SH}(q)\Upsilon_{MP}(q)$. 

![Fig. 1. Underlying dynamics of the received signal power across a route in our basement.](image-url)
where $T_{\text{MP}}(q)$ and $T_{\text{SH}}(q)$ are random variables representing the multipath fading (normalized) and shadowing components, respectively. Furthermore, $T_{\text{PL}}(q)$ is the distance-dependent path loss: $T_{\text{PL}}(q) = \frac{K_{\text{PL}}}{|q - q_0|^h}$, where $K_{\text{PL}}$ represents the path loss power at the distance of 1m from the base station and $n_{\text{PL}}$ denotes the path loss exponent. In this model, the multipath fading component, $T_{\text{MP}}(q)$, has a unit average. In other words, averaging of the signal over multipath fading reveals the underlying shadowing and path loss components. Let $T_{\text{db}}(q) = 10\log_{10}(T(q))$ represent the received signal strength in dB. We have

$$T_{\text{db}}(q) = 10\log_{10}(K_{\text{PL}}) + T_{\text{MP, db}} - 10n_{\text{PL}}\log_{10}(|q - q_0|) + v(q) + \omega(q),$$

where $T_{\text{MP, db}} = 10\mathbb{E}[\log_{10}(T_{\text{MP}}(q))]$ is the average of the multipath fading in dB and $\omega(q) = 10\log_{10}(T_{\text{MP}}(q)) - T_{\text{MP, db}}$ is a zero-mean random variable representing the shadowing effect in dB and $v(q) = 10\log_{10}(T_{\text{SH}}(q))$ is a zero-mean random variable, independent of $\omega(q)$, which denotes the impact of multipath fading in dB, after removing its average. In the communication literature, the distributions of $T_{\text{MP}}(q)$ and $T_{\text{SH}}(q)$ (or equivalently the distributions of $\omega(q)$ and $v(q)$) are well established, based on empirical data [16]. For instance, the Nakagami distribution is shown to be a good match for the distribution of $T_{\text{MP}}(q)$ in several environments [15]. Some experimental measurements have also suggested Gaussian to be a good enough yet simple fit for the distribution of $\omega(q)$ [17,18]. We will take advantage of this Gaussian simplification later in our framework. As for the shadowing variable, lognormal is shown to be a good match for the distribution of $T_{\text{SH}}(q)$ [19]. Then, we have the following zero-mean Gaussian pdf for the distribution of $v(q): f_v(x) = \frac{1}{\sqrt{2\pi \alpha}} e^{-x^2/2\alpha}$, where $\alpha$ is the variance of the shadowing variations around path loss. As for the spatial correlation of shadowing, [19] characterizes an exponentially-decaying spatial correlation function, which is widely used: $\mathbb{E} [v(q_1)v(q_2)] = \alpha e^{-\|q_1 - q_2\|/\beta}$, for $q_1, q_2 \in \mathcal{J}$, where $\alpha$ denotes the shadowing power and $\beta$ is the correlation distance, which controls the spatial correlation of the shadowing component [19]. For more details on the aforementioned modeling of a wireless channel, any graduate book on wireless communications can be consulted [15,20]. In our overview paper [18], we also confirmed this model with several measurements collected by our robots.

**Problem Statement:** Consider the workspace $\mathcal{J}$. We assume that the workspace is discretized into an ordered set of points $\mathcal{P}$. Let vector $\mathbf{x} \in \mathbb{R}^N$ represent the corresponding received signal strength over $\mathcal{P}$, where $N = |\mathcal{P}|$. Consider the case where the received signal strength to the base station is sparsely sampled at positions $\mathcal{Q} = \{q_1, q_2, \ldots, q_k\} \subset \mathcal{J}$ over the workspace, with $K$ representing the total number of gathered measurements. The channel measurements can be gathered by one or a number of cooperative robots, making measurements along their trajectories. Thus, the measurements can be collected at the same time or at different time instants over the workspace. Define $y_i \triangleq T_{\text{db}}(q_i)$ for $1 \leq i \leq K$. Let $\mathbf{Y}_\mathcal{Q} = [y_1, \ldots, y_K]^T \in \mathbb{R}^K$ represent the vector of all the gathered channel measurements (in dB). We then have $\mathbf{Y}_\mathcal{Q} = \mathbf{F}_\mathcal{Q}\mathbf{x}$, where $\mathbf{F}_\mathcal{Q}$ represents a $K \times N$ sampling matrix that corresponds to $\mathcal{Q}$. More specifically, the $i$th row of $\mathbf{F}_\mathcal{Q}$ has all zero entries, except for the entry that corresponds to $q_i$, which becomes one. In this paper, it is our goal to predict the received signal strength, at unvisited locations (set $\mathcal{P} \setminus \mathcal{Q}$), i.e. estimate vector $\mathbf{x}$ from $\mathbf{Y}_\mathcal{Q}$, where $K \ll N$. We can express $\mathbf{Y}_\mathcal{Q}$ as follows, using Eq. (1):

$$\mathbf{Y}_\mathcal{Q} = \begin{bmatrix} 1_K & -\mathbf{D}_\mathcal{Q} \end{bmatrix} \mathbf{\theta} + \mathbf{\Theta}_\mathcal{Q} + \mathbf{\Omega}_\mathcal{Q},$$

Table 1
List of key variables.

| $\mathcal{Y}(q)$ | Received signal strength at position $q$ | $\mathbf{R}_\mathcal{Q}$ | Cov. matrix corresponding to $\mathbf{Y}_\mathcal{Q}$ |
| $g(q)$ | Channel gain at $q$ | $\mathbf{B}_\mathcal{Q}$ | Cov. matrix corresponding to shadowing component |
| $\mathcal{J}$ | Workspace | $\eta_{\text{PL}}$ | Path loss exponent |
| $\mathcal{P}$ | An ordered set of the points of the discretized workspace | $\theta$ | Path loss parameter vector |
| $\mathcal{Q}$ | Sampled set of the workspace | $\alpha$ | Shadowing power |
| $\mathbf{Y}_\mathcal{Q}$ | Gathered channel measurements in dB | $\beta$ | Correlation distance |
| $K$ | Number of gathered channel measurements ($K = |\mathcal{Q}|$) | $\alpha^2$ | Multipath power in dB |
| $\mathbf{\Phi}_\mathcal{Q}$ | Sampling matrix corresponding to $\mathcal{Q}$ | $\mathbf{D}_\mathcal{Q}$ | Distance vector to the base station in dB |
| $\mathbf{x}$ | Vector of the received signal strength corresponding to the points in $\mathcal{P}$ | $\mathbf{\Phi}_\mathcal{Q}(q)$ | Cross covariance between the received signal strength at $q$ and the points in $\mathcal{Q}$ |
| $\mathbf{X}$ | Fourier transform of $\mathbf{x}$ | $\hat{\mathbf{Y}}_{\text{db, Q}}(q)$ | Estimated channel at $q$ |

$\alpha$ as mentioned earlier, we assume that the channel field is not changing with time, in our modeling and prediction. Thus, we only need to consider the spatial variations of the measurements.

As while we pose our framework based on the prediction of the received signal strength, throughout the paper, we use the terms “channel prediction” and “received signal strength prediction” interchangeably.
where \( \mathbf{1}_K \) denotes the vector of all ones, with the length of \( K \), and \( \mathbf{D}_Q \) is the distance vector to the base station in dB (corresponding to \( Q \)): 
\[
\mathbf{D}_Q = \begin{bmatrix}
10 \log_{10}(|q_1 - q_0|) \\
\vdots \\
10 \log_{10}(|q_K - q_0|)
\end{bmatrix}.
\]
Moreover, \( \mathbf{\theta} = [K_{\text{db}} \quad n_{\text{PL}}]^T \) is the vector of the path loss parameters, 
\( \theta = [v_1, \ldots, v_K]^T \) with \( v_i = v(q_i) \) and \( \mathbf{\Sigma}_Q = [\omega_1, \ldots, \omega_K]^T \) with \( \omega_i = \omega(q_i) \), for \( i = 1, \ldots, K \), where \( v \) and \( \omega \) are as defined earlier in this section. Based on the lognormal model for shadowing, \( \mathbf{\theta} \) is a zero-mean Gaussian random vector with the covariance matrix \( \mathbf{R}_Q \in \mathbb{R}^{K \times K} \), where \( [\mathbf{R}_Q]_{ij} = \alpha e^{-||q_i-q_j||/\beta} \), for \( q_i, q_j \in Q \).

The term \( \mathbf{\Omega}_Q \) denotes the impact of multipath fading in the dB domain. As mentioned earlier, some empirical data have shown Gaussian to be a good match for the distribution of \( \omega \) [17]. For instance, Fig. 2 compares the match of both Nakagami and lognormal to the distribution of multipath fading (\( \Gamma_{\text{MP}} \)), for a stationary section of our collected data of Fig. 1, where parameter \( m_{\text{Nak}} \) denotes the shape factor of the Nakagami distribution [15]. As can be seen, Nakagami distribution with \( m_{\text{Nak}} = 1.20 \) provides a considerably good match while lognormal can be acceptable, depending on the required accuracy. Thus, in the model-based estimation framework of Section 4, we take \( \omega_5 \) to have a Gaussian distribution. Furthermore, we take \( \omega_5 \) to be uncorrelated, since there is no single model that can represent its correlation function for all the environments. In addition, multipath fading typically decorrelates considerably fast, and as such its correlation (even if learned) cannot be taken advantage of in our prediction framework. Thus, we take \( \mathbf{\Omega}_Q \) to be an uncorrelated zero-mean Gaussian vector with the covariance of \( \mathbb{E} \{ \mathbf{\Omega}_Q \mathbf{\Omega}_Q^T \} = \sigma^2 \mathbf{I}_{K \times K} \), where \( \mathbf{I}_{K \times K} \) is the \( K \times K \) identity matrix and \( \sigma^2 = \mathbb{E} \{ \omega^2(q) \} \). We then define \( \mathbf{Z}_Q = \mathbf{\theta}_Q + \mathbf{\Omega}_Q \), which is a zero-mean Gaussian vector with the covariance matrix of \( \mathbf{R}_{\text{tot},Q} = \mathbf{R}_Q + \sigma^2 \mathbf{I}_{K \times K} \).

In the next section, we consider a sparsity-based approach to channel prediction, by exploiting the sparsity of the channel in the frequency domain and applying the recent theories from the area of compressive sampling. In Section 4, we then show how the aforementioned channel modeling can be utilized for channel prediction. We show the strengths of both approaches through several experimental results and discuss the underlying tradeoffs. This then forms the basis of our proposed integrated approach in Section 6, an approach that keeps the benefits of both the sparsity-based and model-based methods.

### 3. Sparsity-based estimation of a wireless channel

In this section, we summarize our sparsity-based approach for channel estimation, using a small number of measurements. In this approach, the sparsity of the channel in the frequency domain, together with the recent results in the area of compressive sampling theory, are used for channel estimation based on a considerably incomplete data set. This approach is independent of the previous modeling of the channel. Readers are referred to [12,13] for more details.

#### 3.1. A brief summary of compressive sampling theory [3,4,21,22]

**Definition of sparse and compressible signals:** A **sparse** signal is a signal that can be represented with a small number of non-zero coefficients. Moreover, a **compressible** signal is a signal that has a transformation where most of its energy is in a very few coefficients, making it possible to approximate the rest with zero.

The new theory of compressive sampling [23] shows that, under certain conditions, a compressible signal can be reconstructed using very few observations. Consider a scenario where we are interested in recovering a vector \( \mathbf{x} \in \mathbb{R}^N \). In our case, \( \mathbf{x} \) represents the received signal strength over the field of interest. We refer to the domain of \( \mathbf{x} \) as the primal domain. Let \( \mathbf{z} \in \mathbb{R}^K \), where \( K \ll N \), represent the incomplete linear measurements of vector \( \mathbf{x} \), obtained by the sensors. We have,

\[
\mathbf{z} = \Phi \mathbf{x}.
\]
where $\Phi$ denotes the sampling matrix. Suppose that $\mathbf{x}$ has a sparse representation in another domain: $\mathbf{x} = \Gamma \mathbf{X}$, where $\Gamma$ is an invertible matrix and $\mathbf{X}$ is $S$-sparse, i.e. $|\text{supp}(\mathbf{X})| = S \ll N$. Then we will have $\mathbf{z} = \Psi \mathbf{X}$, where $\Psi = \Phi \times \Gamma$.

In general, solving for vector $\mathbf{X}$, based on the sparse observations $\mathbf{z}$, requires solving a non-convex combinatorial problem. However, if matrix $\Psi$ satisfies the Restricted Isometry Condition (RIC) [23], then solving the following $\ell_1$ optimization problem: $\min \|\mathbf{X}\|_1$, subject to $\mathbf{z} = \Psi \mathbf{X}$, can exactly recover $\mathbf{X}$ from the measurement vector $\mathbf{z}$ [3,24]. While it is not possible to define all the matrices $\Psi$ that satisfy the RIC condition, it is shown that random partial Fourier matrices satisfy the RIC with high probability [25]. The compressive sensing algorithms that reconstruct the signal based on this $\ell_1$ relaxation are typically referred to as “Basis Pursuit” [3].

3.2. Sparsity-based prediction of the channel spatial variations [12,13]

In this part, we are interested in predicting the spatial variations of the received signal strength at unvisited locations, based on sparse measurements. Our analysis of several channel measurements has shown wireless channels to be compressible in the frequency domain for several scenarios. For instance, the solid curve of Fig. 3 (left) shows a sample channel measurement across a street in San Francisco (data is courtesy of Mark Smith [26]). The dashed curve shows the sparsified version of this channel, where only 3% of its ordered Fourier coefficients are retained (ordered decreasingly) while the rest are zeroed. As can be seen, only 3% of the Fourier coefficients can capture the spatial variations of the channel well. Fig. 3 (right) measures the sparsity of the channel in the frequency domain, for different percentages of the retained Fourier coefficients. The $y$-axis shows $-10\log(\text{NMSE})$, where NMSE denotes the Normalized Mean Square Error of the difference between the channel and its sparsified version. Then, the plot characterizes how compressible this channel is. As can be seen, this channel is fairly compressible, i.e., a small percentage of the Fourier coefficients suffices for capturing the signal.

We have also investigated the sparsity of the wireless channels, using other basis, such as wavelet and Legendre [27]. While the channel can possibly be very compressible in the wavelet domain, sampling in the spatial domain and reconstructing based on the wavelet transformation results in a poor quality. This is due to the fact that spatial point-sampling and wavelet basis are not incoherent, resulting in the corresponding $\Psi$ not satisfying the RIC condition [28].

In [27], the authors show that reconstruction based on the random sampling and Legendre basis meets the RIC condition, suggesting a possible recovery strategy if the signal can be compressible in the Legendre basis. Our analysis of several real channel measurements, however, shows that Fourier domain provides a considerably more compressible representation than Legendre basis. Fig. 4, for instance, compares the compressibility of the channel measurement gathered in San Francisco, based on both Fourier and Legendre basis. It can be seen that a wireless signal is considerably more compressible in the Fourier domain. This, accompanied with the fact that the computational complexity of Fourier transformation is also considerably lower, makes Fourier an appropriate domain for our sparsity-based channel reconstruction.

Consider the workspace $\mathcal{F}$. As introduced earlier in Section 2, $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{Y}_{\mathcal{Q}}$ represent the vector of the spatial variations of the channel, corresponding to $\mathcal{P}$, and the vector of all the available channel measurements respectively (both in dB). We have $\mathbf{Y}_{\mathcal{Q}} = \Phi_{\mathcal{Q}} \mathbf{x}$, where $\Phi_{\mathcal{Q}}$ represents a $K \times N$ sampling matrix that corresponds to $\mathcal{Q}$. Then, in the context of compressive sensing (Section 3.1), vector $\mathbf{x}$ of Eq. (3) represents the spatial variations of the channel and $\mathbf{z} = \mathbf{Y}_{\mathcal{Q}}$ denotes the vector of the sparse available channel measurements. Therefore, the sparsity-based estimation of the channel spatial variations, using compressibility in the frequency domain, can be posed as follows:

$$
\hat{\mathbf{x}}_{\text{sparse}} = \arg \min \| \mathbf{X} \|_1 \quad \text{s.t.} \quad \mathbf{Y}_{\mathcal{Q}} = \Phi_{\mathcal{Q}} \Gamma \mathbf{X},
$$

where $\Gamma$ denotes the inverse Fourier matrix and $\mathbf{X}$ represents the Fourier transformation of $\mathbf{x}$. For a 1D case, we have $[\Gamma_{1D}]_{m,n} = \frac{1}{\sqrt{N}} \exp \left( \frac{2\pi i (m-1)(n-1)}{N} \right)$ for $1 \leq m, n \leq N$. For a 2D scenario, consider the case where the size of the discretized
workspace is $J_x \times J_y$, where $J_x \times J_y = N$. Define $k_1(M_1, M_2) \triangleq \lceil \frac{M_1}{M_2} \rceil$, where $\lceil \cdot \rceil$ denotes the smallest integer greater than or equal to the argument and $k_2(M_1, M_2) \triangleq M_1 - M_2 \lceil k_1(M_1, M_2) - 1 \rceil$, for arbitrary variables $M_1$ and $M_2 \in \mathbb{N}$. We have,

$$
\begin{bmatrix}
\Gamma_{2D} \end{bmatrix}_{m,n} = \frac{1}{\sqrt{J_y}} e^{\frac{2\pi i}{M} (k_1(m,J_x)-1)(n,J_x) - \frac{2\pi i}{N} (k_2(m,J_x)-1)(n,J_x)}, \text{ for } 1 \leq m, n \leq N,
$$

### 3.3. Impact of the channel underlying parameters on the variations of channel frequency response

So far, we established that a wireless channel can be considerably compressible in the Fourier domain, based on examining real channel measurements. In this section, our goal is to mathematically characterize channel compressibility in the frequency domain as this directly impacts the performance of our sparsity-based estimator. More specifically, we characterize the impact of the underlying channel parameters, i.e. parameters of path loss, shadowing and multipath, on the variations of the frequency components of a wireless channel. This analysis shows how different parameters can impact the compressibility of the channel and the resulting performance of the sparsity-based estimator. For this analysis, consider the case where a wireless channel is measured along a route. Without loss of generality, we assume that the channel is sampled across x-axis at equally-distanced positions, where the distance between two consecutive sampling positions is $d$ and the base station is located at the origin. Let $T$ denote the set of all the sampled positions. We have $T = \{q_1 = d, q_2 = 2d, \ldots, q_N = Nd\}$ and $D_T = [10 \log_{10}(d), \ldots, 10 \log_{10}(Nd)]^T$. The measurement vector, $Y_T$, can then be represented by the following, for this 1D case, based on Eq. (2),

$$
Y_T = K_{dB} \mathbf{1}_N - n_{PL} D_T + \vartheta_T + \Omega_T,
$$

where $\vartheta_T$ (impact of shadowing) is a zero-mean Gaussian random vector with the covariance matrix $R_T \in \mathbb{R}^{N \times N}$, where $[R_T]_{i,j} = \alpha : e^{-|(i-j)d|^2}$ for $1 \leq i, j \leq N$. Furthermore, for this analysis, we assume that $\Omega_T$ (impact of multipath fading) is a zero-mean Gaussian vector with the covariance matrix $\sigma^2 I_{N \times N}$ (as discussed earlier in Section 2). Next, we characterize the impact of different channel parameters on the variations of the frequency response of a wireless channel. Let $\Gamma^{-1}$ denote the 1D Fourier transform matrix with entries $[\Gamma^{-1}]_{m,n} = \frac{1}{\sqrt{N}} e^{-j \frac{2\pi}{N} (m-1)(n-1)}$ for $1 \leq m, n \leq N$. We have the following for the frequency response of the channel:

$$
Y_{T, \varphi} \triangleq \Gamma^{-1} Y_T = K_{dB} \Gamma^{-1} \mathbf{1}_N - n_{PL} \Gamma^{-1} D_T + \Gamma^{-1} \vartheta_T + \Gamma^{-1} \Omega_T
$$

$$
= \sqrt{N} K_{dB} \mathbf{e}_1 - n_{PL} \Gamma^{-1} D_T + \Gamma^{-1} \vartheta_T + \Gamma^{-1} \Omega_T
$$

where $\mathbf{e}_1$ denotes a unit vector in $\mathbb{R}^N$, with all entries zero except for the first one. We have, $Y_{T, \varphi} \sim \mathcal{N} \left( \sqrt{N} K_{dB} \mathbf{e}_1 - n_{PL} \Gamma^{-1} D_T, \alpha \Gamma^{-1} R_{\text{norm}, T} \Gamma^{-1} + \sigma^2 I_{N \times N} \right)$. As can be seen, $K_{dB}$ only affects the dc component of the frequency domain. Moreover, as $n_{PL}$ increases, the absolute value of each frequency coefficient corresponding to the second part of the path loss term in Eq. (6) increases. This implies that as $n_{PL}$ increases, on average, channel frequency response will become less compressible. As $\alpha$ and/or $\sigma^2$ increase, the variation of each of the Fourier components around its mean increases, as expected, implying lower chance of compressibility.
In order to understand the impact of $\beta$, we next characterize the variations of each component of $\Gamma^{-1}\vartheta_T$ as a function of $\beta$. For $1 \leq k \leq N$, we have the following for the normalized variance of the $k$th element of $\Gamma^{-1}\vartheta_T$.

$$
\left[ \Gamma^{-1}R_{\text{norm},T} \Gamma^{-H} \right]_{k,k} = \frac{1}{N} \sum_{1 \leq m,n \leq N} e^{-j\frac{2\pi}{N} (k-1)(m-1)} e^{-\frac{imn}{N}} e^{\frac{2\pi i}{N} (k-1)(m-1)}
\begin{array}{l}
\frac{1}{N} \sum_{0 \leq m,n \leq N-1} e^{-j\frac{2\pi}{N} (k-1)(m-n)} \rho^{|m-n|}
\end{array}
= 1 + \frac{1}{N} \sum_{i=1}^{N-1} (N-i) \left( e^{-j\frac{2\pi}{N} (k-1)i} + e^{j\frac{2\pi}{N} (k-1)i} \right) \rho^i
= 1 + \frac{1}{N} \sum_{i=1}^{N-1} (N-i) \left[ \left( \frac{\rho}{\lambda_k} \right)^i + (\rho \lambda_k)^i \right],
$$

where $\rho = e^{-\frac{\alpha}{k}}$ and $\lambda_k = e^{\frac{2\pi i}{N} (k-1)}$.

**Lemma 1.** For $\zeta \in (-1, 1)$, we have,

$$
\sum_{i=1}^{N-1} (N-i) \zeta^i = \frac{N \zeta}{1-\zeta} - \frac{\zeta - \zeta^{N+1}}{(1-\zeta)^2}.
$$

**Proof.** We have,

$$
\sum_{i=1}^{N-1} (N-i) \zeta^i = N \sum_{i=1}^{N-1} \zeta^i - \sum_{i=1}^{N-1} i \zeta^i = N \frac{\zeta - \zeta^N}{1-\zeta} - \left[ \frac{\zeta - \zeta^N}{(1-\zeta)^2} - \frac{(N-1) \zeta^N}{1-\zeta} \right]
= \frac{\zeta - N \zeta^N + \zeta^N}{(1-\zeta)^2}.
$$

Using Lemma 1 and the fact that $\rho \neq 1$ and $\lambda_k^N = (\frac{1}{\lambda_k})^N = 1$, we have:

$$
\left[ \Gamma^{-1}R_{\text{norm},T} \Gamma^{-H} \right]_{k,k} = 1 + \frac{1}{N} \sum_{i=1}^{N-1} (N-i) \left( \frac{\rho}{\lambda_k} \right)^i + \frac{1}{N} \sum_{i=1}^{N-1} (N-i)(\rho \lambda_k)^i
\begin{array}{l}
= 1 + \frac{1-\rho^N}{N} \frac{\rho}{\lambda_k} - \left( \frac{\rho}{\lambda_k} \right)^2 + \frac{1-\rho^N}{N} \rho \lambda_k - (\rho \lambda_k)^2
\end{array}
\begin{array}{l}
= 1 + \rho \left( 1 - \frac{1-\rho^N}{N} \right) \frac{1}{\lambda_k^2} + \frac{\lambda_k}{(1-\rho) \lambda_k^2}
\end{array}
\begin{array}{l}
= 1 + \rho \left( 1 - \frac{1-\rho^N}{N} \right) \frac{1}{(1-\rho^2) \lambda_k^2} - \rho^2 \frac{1}{\lambda_k^2} + \frac{\lambda_k^2}{(1-\rho) \lambda_k^2}
\end{array}
= 1 + \rho \left( \frac{1-\rho^N}{N} \right) \frac{(1+\rho^2) \lambda_k - 4\rho}{(1+\rho^2) \lambda_k^2} - \rho^2 \frac{\lambda_k^2}{(1+\rho^2) \lambda_k^2} - \rho^2 \frac{2(1-\rho^2)}{(1+\rho^2) \lambda_k^2}.
$$

where $\lambda_k = \lambda_k + \frac{1}{\lambda_k} = 2 \cos \left( \frac{2\pi i}{N} (k-1) \right)$. Let $f_\rho : [-2, 2] \to \mathbb{R}^+$ be defined as follows

$$
f_\rho(A) \triangleq 1 + \rho \left( \frac{1-\rho^N}{N} \right) \frac{(1+\rho^2)A - 4\rho}{(1+\rho^2) \lambda_k^2} - \rho^2 \frac{\lambda_k^2}{(1+\rho^2) \lambda_k^2} - \rho^2 \frac{2(1-\rho^2)}{(1+\rho^2) \lambda_k^2}
\text{ for } \rho \in (0, 1).
$$

Then, the following theorem characterizes the variance of the Fourier transformation of the shadowing term $(\Gamma^{-1}\vartheta_T)$.

**Theorem 1.** Let $g_k \triangleq \left[ \Gamma^{-1}R_{\text{H}} \Gamma^{-H} \right]_{k,k} = \alpha f_\rho(\lambda_k)$ denote the variance of the $k$th Fourier transform component of $\vartheta_T$, where $\beta \in (0, \infty)$. We have $g_k \geq g_{k+1}$ for $1 \leq k \leq \left\lfloor \frac{N-1}{2} \right\rfloor$. 

For \( 1 \leq k \leq \left\lceil \frac{N-1}{2} \right\rceil \), it can be easily confirmed that \( A_k \geq A_{k+1} \). To prove the theorem, it suffices to show that \( f_\rho(A) \) is an increasing function of \( A \). Taking the derivative with respect to \( A \) results in:

\[
\frac{d}{dA} f_\rho(A) = \rho \left( 1 - \frac{1 - \rho^N}{N} \right) (1 + \rho^2)(1 + \rho^2 - \rho A) + 2\rho\left( (\rho^3 + 1 - \rho A) + \rho(A^2 - 4) \right) (1 + \rho^2 - \rho A)^3 \\
-\rho^2(2(1 + \rho^2 - \rho A)(1 + \rho^2 - \rho A) + 2\rho(A^2 - 4)) (1 + \rho^2 - \rho A)^3 \\
= \frac{(1 + \rho^2 - \rho A)(\rho(1 + \rho^2 + 2\rho A) - \rho^2 \times 2(A + \rho))}{(1 + \rho^2 - \rho A)^3} \\
- \frac{1 - \rho^N}{\rho} \frac{(1 + \rho^2 - \rho A)(1 + \rho^2 + 2\rho A) + 2\rho^2(A^2 - 4)}{(1 + \rho^2 - \rho A)^3} \\
= \frac{(\rho - \rho^3)(1 + \rho^2 - \rho A) + \frac{1 - \rho^N}{\rho} (\rho(1 + \rho^3) + 1 + \rho^2 - \rho A) - 2\rho(1 - \rho^2)^2}{(1 + \rho^2 - \rho A)^3}. \tag{10}
\]

For \( A \in [-2, 2] \), we have \( A^2 - 4 \leq 0 \), resulting in \( 1 - \rho^2 > 0 \) for all \( \rho \in (0, 1) \). Therefore, it suffices to show that the numerator is positive. We have, \( (\rho - \rho^3)(1 + \rho^2 - \rho A) + \frac{1 - \rho^N}{\rho} (\rho(1 + \rho^3)(1 + \rho^2 - \rho A) - 2\rho(1 - \rho^2)^2) \geq \rho(1 - \rho^2)^2 + \frac{1 - \rho^N}{\rho} (1 - \rho^2)(3\rho^2 - 1) \). For \( \frac{1}{\sqrt{3}} \leq \rho < 1 \), it can be easily confirmed that the right side of the above inequality is greater than zero. For \( 0 < \rho < \frac{1}{\sqrt{3}} \), we have.

\[
\rho(1 - \rho^2)^2 + \frac{1 - \rho^N}{\rho} (1 - \rho^2)(3\rho^2 - 1) \geq \rho(1 - \rho^2)^2 + \rho(1 - \rho^2)(3\rho^2 - 1) = 2\rho^3(1 - \rho^2) > 0,
\]

which proves the theorem. \( \square \)

For \( \beta = 0 \), it can be easily confirmed that \( g_k = \alpha \) for \( 1 \leq k \leq N \). Moreover, for \( \beta = \infty \), we have \( g_1 = N\alpha \) and \( g_k = 0 \) for \( 2 \leq k \leq N \). In summary, the analysis of this part implies the following: as \( \alpha \) and/or \( \sigma^2 \) increase, the probability of having a less compressible channel increases. As \( n_{HI} \) increases, channel becomes less sparse (in the frequency domain) on average. As for the impact of \( \beta \), while our derivations are toward establishing that as \( \beta \) increases, the shadowing component becomes more compressible, more analysis is required to complete the proof. Thus, we complement this part with a simulation result and leave the full proof to our future work. Fig. 5 characterizes the sparsity of the shadowing component, for different correlation distances (\( \beta \)s). The \( y \)-axis measures the sparsity, i.e. the inverse of the Normalized Mean Square Error (in dB) between the shadowing component and its sparsified version, as a function of the percentage of the retained Fourier coefficients. The figure shows that as the correlation distance increases, the shadowing component becomes more compressible.

### 4. Model-based estimation of a wireless channel

In the previous section, we showed how channel sparsity in the frequency domain can be used for estimating channel spatial variations. Our sparsity-based approach does not utilize any model for the spatial variations of the channel.
Alternatively, the probabilistic modeling of Section 2, can be utilized to estimate channel spatial variations, based on a small number of measurements. In this section, we summarize our model-based channel estimation framework. See our previous work [14,29] for more details. Consider Eq. (1) of Section 2, where the channel over the workspace, is expressed as a function of its three underlying dynamics. Furthermore, consider Eq. (2) for an expression for the vector of all the sparse available measurements. These equations, and their corresponding details, form the basis of our probabilistic channel assessment framework. More specifically, we are interested in utilizing the aforementioned probabilistic channel modeling and predict the channel at all the unvisited locations, \( q \in \mathcal{J} \setminus \mathcal{Q} \), based on the available measurement vector \( Y_{\mathcal{Q}} \).

4.1. Estimation of the channel underlying parameters

In our model-based probabilistic framework, we first need to estimate the parameters of the underlying model (\( \theta, \alpha, \beta \) and \( \sigma^2 \)) and then use these parameters to estimate the channel. Let \( f_{Y_{\mathcal{Q}}}(Y_{\mathcal{Q}}|\theta, \alpha, \beta, \sigma^2) \) denote the conditional pdf of \( Y_{\mathcal{Q}} \), given the parameters \( \theta, \alpha, \beta \) and \( \sigma^2 \). Under the assumption of independent multipath fading variables, Eq. (2) will result in the following:

\[
f_{Y_{\mathcal{Q}}}(Y_{\mathcal{Q}}|\theta, \alpha, \beta, \sigma^2) = \frac{1}{(2\pi)^{K/2} \det[\alpha R_{\text{norm}, \mathcal{Q}}(\beta) + \sigma^2 I_{K \times K}]} \exp \left\{-\frac{1}{2} (Y_{\mathcal{Q}} - H_{\mathcal{Q}} \theta)^T (\alpha R_{\text{norm}, \mathcal{Q}}(\beta) + \sigma^2 I_{K \times K})^{-1} (Y_{\mathcal{Q}} - H_{\mathcal{Q}} \theta) \right\},
\]

where \( R_{\text{norm}, \mathcal{Q}} = \frac{1}{2} R_{\mathcal{Q}} \) denotes the normalized version of \( R_{\mathcal{Q}} \) (the covariance matrix corresponding to the shadowing component). Next, we characterize the Maximum Likelihood (ML) estimation of the underlying channel parameters. We have the following ML estimation of the parameters:

\[
[\hat{\theta}_{\text{ML}}, \hat{\alpha}_{\text{ML}}, \hat{\beta}_{\text{ML}}, \hat{\sigma}^2_{\text{ML}}] = \arg\max_{\theta, \alpha, \beta, \sigma^2} \ln(f_{Y_{\mathcal{Q}}}(Y_{\mathcal{Q}}|\theta, \alpha, \beta, \sigma^2)) = \arg\min_{\theta, \alpha, \beta, \sigma^2} (Y_{\mathcal{Q}} - H_{\mathcal{Q}} \theta)^T (\alpha R_{\text{norm}, \mathcal{Q}}(\beta) + \sigma^2 I_{K \times K})^{-1} (Y_{\mathcal{Q}} - H_{\mathcal{Q}} \theta) + \ln(\det[\alpha R_{\text{norm}, \mathcal{Q}}(\beta) + \sigma^2 I_{K \times K}]),
\]

which results in:

\[
\hat{\theta}_{\text{ML}} = \left( H_{\mathcal{Q}}^T (\alpha \hat{\alpha}_{\text{ML}} R_{\text{norm}, \mathcal{Q}}(\hat{\beta}_{\text{ML}}) + \hat{\sigma}^2_{\text{ML}} I_{K \times K})^{-1} H_{\mathcal{Q}} \right)^{-1} H_{\mathcal{Q}}^T R_{\text{norm}, \mathcal{Q}}(\hat{\beta}_{\text{ML}}) Y_{\mathcal{Q}},
\]

\[
\hat{\alpha}_{\text{ML}} = \frac{1}{K} (Y_{\mathcal{Q}} - H_{\mathcal{Q}} \hat{\theta}_{\text{ML}, \sigma^2=0})^T R_{\text{norm}, \mathcal{Q}}^{-1}(\hat{\beta}_{\text{ML}, \sigma^2=0}) (Y_{\mathcal{Q}} - H_{\mathcal{Q}} \hat{\theta}_{\text{ML}, \sigma^2=0}),
\]

\[
\hat{\beta}_{\text{ML}, \sigma^2=0} = \arg\min_{\beta} \left[ Y_{\mathcal{Q}}^T P_{\mathcal{Q}, \text{ML}}(\beta) R_{\text{norm}, \mathcal{Q}}^{-1}(\beta) P_{\mathcal{Q}, \text{ML}}(\beta) Y_{\mathcal{Q}} \right]^{-1} \ln(\det[\alpha R_{\text{norm}, \mathcal{Q}}(\beta)]),
\]

where \( P_{\mathcal{Q}, \text{ML}}(\beta) = I_{K \times K} - H_{\mathcal{Q}} (H_{\mathcal{Q}}^T R_{\text{norm}, \mathcal{Q}}^{-1}(\beta) H_{\mathcal{Q}})^{-1} H_{\mathcal{Q}}^T R_{\text{norm}, \mathcal{Q}}^{-1}(\beta) \). As can be seen, in order to estimate \( \theta \) and \( \alpha \), we first need to estimate \( \beta \), which is computationally challenging. Furthermore, finding the ML estimation of the channel parameters for the general case, where \( \sigma^2 \neq 0 \), is computationally complex. Therefore, we next devise a suboptimum but simpler estimation strategy. Let \( \chi = \alpha + \sigma^2 \) denote the sum of the shadowing and multipath powers. A Least Square (LS) estimation of \( \theta \) and \( \chi \) then results in:

\[
\hat{\theta}_{\text{LS}} = (H_{\mathcal{Q}}^T H_{\mathcal{Q}})^{-1} H_{\mathcal{Q}}^T Y_{\mathcal{Q}},
\]

\[
\hat{\chi}_{\text{LS}, \theta=\hat{\theta}_{\text{LS}}} = \frac{1}{K} Y_{\mathcal{Q}}^T (I_{K \times K} - H_{\mathcal{Q}} (H_{\mathcal{Q}}^T H_{\mathcal{Q}})^{-1} H_{\mathcal{Q}}) Y_{\mathcal{Q}} = \frac{1}{K} Y_{\mathcal{Q}}^T (I_{K \times K} - H_{\mathcal{Q}} (H_{\mathcal{Q}}^T H_{\mathcal{Q}})^{-1} H_{\mathcal{Q}}) Y_{\mathcal{Q}},
\]

where \( H_{\mathcal{Q}} \) is full rank, except for the case where the samples are equally-distanced from the base station. Since such a special case is very unlikely, we assume that \( H_{\mathcal{Q}} \) is full rank throughout the paper unless otherwise is stated. We refer to this suboptimal approach as LS throughout the paper. We next discuss a more practical but suboptimum strategy to estimate \( \beta \). Let \( \mathcal{I} = \{(i,j)\mid q_i, q_j \in \mathcal{Q} \text{ such that } \|q_i - q_j\| = l\} \) denote the pairs of points in \( \mathcal{Q} \) which are located at distance \( l \) from each other. Let \( Y_{\mathcal{Q}, \text{cent}, \text{LS}} = (I_{K \times K} - H_{\mathcal{Q}} (H_{\mathcal{Q}}^T H_{\mathcal{Q}})^{-1} H_{\mathcal{Q}}) Y_{\mathcal{Q}} \) represent the centered version of the measurement vector, when path loss parameters are estimated using the LS estimator of Eq. (14). Define \( \tilde{f}_{\mathcal{Q}}(l) \triangleq \frac{1}{|\mathcal{I}|} \sum_{(i,j) \in \mathcal{I}} |Y_{\mathcal{Q}, \text{cent}, \text{LS}}(i)| |Y_{\mathcal{Q}, \text{cent}, \text{LS}}(j)| \) to be
the numerical estimate of the spatial correlation function at distance $l$, where $|i|$ represents the cardinality of the argument set and $[i,j]$ denotes the $i$th element of the argument vector. We have $[\hat{\alpha}_{LS}, \hat{\beta}_{LS}] = \arg \min_{\alpha, \beta} \sum_{l \in \mathcal{L}_Q} w(l) \left[ \ln(\alpha e^{-l/\beta}) - \ln(\hat{r}_Q(l)) \right]^2$, where $\mathcal{L}_Q = \{|0 < \hat{r}_Q(l) < \hat{\chi}_{LS} = \hat{\chi}_{LS} + \hat{\chi}_{LS}|\}$ and $w(l)$ can be chosen based on our assessment of the accuracy of the estimation of $\hat{r}_Q(l)$. For instance, if we have very few pairs of measurements at a specific distance, then the weight should be smaller. Let $\mathcal{L}_Q = \{l_1, l_2, \ldots, l_{|\mathcal{L}_Q|}\}$ denote an ordered set of all the possible distances among the measurement points. We have the following Least Square estimator of $\alpha$ and $\beta$:

\[
\begin{bmatrix}
\ln(\hat{\alpha}_{LS}) \\
\frac{1}{\hat{\beta}_{LS}}
\end{bmatrix} = \left( \mathbf{M}^T_{\mathcal{L}_Q} \mathbf{W}_{\mathcal{L}_Q} \mathbf{M}_{\mathcal{L}_Q} \right)^{-1} \mathbf{M}^T_{\mathcal{L}_Q} \mathbf{W}_{\mathcal{L}_Q} \mathbf{b}
\]

where $\mathbf{M}_{\mathcal{L}_Q} = \begin{bmatrix} 1 & -l_1 \\ \vdots & \vdots \\ 1 & -l_{|\mathcal{L}_Q|} \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} \ln(\hat{r}_Q(l_1)) \\ \vdots \\ \ln(\hat{r}_Q(l_{|\mathcal{L}_Q|})) \end{bmatrix}$.

4.2. Model-based estimation of channel spatial variations

Once the underlying parameters of our model are estimated, channel at any arbitrary position $q \in \mathcal{J}$ can be modeled and estimated as follows [14, 29].

**Lemma 2.** Based on the available measurements $(\mathbf{Y}_Q)$ and conditioned on the channel parameters, a Gaussian distribution with mean $\hat{\mathbf{r}}_{db, Q}(q) = \mathbb{E} \{ \mathbf{r}_{db}(q) \mid \mathbf{Y}_Q, \theta, \alpha, \beta, \sigma^2 \}$ and variance $\sigma_{db, Q}(q) = \mathbb{E} \left\{ (\mathbf{r}_{db}(q) - \hat{\mathbf{r}}_{db, Q}(q))^2 \mid \theta, \alpha, \beta, \sigma^2 \right\}$ can best characterize the channel at an unvisited position $q \in \mathcal{J} \setminus \mathcal{Q}$. We then have

\[
\hat{\mathbf{r}}_{db, Q}(q) = \mathbf{h}^T(q) \theta + \mathbf{\phi}_Q(q) \mathbf{R}_{tot, Q}^{-1} (\mathbf{Y}_Q - \mathbf{H}_Q \theta)
\]

\[
\sigma_{db, Q}(q) = \alpha + \sigma^2 - \mathbf{\phi}_Q(q) \mathbf{R}_{tot, Q}^{-1} \mathbf{\phi}_Q(q).
\]

where $\mathbf{h}(q) = \begin{bmatrix} 1 & -D_{|q|} \\ \vdots & \vdots \\ 1 & -D_{|q|} \end{bmatrix}, D_{|q|} = 10 \log_{10}(\|q - q_0\|)$ and $\mathbf{\phi}_Q(q) = \alpha \left[ e^{-\frac{\|q - q_0\|}{\beta}}, \ldots, e^{-\frac{\|q - q_0\|}{\beta}} \right]^T$ denotes the cross covariance between $\mathcal{Q}$ and $q$. Therefore, the Minimum Mean Square Error (MMSE) estimation of $\mathbf{r}_{db}(q)$, assuming perfect estimation of the underlying parameters, is given by $\hat{\mathbf{r}}_{db, Q}(q)$.

**Proof.** See [29].

We then have the following, for channel prediction at an unvisited location $q \in \mathcal{J} \setminus \mathcal{Q}$, by considering the true estimated parameters:

\[
\hat{\mathbf{r}}_{db, Q}(q) = \mathbb{E} \{ \mathbf{r}_{db}(q) \mid \mathbf{Y}_Q, \theta = \hat{\theta}, \alpha = \hat{\alpha}, \beta = \hat{\beta}, \sigma^2 = \hat{\sigma}^2 \} = \mathbf{h}^T(q) \hat{\theta} + \mathbf{\phi}_Q(q) \hat{\mathbf{R}}_{tot, Q}^{-1} (\mathbf{Y}_Q - \mathbf{H}_Q \hat{\theta})
\]

where $\mathbf{\phi}_Q(q) = \left[ \hat{\alpha} e^{-\|q - q_0\|/\hat{\beta}}, \ldots, \hat{\alpha} e^{-\|q - q_0\|/\hat{\beta}} \right]^T$, $\hat{\mathbf{R}}_{tot, Q} = \hat{\alpha} \mathbf{R}_{norm, Q}(\hat{\beta}) + \hat{\sigma}^2 \mathbf{I}_{K \times K}$ and $\hat{\mathbf{r}}_{db, Q}$ denotes the estimated channel using the estimated underlying parameters.

**Remark 1.** Our probabilistic model-based channel estimator is based on predicting the shadowing and path loss components of the channel since the framework assumes uncorrelated multipath fading. Our simulation results, however, are with real channel measurements. The multipath fading components of these real measurements experience the natural correlation that is dictated by the environment.

**Remark 2.** Lemma 2 can also be derived, in the context of a kriging estimator [30].

5. Channel prediction and the underlying tradeoffs

In this section, we show the performance of the aforementioned approaches for channel estimation based on a small number of measurements. As we shall see, each approach has its own strength that can result in a better reconstruction, depending on the scenario.

Fig. 6 shows the performance of both the sparsity-based and model-based approaches for the reconstruction of the channel in Fig. 3 (left), where the $x$-axis shows the percentage of the measurements gathered (as a % of the whole area of interest). In this case, the gathered measurements are randomly distributed over the workspace. For the model-based
approach, our LS estimator of Section 4 is used for estimating the underlying parameters. This is then followed by utilizing Eq. (17) for channel prediction. It can be seen that when the number of measurements is small (less than 13.5% in this figure), the sparsity-based approach outperforms the model-based one. This makes sense as the model-based approach needs to estimate the underlying parameters. For a very small number of measurements, the error in the estimation of these parameters can be high, resulting in a performance degradation in the overall estimation. As the number of measurements increases, the model-based approach then outperforms the sparsity-based one in this case.

The model-based approach is also sensitive to the accuracy of the underlying model. In order to see this, Fig. 7 (left) shows another channel measurement in San Francisco [26]. It can be seen that this channel cannot be well characterized by only one path loss trend. As a result, we expect that the performance of our model-based approach degrade since it assumes only one path loss trend in the area of interest (see Eq. (2)). Fig. 7 (right) shows the performance of channel reconstruction in this case. It can be seen that the sparsity-based approach outperforms the model-based one in this case, due to a modeling inaccuracy of the model-based approach.

The performance of the sparsity-based approach, on the other hand, depends heavily on the compressibility of the channel in the frequency domain. For both Figs. 6 and 7 (right), the channel is considerably compressible in the Fourier domain, which is evident from the good performance of the sparsity-based approach. There could, however, be cases where the spatial variations of the channel are not that compressible in the area of interest. Fig. 8 (left) shows a 2D channel, in a hallway in our basement. Fig. 8 (right) shows the sparsity of this channel in the same way that we measured the sparsity for Fig. 3. As can be seen, this channel is not that sparse. Fig. 9 shows the 2D reconstruction of this channel. It can be seen that the performance of both approaches degrades considerably, as compared to the previous channels. This area experiences considerable multipath fading and negligible shadowing, which reduces channel compressibility. Thus, the model-based approach outperforms the sparsity one unless almost half of the area is sampled. In summary, both approaches have their strengths and can be useful in estimating a wireless channel, based on a small number of measurements. However, depending on the scenario and the percentage of the available measurements, one of the approaches may outperform the other one. Thus, in the next section, we propose an integrated approach which takes advantage of both sparsity in the frequency domain and probabilistic characterization in the spatial domain.
6. An Integrated sparsity and model-based framework for estimating channel spatial variations

So far, we discussed a sparsity-based and a model-based approach for estimating the spatial variations of a wireless channel. In Section 5, we showed the underlying tradeoffs between the two approaches and discussed the strengths and weaknesses of each. In this section, we propose a framework that integrates the strengths of both approaches, in order to achieve a more robust channel estimator with a better performance.

For the model-based approach, its performance is directly affected by the estimation of the underlying model parameters, as we saw in Section 5. Thus, in this section we also show how the sparsity of the channel in frequency domain can further be utilized to improve the estimation of the underlying model parameters.

6.1. An integrated model and sparsity-based estimator

Define $Q^c$ as the set of all the positions of $\mathcal{P}$ where channel needs to be estimated: $Q^c \triangleq \mathcal{P} \setminus Q = \{q_1', q_2', \ldots, q_{N-K}'\}$. Let $\tilde{Y}_{Q^c}$ denote the channel values at positions corresponding to $Q^c$, which are not directly measured. Based on Eq. (17), we have the following for the probability distribution of $Y_{Q^c}$, conditioned on all the gathered measurements and the underlying parameters: $f(Y_{Q^c} | Y_Q, \theta, \alpha, \beta, \sigma^2) \sim \mathcal{N}(\tilde{Y}_{Q^c}, R_{tot,Q^c}^{\theta,\alpha,\beta,\sigma^2})$ with

\[
\tilde{Y}_{Q^c} \triangleq \mathbb{E}[Y_{Q^c} | Y_Q, \theta, \alpha, \beta, \sigma^2] = H_{Q^c} \theta + \Sigma_{Q^c}^{-1}(Y_Q - H_Q \theta)
\]

and

\[
R_{tot,Q^c}^{\theta,\alpha,\beta,\sigma^2} \triangleq \mathbb{E}[(Y_{Q^c} - \tilde{Y}_{Q^c})(Y_{Q^c} - \tilde{Y}_{Q^c})^T | \theta, \alpha, \beta, \sigma^2] = R_{tot,Q^c} - \Sigma_{Q^c}^{\theta,\alpha,\beta,\sigma^2} R_{tot,Q^c}^{-1} \Sigma_{Q^c},
\]

where $\Sigma_{Q^c} = \text{COV}(Y_Q, Y_{Q^c}) \in \mathbb{R}^{K \times (N-K)}$ with entries $[\Sigma_{Q^c}]_{ij} = \alpha e^{-\frac{|q_i - q_j'|}{\beta}}$ for $1 \leq i \leq K$ and $1 \leq j \leq N - K$. Let $\tilde{Y}_{Q^c,ML}$ denote the ML estimation of $Y_{Q^c}$. We then have:
\[ \hat{Y}_{Q_c, \text{ML}} = \arg \max_{\hat{Y}_{Q_c}} f(\hat{Y}_{Q_c} | Y_Q, \theta, \alpha, \beta, \sigma^2) = \arg \max_{\hat{Y}_{Q_c}} \int f(\hat{Y}_{Q_c} | Y_Q, \theta, \alpha, \beta, \sigma^2) \]

\[ = \arg \max_{\hat{Y}_{Q_c}} \left( \frac{1}{2} \left( \hat{Y}_{Q_c} - \bar{Y}_{Q_c} \right)^T R^{-1}_{\text{tot}, Q_c} \left( \hat{Y}_{Q_c} - \bar{Y}_{Q_c} \right) \right). \tag{20} \]

Clearly, we have \( \hat{Y}_{Q_c, \text{ML}} = \hat{Y}_{Q_c} \) if no information on channel sparsity is utilized. Eq. (20) is equivalent to the following optimization problem, as a function of the Fourier coefficients of the channel (\(X\)):

\[ \hat{X}_{\text{ML}} = \arg \min_{\hat{X}} \left( \tau \|X\|_1 + \|R^{-\frac{1}{2}}_{\text{tot}, Q_c} \hat{Y}_{Q_c} - \bar{Y}_{Q_c} \|_2^2 \right) \]

\[ \text{s.t. } Y_Q = \Phi \hat{X} \tau. \]

where \( \Phi_Q \) denotes the corresponding sampling matrix, as defined in Section 2 and \( \tau \) is the inverse Fourier matrix (see Eq. (4)). \( \Phi_Q \) is then defined in a similar manner. By integrating this estimator with the sparsity-based one of Eq. (4), we have,

\[ \hat{X}_{\text{integrated}} = \arg \min_{\hat{X}} \left( \tau \|X\|_1 + \|R^{-\frac{1}{2}}_{\text{tot}, Q_c} \hat{Y}_{Q_c} - \bar{Y}_{Q_c} \|_2^2 \right) \]

\[ \text{s.t. } Y_Q = \Phi \hat{X} \tau. \]

where \( \tau \) is a weighting coefficient. This estimator has optimality in both \( \ell_1 \) and Maximum Likelihood senses. Eq. (21) can also be posed as the following unconstrained optimization problem to account for measurement noise:

\[ \hat{X}_{\text{integrated}} = \arg \min_{\hat{X}} \left( \tau \|X\|_1 + \|A_{\text{integrated}} \hat{X} - b_{\text{integrated}}\|_2^2 \right) \]

where

\[ A_{\text{integrated}} = \left[ R^{-\frac{1}{2}}_{\text{tot}, Q_c} \hat{Y}_{Q_c} \right], \quad b_{\text{integrated}} = \left[ R^{-\frac{1}{2}}_{\text{tot}, Q_c} \hat{Y}_{Q_c} \right]. \]

and \( \tau \) is a weighting coefficient. \( \tau \) can be assigned to give more or less emphasis to the sparsity part (\( \ell_1 \) optimization).

Furthermore, in [31], the authors show that for a general \( \ell_1 - \ell_2 \) problem, i.e. \( \arg \min_{\hat{X}} \left( \tau \|X\|_1 + \|A \hat{X} - b\|_2^2 \right) \), we should have \( \tau < 2\|A^T b\|_\infty \). Otherwise, the unique solution will be the zero vector. This gives us a range for valid values of \( \tau \). In some of the optimization literature and papers that have such an \( \ell_1 - \ell_2 \) problem, a pre-determined coefficient is found, by assuming some a priori information about the signal [32]. However, we do not assume any a priori information to optimize \( \tau \). We can simply choose \( \tau \) to be a fraction of the maximum allowed value. Alternatively, an adaptive weight, based on the percentage of available channel samples and the estimated underlying parameters, can also be utilized.

6.2. Estimation of the underlying model parameters using channel sparsity in the frequency domain

In the previous part, we assumed that the underlying parameters of the probabilistic model are estimated, using the ML or LS approach of Section 4. As noted in Section 5, if enough channel samples are collected, the underlying parameters can be estimated with a good enough accuracy. However, at low enough sampling rates, the error in the estimation of the underlying parameters may not be negligible. Thus, in this part, we show how the sparsity of the channel can further be used to improve the estimation of the underlying parameters.

Case of negligible multipath fading – \( \sigma^2 \approx 0 \): Depending on the environment, multipath fading can be negligible, as compared to the shadowing and path loss terms. We start by considering this case. Under this assumption, we can apply the ML estimator of channel parameters, as developed in Section 4 (Eq. (13)), and write an expression for the ML estimation of the channel, only as a function of the correlation distance \( \beta \), as follows:

\[ \hat{Y}_{Q_c, \text{ML}, \sigma^2=0}(\beta) \triangleq \mathbb{E} \left\{ Y_{Q_c} | Y_Q, \theta = \hat{\theta}_{\text{ML}, \sigma^2=0}, \alpha = \hat{\alpha}_{\text{ML}, \sigma^2=0}, \beta, \sigma^2 = 0 \right\} \]

\[ = H_{Q_c} \hat{\theta}_{\text{ML}, \sigma^2=0} + \Sigma^T_{\text{norm}, Q_c} (\beta) R^{-1}_{\text{norm}, Q_c} (\beta) (Y_Q - H_c \hat{\theta}_{\text{ML}, \sigma^2=0}) \]

\[ = \left( H_{Q_c} - \Sigma^T_{\text{norm}, Q_c} (\beta) R^{-1}_{\text{norm}, Q_c} (\beta) H_{Q_c} \right)^{-1} H_{Q_c}^T R^{-1}_{\text{norm}, Q_c} (\beta) Y_Q, \]

where \( \Sigma_{\text{norm}, Q_c} = \frac{1}{\alpha} \Sigma_{Q_c} \), with entries \( \left[ \Sigma_{\text{norm}, Q_c} \right]_{ij} = e^{-\left|\nu_{ij}^2\right|^{\alpha}} \) for \( 1 \leq i \leq K \) and \( 1 \leq j \leq N - K \). By considering the channel over the whole field, including both measured and estimated points, we will have: \( \hat{x}(\beta) = \)}
\(\Phi_Q^T Y_Q + \Phi_Q^T \hat{Y}_{\text{ML}, \sigma^2=0}(\beta)\), where \(\hat{x}\) is a vector of channel signal strengths over the whole field. Next, we utilize the sparsity of \(x\) in the frequency domain in order to estimate \(\beta\). We have,

\[
\hat{\beta}_{\text{sparsity}, \sigma^2=0} = \arg \min_\beta \| \Gamma^{-1} \hat{x}(\beta) \|_1 = \arg \min_\beta \| \Gamma^{-1} \Phi_Q^T Y_Q + \Gamma^{-1} \Phi_Q^T \hat{Y}_{\text{ML}, \sigma^2=0}(\beta) \|_1.
\]

(23)

No closed-form expression, however, exists for the optimum \(\beta\) in this case. Once \(\beta\) is estimated from Eq. (23), \(\alpha\) and \(\theta\) can be immediately estimated as follows (see Section 4):

\[
\hat{\theta}_{\text{ML}, \sigma^2=0} = \left( H_Q^T \text{R}^{-1}_{\text{norm}, Q} (\hat{\beta}_{\text{sparsity}, \sigma^2=0}) H_Q \right)^{-1} H_Q^T \text{R}^{-1}_{\text{norm}, Q} (\hat{\beta}_{\text{sparsity}, \sigma^2=0}) Y_Q,
\]

\[
\hat{\alpha}_{\text{ML}, \sigma^2=0} = \frac{1}{K} (Y_Q - H_Q \hat{\theta}_{\text{ML}, \sigma^2=0})^T \text{R}^{-1}_{\text{norm}, Q} (\hat{\beta}_{\text{sparsity}, \sigma^2=0}) (Y_Q - H_Q \hat{\theta}_{\text{ML}, \sigma^2=0}).
\]

Once the underlying parameters are estimated, we can apply the integrated estimator of Eq. (22).

Case of non-negligible multipath: If \(\sigma^2 \neq 0\), there is no closed-form expression that can express all the underlying parameters as a function of one of them, as was done in the previous part. Furthermore, the ML estimation of the underlying parameters was for the case of \(\sigma^2 = 0\) in Section 4. Thus, in this case we consider the LS estimator of the underlying parameters. We can write the following for the estimated channel, as a function of \(\alpha, \beta, \sigma^2\) and the LS estimation of the path loss parameters:

\[
\hat{Y}_{\text{ML}, LS}(\alpha, \beta, \sigma^2) \triangleq \mathbb{E} \{ Y_Q | Y_Q, \theta = \hat{\theta}_{\text{LS}}, \alpha, \beta, \sigma^2 \} = \left( \left[ H_Q \alpha \Sigma_{\text{norm}, Q}^T (\beta) \left[ \alpha \text{R}_{\text{norm}, Q}(\beta) + \sigma^2 I_{K \times K} \right]^{-1} H_Q \right] \left[ H_Q^T H_Q \right]^{-1} H_Q \right)^T \hat{\theta}_{\text{LS}} + \alpha \Sigma_{\text{norm}, Q}^T (\beta) \left[ \alpha \text{R}_{\text{norm}, Q}(\beta) + \sigma^2 I_{K \times K} \right]^{-1} \hat{\theta}_{\text{LS}}.
\]

(24)

Similar to the previous part, this results in the following for the sparsity-based estimation of \(\beta\), assuming that \(\alpha\) and \(\sigma^2\) are known:

\[
\hat{\beta}_{\text{sparsity}}(\alpha, \sigma^2) = \arg \min_\beta \| \Gamma^{-1} \Phi_Q^T Y_Q + \Gamma^{-1} \Phi_Q^T \hat{Y}_{\text{ML}, LS}(\alpha, \beta, \sigma^2) \|_1.
\]

(25)

Assuming an estimated \(\beta\), we can then estimate \(\alpha\) and \(\sigma^2\), using an LS estimator:

\[
\left[ \hat{\alpha}_{\text{LS}}(\beta), \hat{\sigma}_{\text{LS}}^2(\beta) \right] = \arg \min_{\alpha, \sigma^2 \geq 0} \sum_{l \in L_Q^c} w'(l) \left[ \alpha e^{-\beta l} + \sigma^2 \delta(l) - \hat{r}_Q(l) \right]^2
\]

(26)

where \(\delta(.)\) denotes the dirac delta function and \(L_Q^c = \{ l | 0 < \hat{r}_Q(l) \leq \hat{x}_{\text{LS}}(\hat{\theta}_{\text{LS}}) \}. The weights \(w'(l)\) can be chosen based on our assessment of the accuracy of the estimation of \(F_Q(l)\). By iteratively solving the equations given by (25) and (26), we can estimate the underlying channel parameters. After estimating the parameters, the integrated estimator of Eq. (22) can be applied to reconstruct the channel.

Next, we compare the performance of the integrated approach to that of the sparsity-based and model-based ones. To solve the convex problem of Eq. (22), we use SpaRSA [33]. SpaRSA is an efficient iterative solver for minimizing an objective function that is a weighted sum of a quadratic \((\ell_2)\) error term and a sparsity regularizer \((\ell_1)\). In each iteration, it solves an optimization subproblem, involving a quadratic term with a diagonal Hessian, in combination with the original sparsity regularizer. The readers are referred to [33] for more details.

First, we show the performance of the integrated approach for two chunks of a channel across a street in San Francisco [26]. The first chunk, channel A, is as shown in Fig. 10 (left) whereas channel B is shown in Fig. 10 (right). Fig. 11 compares the sparsity level of these two chunks. As can be seen, channel B is more sparse than channel A. Fig. 12 shows the performance of the integrated approach for both channel A (left) and B (right) and compares them with that of the original sparsity and model-based approaches. As for estimating the underlying parameters, if the number of gathered measurements is high enough such that channel spatial correlation can be properly estimated, then the LS estimator of Section 4 is used to estimate all the underlying parameters. If the number of available channel measurements is very low, on the other hand, the proposed sparsity approach of this section is utilized to estimate the underlying parameters. For this channel, we assumed that multipath is negligible, when estimating the underlying parameters. Furthermore, an adaptive weight, inversely proportional to the number of available channel samples, is used for the \(\ell_1\) term. As can be seen, the integrated approach outperforms the original approaches considerably and can provide more than 10 dB performance improvement, depending on the % of available measurements (similar performance improvements can also be seen with a fixed weight). By comparing the left and right figures, it can further be seen that the sparsity-based approach provides a better performance for channel B since it is sparse. Thus, channel A benefits more from the integrated approach. Next consider the 2D channel of Fig. 8 (left). Fig. 13 shows the performance of our integrated approach for this channel. This is an indoor channel that experiences considerable multipath. Thus, we cannot assume that \(\sigma^2 \approx 0\). For low sampling rates, where the underlying parameters cannot be accurately estimated by the LS approach of Section 4, we use our proposed
Fig. 10. Two chunks of the channel measurement of Fig. 3, which are collected across a street in San Francisco [26]—(left) channel A and (right) channel B.

Fig. 11. Characterizing the sparsity of channel A and B. The y-axis shows the inverse of the Normalized Mean Square Error (in dB) between the channel and its sparsified version, as a function of the % of the retained Fourier coefficients.

Fig. 12. Performance of the proposed integrated sparsity and model-based approach for (left) channel A (Fig. 10 (left)) and (right) channel B (Fig. 10 (right)).

7. Conclusion

In this paper, we considered estimating the spatial variations of a wireless channel, based on a small number of measurements. We considered two approaches: a sparsity-based approach and a probabilistic model-based one. The former
is model-free and utilizes the compressibility of the channel in the frequency domain, while the latter uses a multi-scale probabilistic modeling of the spatial variations of the channel. Motivated by our analysis of the underlying tradeoffs between the two approaches, we proposed an integrated channel prediction framework. In this framework, we showed how to utilize both channel sparsity in the frequency domain and probabilistic characterization in the spatial domain, in order to build a channel estimator that can keep the benefits of both approaches. We furthermore tested our framework using both outdoor and indoor channel measurements. The results confirmed the superior performance of the proposed integrated approach.

References

[27] H. Rauhut, R. Ward, Sparse Legendre expansions via\texttt{l}\textsubscript{1}-minimization, Preprint.
Mehrzad Malmirchegini received the B.S. degree in electrical engineering from Iran University of Science and Technology, Tehran, Iran, in 2004 and the M.S. degree from Sharif University of Technology, Tehran, Iran in 2007. He joined the UNM Cooperative Network Lab in 2007, where he is currently a Ph.D. student. From 2004 to 2007, he was a research assistant in the Multi Media Lab at Sharif University. In the summer of 2010, he did summer internship at Motile Robotics Incorporated, Joppa, Maryland. He also did an internship with Samsung Telecommunications America, Dallas, Texas during the summer of 2011. Mehrzad’s current research interests include wireless sensor networks, mobile communications and signal processing.

Yasamin Mostofi received the B.S. degree in electrical engineering from Sharif University of Technology, Tehran, Iran, in 1997, and the M.S. and Ph.D. degrees in the area of wireless communication systems from Stanford University, Stanford, CA, in 1999 and 2004, respectively. She is currently an assistant professor in the Department of Electrical and Computer Engineering at the University of New Mexico. Prior to that, she was a postdoctoral scholar in control and dynamical systems at the California Institute of Technology from 2004 to 2006. Her current research lies at the intersection of the two areas of communications and control in mobile sensor networks. Current research projects include communication-aware navigation and decision making in robotic networks, compressive sensing and control, obstacle mapping, robotic routers, and cooperative information processing. Dr. Mostofi is the recipient of the Presidential Early Career Award for Scientists and Engineers (PECASE) and the National Science Foundation (NSF) CAREER award. She also received Bellcore fellow-advisor award from Stanford Center for Telecommunications in 1999. She won the 2008-2009 ECE Distinguished Researcher Award at the University of New Mexico. She serves on the Control Systems Society conference editorial board since 2008. She is a member of the IEEE.