Massive MIMO mmWave Channel Estimation Using Approximate Message Passing and Laplacian Prior

Faouzi Bellili, Foad Sohrabi, and Wei Yu
Department of Electrical and Computer Engineering
University of Toronto, Toronto, Ontario M5S 3G4, Canada
Emails: faouzi.bellili@utoronto.ca, {fsohrabi, weiyu}@ece.utoronto.ca

***Abstract*—This paper tackles the problem of channel estimation in mmWave large-scale communication systems. To leverage the sparsity of mmWave MIMO channels in the beam domain, we use discrete Fourier transform (DFT) precoding and combining and recast the channel estimation problem as a compressed sensing (CS) problem. The generalized approximate message passing (GAMP) algorithm is then used to find the minimum mean square estimate (MMSE) of each entry of the unknown mmWave MIMO channel matrix. Unlike the existing works, this paper models the angular-domain channel coefficients by a Laplacian prior and accordingly establishes the closed-form expressions for all the statistical quantities that need to be updated iteratively by GAMP. Further, to render the proposed algorithm fully automated, we develop an expectation-maximization (EM)-based procedure which can be readily embedded within GAMP's iteration loop in order to learn the unknown scale parameter of the underlying Laplacian prior along with the noise variance. Numerical results indicate that the proposed EM-GAMP algorithm under a Laplacian prior yields substantial improvements both in terms of channel estimation accuracy and computational complexity as compared to the existing methods that advocate a Gaussian mixture (GM) prior.**

I. Introduction

Massive MIMO technology in which the transceivers are equipped with large-scale antennae arrays has recently attracted considerable research interests [1], [2], especially in the mmWave spectrum due to its shorter wavelength which allows more antennas to be packed in the same physical dimension [3]. One of the crucial requirements for mmWave massive MIMO systems is the acquisition of high-quality channel state information (CSI). This is, however, a challenging task particularly due to the large number of antennas at both ends of the communication link. In fact, the direct application of traditional MIMO channel estimation techniques lead to prohibitive overhead [4]. Fortunately, many recent channel measurement campaigns [5], [6] have revealed that the number of scatterers in mmWave frequencies is limited, i.e., the signal propagates from the transmitter to the receiver only through a small number of path clusters. This has motivated the use of various compressed sensing (CS) techniques to accurately estimate massive MIMO mmWave channels from relatively short pilot sequences by capitalizing on the beam-domain sparsity (see [4], [7] and references therein).

In this paper, we use the generalized approximate message-passing (GAMP) [8] for CS to estimate the single-user massive MIMO mmWave channel. Unlike the original approximate message-passing (AMP) algorithm [9], GAMP is able to accommodate arbitrary priori distributions on the components of the unknown sparse vector, further it is applicable to both linear and nonlinear observation models. This algorithm has already been applied to estimate single- and multi-user massive MIMO mmWave channels in [7] and [10], respectively. In both works, GAMP is used in conjunction with a Gaussian mixture (GM) prior on the channel coefficients.

This paper makes an observation that the use of GAMP with a Laplacian prior on the beam-domain coefficients of massive MIMO mmWave channels can lead to better channel estimation performance. Our use of the Laplacian prior is in part motivated by its wide success in Bayesian image reconstruction practices where it is widely recognized [11] that the sparsity of the discrete cosine coefficients (DCT) coefficients of natural images is well captured by a Laplacian prior. In this paper, we make a case that the Laplacian prior is also a good model for capturing the sparsity of DFT (i.e., beam-domain) coefficients of massive MIMO mmWave channels.

This paper further recognizes that the model parameters required by GAMP need to be estimated in real time in practical implementation. These include the noise variance and the scale parameter of the postulated Laplacian prior which captures the large-scale fading coefficient between the transmitter and the receiver. To address this issue, this paper devises a computationally efficient approach that learns these parameters as well using the expectation-maximization (EM) principle. The proposed EM-based approach comes with almost no additional cost since all the statistical quantities it requires are already available as by-products of GAMP while trying to reconstruct the unknown channel. Simulation results suggest that the Laplacian prior indeed leads to enhanced reconstruction performance, as compared to the GM prior advocated in [7], while speeding up the convergence of GAMP at the same time.

II. System Model

Consider a massive MIMO mmWave communication system wherein the transmitter and the receiver are equipped with M_t and M_r antennas, respectively. The components of each $\{k^{th}\}_{k=0}^{M-1}$ pilot symbol, $\mathbf{b}(k) \triangleq [b_1(k), b_2(k), \ldots, b_{M_t}(k)]^{\mathsf{T}}$, are drawn independently with $b_i(k) \sim \mathcal{CN}(0, 1/\sqrt{M_t})\ \forall i, k$. Under the block fading narrowband assumption[1], the received signal, $\mathbf{y}(k) \in \mathbb{R}^{M_r}$, can be modeled as follows:

$$\mathbf{y}(k) \;=\; \mathbf{H}\mathbf{b}(k) \;+\; \mathbf{w}(k), \quad k = 0, 1, \ldots, K-1. \tag{1}$$

Here, $\mathbf{H} \in \mathbb{C}^{M_r \times M_t}$ is the unknown channel matrix and $\mathbf{w}(k) \triangleq [w_1(k), w_2(k), \ldots w_{M_r}(k)]^{\mathsf{T}}$ is the additive white Gaussian noise vector with $w_i(k) \sim \mathcal{CN}(0, 2\sigma_w^2)\ \forall i, k$. The mmWave channels are typically modeled with few (say P) path clusters each containing a small number of sub-paths, e.g., for uniform linear array (ULA) configurations, we have:

$$\mathbf{H} \;=\; \sum_{p=1}^{P} \sum_{q=1}^{Q_p} \alpha_{p,q} \mathbf{a}_r(\Omega_{p,q}^r) \mathbf{a}_t(\Omega_{p,q}^t)^{\mathsf{H}}. \tag{2}$$

[1] We also assume that each antenna element is equipped with a dedicated RF chain. Generalization of the results to hybrid structures is left to a future work.

where $\alpha_{p,q}$ is the gain of the q^{th} sub-path within the p^{th} cluster, $\Omega^t_{p,q} = \cos(\phi_{p,q})$ and $\Omega^r_{p,q} = \cos(\theta_{p,q})$ are its directional cosine with respect to the transmit and receive antennae arrays, respectively. Moreover, $\mathbf{a}_t(\Omega)$ and $\mathbf{a}_r(\Omega)$ are the transmit and receive array response vectors. The goal of this paper is to estimate $\mathbf{H}$ given the set of observations $\{\mathbf{y}(k)\}_{k=0}^{K-1}$ and training symbols $\{\mathbf{b}(k)\}_{k=0}^{K-1}$.

The channel matrix as expressed in (2) is not visibly sparse. Expressing it in the angular domain, however, reveals that it has indeed very few dominant entries. To see this, let $\mathbf{U}_r$ and $\mathbf{U}_t$ denote the $M_r \times M_r$ and $M_t \times M_t$ unitary DFT matrices and consider the angular-domain representation of $\mathbf{H}$ [12]:

$$\widetilde{\mathbf{H}} = \mathbf{U}_r^{\mathsf{H}} \mathbf{H} \mathbf{U}_t = \sum_{p=1}^{P} \sum_{q=1}^{Q_p} \alpha_{p,q} \big[\mathbf{U}_r^{\mathsf{H}} \mathbf{a}_r(\Omega^r_{p,q})\big] \big[\mathbf{U}_t^{\mathsf{H}} \mathbf{a}_t(\Omega^t_{p,q})\big]^{\mathsf{H}}. \quad (3)$$

Intuitively, in (3), the columns of $\mathbf{U}_r$ and $\mathbf{U}_t$ act as transmit and receive beamforming vectors, respectively, capturing how much energy is present along their associated transmit/receive beams. If the angular spread is small (which is the case in mmWave bands), then each cluster have most of its energy along one particular pair of transmit/receive beamforming vectors, thereby reflecting the sparsity of the channel in the angular domain. Note, however, that $\widetilde{\mathbf{H}}$ is not *exactly* sparse but rather *approximately* sparse due the spectral leakage phenomenon.

To exploit the virtual model in (3), we apply the DFT precoder, $\mathbf{U}_t$, and transmit $\mathbf{U}_t \mathbf{b}(k)$ instead of $\mathbf{b}(k)$. We also combine the corresponding received noisy vector using $\mathbf{U}_r^{\mathsf{H}}$. In this way, the observation model is as follows:

$$\widetilde{\mathbf{y}}(k) \;=\; \mathbf{U}_r^{\mathsf{H}} \mathbf{H} \mathbf{U}_t \mathbf{b}(k) + \mathbf{U}_r^{\mathsf{H}} \mathbf{w}(k) \;=\; \widetilde{\mathbf{H}}\, \mathbf{b}(k) + \widetilde{\mathbf{w}}(k), \quad (4)$$

where $\widetilde{\mathbf{w}}(k) \triangleq \mathbf{U}_r^{\mathsf{H}} \mathbf{w}(k)$ is the resulting combined noise which has exactly the same statistics as $\mathbf{w}(k)$ since $\mathbf{U}_r$ is a unitary matrix. Consequently, the space-time observation matrix $\widetilde{\mathbf{Y}} = [\widetilde{\mathbf{y}}(0)\ \widetilde{\mathbf{y}}(1)\ \ldots\ \widetilde{\mathbf{y}}(K-1)]$ is given by:

$$\widetilde{\mathbf{Y}} \;=\; \widetilde{\mathbf{H}} \mathbf{B} \;+\; \widetilde{\mathbf{W}}, \quad (5)$$

with the matrices $\mathbf{B}$ and $\widetilde{\mathbf{W}}$ being constructed in the same way as $\mathbf{Y}$. Now, by defining $\widetilde{\mathbf{y}} \triangleq \mathrm{vec}(\widetilde{\mathbf{Y}})$, $\widetilde{\mathbf{x}} \triangleq \mathrm{vec}(\widetilde{\mathbf{H}})$, $\widetilde{\mathbf{w}} \triangleq \mathrm{vec}(\widetilde{\mathbf{W}})$, and $\widetilde{\mathbf{A}} \triangleq \mathbf{B}^{\mathsf{T}} \otimes \mathbf{I}_{M_r}$, with $\otimes$ returning the Kronecker product, it follows from (5) that:

$$\widetilde{\mathbf{y}} \;=\; \widetilde{\mathbf{A}} \widetilde{\mathbf{x}} \;+\; \widetilde{\mathbf{w}}, \quad (6)$$

Recall here that the vector $\widetilde{\mathbf{x}}$ is approximately sparse due to the approximate sparsity of $\widetilde{\mathbf{H}}$. This paper captures the underlying sparsity by a Laplacian prior. Since the Laplace distribution is defined for real-valued RVs only, we transform the complex model in (6) as follows:

$$\underbrace{\begin{bmatrix} \Re\{\widetilde{\mathbf{y}}\} \\ \Im\{\widetilde{\mathbf{y}}\} \end{bmatrix}}_{\triangleq \mathbf{y}} = \underbrace{\begin{bmatrix} \Re\{\widetilde{\mathbf{A}}\} & -\Im\{\widetilde{\mathbf{A}}\} \\ \Im\{\widetilde{\mathbf{A}}\} & \Re\{\widetilde{\mathbf{A}}\} \end{bmatrix}}_{\triangleq \mathbf{A}} \underbrace{\begin{bmatrix} \Re\{\widetilde{\mathbf{x}}\} \\ \Im\{\widetilde{\mathbf{x}}\} \end{bmatrix}}_{\triangleq \mathbf{x}} + \underbrace{\begin{bmatrix} \Re\{\widetilde{\mathbf{w}}\} \\ \Im\{\widetilde{\mathbf{w}}\} \end{bmatrix}}_{\triangleq \mathbf{w}}.$$

We recognize here the well-known *inverse problem*, i.e., reconstruct a sparse vector from the fewest possible number of noisy linear observations, $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$, in which $\mathbf{y} \in \mathbb{R}^N$, $\mathbf{A} \in \mathbb{C}^{M \times N}$, and $\mathbf{x} \in \mathbb{C}^N$, with $M \triangleq 2M_r K$ and $N \triangleq 2M_r M_t$.

Algorithm 1 Sum-Product GAMP for MMSE Estimation

Require: $\mathbf{A} \in \mathbb{R}^{M \times N}$; $\mathbf{y} \in \mathbb{R}^M$; $\boldsymbol{\alpha}$; σ_w^2, $p_{\mathcal{X}}(x;\boldsymbol{\alpha})$, precision tolerance ($\epsilon$), maximum # of iterations (T_{MAX})

Ensure: MMSE estimates $\{\widehat{x}_n\}_{n=1}^N$ for $\{x_n\}_{n=1}^N$

Initialization
$t \leftarrow 1$
$\forall n:\ \widehat{x}_n(t) \;=\; \int_x x\, p_{\mathcal{X}}(x;\boldsymbol{\alpha})\, \mathrm{d}x$
$\forall n:\ \mu_n^x(t) \;=\; \int_x |x - \widehat{x}_n(t)|^2 p_{\mathcal{X}}(x;\boldsymbol{\alpha})\, \mathrm{d}x$
$\forall m:\ \widehat{s}_m(t-1) \;=\; 0$
repeat
  $\forall m:\ \mu_m^p(t) \;=\; \sum_{n=1}^N |\mathbf{A}_{m,n}|^2 \mu_n^x(t)$
  $\forall m:\ \widehat{p}_m(t) \;=\; \sum_{n=1}^N \mathbf{A}_{m,n} \widehat{x}_n(t) \;-\; \mu_m^p(t) \widehat{s}_m(t-1)$
  $\forall m:\ \mu_m^z(t) \;=\; \mathsf{var}_{\mathcal{Z}_m|\mathcal{Y}}\{z_m | \mathbf{y}\,;\, \widehat{p}_m(t), \mu_m^p(t), \boldsymbol{\theta}\}$
  $\forall m:\ \widehat{z}_m(t) \;=\; \mathbb{E}_{\mathcal{Z}_m|\mathcal{Y}}\{z_m | \mathbf{y}\,;\, \widehat{p}_m(t), \mu_m^p(t), \boldsymbol{\theta}\}$
  $\forall m:\ \mu_m^s(t) \;=\; [1 - \mu_m^z(t)/\mu_m^p(t)]\,/\mu_m^p(t)$
  $\forall m:\ \widehat{s}_m(t) \;=\; [\widehat{z}_m(t) - \widehat{p}_m(t)]/\mu_m^p(t)$
  $\forall n:\ \mu_n^r(t) \;=\; \big(\sum_{n=1}^N |\mathbf{A}_{m,n}|^2 \mu_m^s(t)\big)^{-1}$
  $\forall n:\ \widehat{r}_n(t) \;=\; \widehat{x}_n(t) \;+\; \mu_n^x(t) \sum_{m=1}^M \mathbf{A}_{m,n}^* \widehat{s}_m(t)$
  $\forall n:\ \mu_n^x(t+1) \;=\; \mathsf{var}_{\mathcal{X}_n|\mathcal{Y}}\{x_n | \mathbf{y}\,;\, \widehat{r}_n(t), \mu_n^r(t), \boldsymbol{\theta}\}$
  $\forall n:\ \widehat{x}_n(t+1) \;=\; \mathbb{E}_{\mathcal{X}_n|\mathcal{Y}}\{x_n | \mathbf{y}\,;\, \widehat{r}_n(t), \mu_n^r(t), \boldsymbol{\theta}\}$
  $t \leftarrow t+1$
until $\big\|\widehat{\mathbf{x}}(t+1) \;-\; \widehat{\mathbf{x}}(t)\big\|^2 \le \epsilon \big\|\widehat{\mathbf{x}}(t)\big\|^2$ or $t > T_{\text{MAX}}$

III. Proposed mmWave Channel Estimation Algorithm

A. GAMP for Bayesian Sparse Signal Reconstruction

This paper uses the GAMP framework [8] to find the MMSE estimate of $\mathbf{x}$ while accommodating a prior on its components $\{x_n\}_{n=1}^N$. Assume that x_n's are independent and identically distributed (i.i.d.) with a common prior distribution $p_{\mathcal{X}}(x;\boldsymbol{\alpha})$ that is parameterized by an unknown parameter vector $\boldsymbol{\alpha}$. After a linear transformation $\mathbf{z} \triangleq \mathbf{A}\mathbf{x}$ and propagation through a probabilistic channel, we have:

$$p_{\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{Z}}}\left(\mathbf{y}|\mathbf{z};\sigma_w^2\right) \;=\; \textstyle\prod_{m=1}^M p_{\mathcal{Y}_m|\mathcal{Z}_m}\left(y_m|z_m;\sigma_w^2\right). \quad (7)$$

For now we assume $\boldsymbol{\theta} \triangleq [\boldsymbol{\alpha}\ \sigma_w^2]$ to be perfectly known. Later in Section IV, we describe an efficient procedure that allows learning of $\boldsymbol{\theta}$ online as well. Given the knowledge of $\mathbf{y}$, $\mathbf{A}$, and $\boldsymbol{\theta}$, GAMP runs iteratively according to the algorithmic description provided in Algorithm 1.

B. Modeling the Angular-Domain Channel Coefficients

In conventional sub-3 GHz MIMO communications, the entries of $\mathbf{H}$ are usually modeled by a Gaussian distribution due to rich scattering. In this case, the entries of $\widetilde{\mathbf{H}}$ in (3), or equivalently the components of $\mathbf{x}$, also follow the same Gaussian distribution since $\mathbf{U}_r$ and $\mathbf{U}_t$ are unitary matrices. Due to the sparse scattering environment in mmWave bands, however, a more appropriate statistical model for the angular-domain channel coefficients needs to be specified. In this context, [7] and [10] use the i.i.d. GM prior to model $\mathbf{x}$:

$$p_{\mathcal{X}}(x;\boldsymbol{\alpha}) \;=\; \textstyle\sum_{l=1}^L \omega_l \mathcal{N}(x;\eta_l,\nu_l), \quad (8)$$

with $\boldsymbol{\alpha} \triangleq [\omega_1, \ldots \omega_L, \eta_1 \ldots \eta_L, \nu_1, \ldots, \nu_l]^{\mathsf{T}}$. In this paper, we propose to use a Laplacian prior with scale parameter b:

$$p_{\mathcal{X}}(x;b) \;=\; \frac{1}{2b} e^{-\frac{|x|}{b}}. \quad (9)$$

In the next subsection, we derive the explicit expression for the posterior means and variances involved in Lines 15 and 16 of Algorithm 1 under the Laplacian prior in (9).

C. Derivation of the Posterior Means and Variances

First, since $\mathbf{A}$ is a large-size matrix, then owing to the central limit theorem, GAMP models the components of $\mathbf{z} \triangleq \mathbf{A}\mathbf{x}$, by a Gaussian distribution with mean $\widehat{p}_m$ and variance μ_m^p (in lines 7 and 8 of Algorithm 1), i.e., $z_m \sim \mathcal{N}(z_m; \widehat{p}_m, \mu_m^p)$. Therefore, it can be shown that the posterior mean and variance of $\mathcal{Z}_m$ (in lines 9 and 10 of Algorithm 1) are given by:

$$\widehat{z}_m(t) = \frac{\mu_m^p y_m + \sigma_w^2 \widehat{p}_m}{\mu_m^p + \sigma_w^2} \quad \text{and} \quad \mu_m^z(t) = \frac{\mu_m^p \sigma_w^2}{\mu_m^p + \sigma_w^2}.$$

Next, we establish the expressions for the posterior mean and variance, $\widehat{x}_n(t+1)$ and $\mu_n^x(t+1)$ (in Lines 16 and 15 of Algorithm 1), respectively. Given any prior, $p_{\mathcal{X}}(x_n;\boldsymbol{\alpha})$, on x_n's GAMP approximates their marginal posteriors by:

$$p_{\mathcal{X}_n|\boldsymbol{\mathcal{Y}}}(x_n|\mathbf{y};\widehat{r}_n,\mu_n^r,\boldsymbol{\theta}) \;=\; \frac{p_{\mathcal{X}}(x_n;\boldsymbol{\alpha})\mathcal{N}(x_n;\widehat{r}_n,\mu_n^r)}{\int_x p_{\mathcal{X}}(x;\boldsymbol{\alpha})\mathcal{N}(x;\widehat{r}_n,\mu_n^r)}. \quad (10)$$

Using the Laplacian prior, $p_{\mathcal{X}}(x_n;b)$, in (9) we first show that:

$$\begin{aligned} &p_{\mathcal{X}}(x_n;b)\mathcal{N}(x_n;\widehat{r}_n,\mu_n^r) \\ &\qquad = \tfrac{1}{2b}e^{-\alpha_n(x_n)}\mathcal{N}\big(x_n;\gamma_n(x_n),\,\mu_n^r\big), \end{aligned} \quad (11)$$

in which the functions $\alpha_n(x)$ and $\gamma_n(x)$ are given by:

$$\alpha_n(x) \;=\; \mathrm{sgn}(x)\,\widehat{r}_n/b \;-\; \mu_n^r/(2b^2), \quad (12)$$

$$\gamma_n(x) \;=\; \widehat{r}_n \;-\; \mathrm{sgn}(x)\,\mu_n^r/b. \quad (13)$$

with $\mathrm{sgn}(.)$ being the standard signum function. Plugging (11) back into (10), it follows that:

$$p_{\mathcal{X}_n|\boldsymbol{\mathcal{Y}}}(x_n|\mathbf{y};\widehat{r}_n,\mu_n^r,\boldsymbol{\theta}) \;=\; \frac{e^{-\alpha_n(x_n)}}{2b\psi_n}\mathcal{N}\big(x_n;\,\gamma_n(x_n),\,\mu_n^r\big), \quad (14)$$

in which the normalization factor ψ_n is given by:

$$\psi_n \;=\; \frac{1}{2b}\int_{-\infty}^{+\infty} e^{-\alpha_n(x)}\mathcal{N}\big(x;\,\gamma_n(x),\,\mu_n^r\big)\,\mathrm{d}x. \quad (15)$$

Using (12) and (13) in (15), it can be shown that:

$$\psi_n \;=\; \frac{1}{2b}\left[e^{-\alpha_n^-}\mathsf{Q}\left(\frac{\gamma_n^-}{\sqrt{\mu_n^r}}\right) \;+\; e^{-\alpha_n^+}\mathsf{Q}\left(-\frac{\gamma_n^+}{\sqrt{\mu_n^r}}\right)\right], \quad (16)$$

in which $\mathsf{Q}(.)$ is the tail probability of the standard normal distribution and the quantities, α_n^-, α_n^+, γ_n^-, and γ_n^+ are:

$$\alpha_n^- \;=\; -\frac{\widehat{r}_n}{b} \;-\; \frac{\mu_n^r}{2b^2} \quad \text{and} \quad \alpha_n^+ \;=\; \frac{\widehat{r}_n}{b} \;-\; \frac{\mu_n^r}{2b^2}, \quad (17)$$

$$\gamma_n^- \;=\; \widehat{r}_n \;+\; \frac{\mu_n^r}{b} \qquad \text{and} \quad \gamma_n^+ \;=\; \widehat{r}_n \;-\; \frac{\mu_n^r}{b}. \quad (18)$$

Now, owing to (14), the posterior mean, $\widehat{x}_n(t+1)$, is given by:

$$\widehat{x}_n(t+1) = \frac{1}{2b\psi_n}\int_{\mathbb{R}} x_n e^{-\alpha_n(x_n)}\mathcal{N}\big(x_n;\gamma_n(x_n),\mu_n^r\big)\,\mathrm{d}x_n.$$

Then, by using (12)-(13) along with (17)-(18) and resorting to some algebraic manipulations, we show that:

$$\begin{aligned} &\widehat{x}_n(t+1) \\ &\quad = \frac{1}{2b\psi_n}\Big[e^{-\alpha_n^+}\Phi_1\big(\gamma_n^+,\mu_n^r\big) \;-\; e^{-\alpha_n^-}\Phi_1\big(-\gamma_n^-,\mu_n^r\big)\Big], \end{aligned} \quad (19)$$

in which $\Phi_1(\gamma,\mu)$ is defind as follows:

$$\Phi_1\big(\gamma,\mu\big) \;\triangleq\; \frac{1}{\sqrt{2\pi\mu}}\int_0^{+\infty} t e^{-\frac{(t-\gamma)^2}{2\mu}}\,dt. \quad (20)$$

Moreover, it can be shown that $\Phi_1\big(\gamma,\mu\big)$ is expressed as:

$$\Phi_1\big(\gamma,\mu\big) \;=\; \gamma\mathsf{Q}\left(-\gamma/\sqrt{\mu}\right) \;+\; \frac{\mu}{\sqrt{2\pi\mu}}e^{-\frac{\gamma^2}{2\mu}}. \quad (21)$$

Plugging (21) in (19), it follows that:

$$\begin{aligned} &\widehat{x}_n(t+1) \\ &\quad = \frac{1}{2b\psi_n}\Bigg[e^{-\alpha_n^-}\gamma_n^-\mathsf{Q}\left(\frac{\gamma_n^-}{\sqrt{\mu_n^r}}\right) \;+\; e^{-\alpha_n^+}\gamma_n^+\mathsf{Q}\left(\frac{-\gamma_n^+}{\sqrt{\mu_n^r}}\right) \\ &\qquad + \frac{\mu_n^r}{\sqrt{2\pi\mu_n^r}}\left(e^{-\alpha_n^+-\frac{(\gamma_n^+)^2}{2\mu_n^r}} \;-\; e^{-\alpha_n^--\frac{(\gamma_n^-)^2}{2\mu_n^r}}\right)\Bigg]. \end{aligned} \quad (22)$$

Besides, from (17)-(18), we establish the following identity:

$$\alpha_n^+ \;+\; \frac{(\gamma_n^+)^2}{2\mu_n^r} \;=\; \alpha_n^- \;+\; \frac{(\gamma_n^-)^2}{2\mu_n^r} \;=\; \frac{\widehat{r}_n^2}{2\mu_n^r}. \quad (23)$$

This cancels the last two terms in (22) thereby leading to:

$$\begin{aligned} &\widehat{x}_n(t+1) \\ &\quad = \frac{1}{2b\psi_n}\left[e^{-\alpha_n^-}\gamma_n^-\mathsf{Q}\left(\frac{\gamma_n^-}{\sqrt{\mu_n^r}}\right) + e^{-\alpha_n^+}\gamma_n^+\mathsf{Q}\left(\frac{-\gamma_n^+}{\sqrt{\mu_n^r}}\right)\right]. \end{aligned}$$

The posterior variance in Line 15 of Algorithm 1 is given by:

$$\widehat{\mu}_n^x(t+1) \;=\; \sigma_{\mathcal{X}_n}^2(t+1) \;-\; \widehat{x}_n(t+1)^2, \quad (24)$$

where $\sigma_{\mathcal{X}_n}^2(t+1)$ is the posterior second moment of $\mathcal{X}_n$:

$$\sigma_{\mathcal{X}_n}^2(t+1) \;\triangleq\; \mathsf{E}_{\mathcal{X}_n|\boldsymbol{\mathcal{Y}}}\big\{x_n^2\big|\mathbf{y};\widehat{r}_n(t),\mu_n^r(t),\boldsymbol{\theta}\big\}. \quad (25)$$

Using the posterior distribution in (14), it follows that:

$$\sigma_{\mathcal{X}_n}^2(t+1) = \frac{1}{2b\psi_n}\int_{\mathbb{R}} x_n^2 e^{-\alpha_n(x_n)}\mathcal{N}\Big(x_n;\gamma_n(x_n),\mu_n^r\Big)\mathrm{d}x_n.$$

whose analytical expression is also established (using equivalent algebraic manipulations) as follows:

$$\begin{aligned} \sigma_{\mathcal{X}_n}^2(t+1) \;=\; \frac{1}{2b\psi_n}\Bigg[&\Big((\gamma_n^+)^2+\mu_n^r\Big)e^{-\alpha_n^+}\mathsf{Q}\left(\frac{-\gamma_n^+}{\sqrt{\mu_n^r}}\right) \\ &+\Big((\gamma_n^-)^2+\mu_n^r\Big)e^{-\alpha_n^-}\mathsf{Q}\left(\frac{\gamma_n^-}{\sqrt{\mu_n^r}}\right) \\ &-\frac{2(\mu_n^r)^2}{b\sqrt{2\pi\mu_n^r}}e^{-\frac{\widehat{r}_n^2}{2\mu_n^r}}\Bigg], \end{aligned}$$

Plugging this result back in (24) yields the required $(t+1)th$ update for the nth posterior variance, $\widehat{\mu}_n^x(t+1)$.

IV. Learning the Scale Paramater of the Laplacian Prior and the Noise Variance

We now address the issue of how to find the maximum likelihood estimate (MLE) of $\boldsymbol{\theta} \triangleq [b\ \sigma_w]$ given by:

$$\widehat{\boldsymbol{\theta}} \;=\; \underset{\boldsymbol{\theta}}{\mathrm{argmax}}\ \ln p_{\boldsymbol{\mathcal{Y}}}(\mathbf{y};\boldsymbol{\theta}). \quad (26)$$

The direct maximization of the log-likelihood function, $\ln p_{\boldsymbol{\mathcal{Y}}}(\mathbf{y};\boldsymbol{\theta})$, is analytically intractable. This paper proposes to

use the EM concept [13] in order to find the MLEs approximately. Observe that GAMP already returns the adequate posterior probabilities that are required by the EM algorithm in order to update its estimates. Hence, the implementation of the EM algorithm incurs little additional cost. We define the so-called *incomplete* and *complete* data sets (in EM terminology), as $\mathbf{y}$ and $\mathbf{v} \triangleq [\mathbf{y}, \mathbf{x}]^T$, respectively. Then, starting with some initial guess, $\widehat{\boldsymbol{\theta}}^{(0)}$, the EM algorithm updates the MLEs iteratively by alternating between the following two main steps:

• **Expectation step (E-STEP)**: Find the average log-likelihood of the complete data:

$$Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_k) \;=\; \int_{\mathbf{v}} \ln p_{\boldsymbol{\mathcal{V}}}(\mathbf{v}; \boldsymbol{\theta}) p_{\boldsymbol{\mathcal{V}}|\boldsymbol{\mathcal{Y}}}\big(\mathbf{v}|\mathbf{y}; \widehat{\boldsymbol{\theta}}_k\big)\, \mathrm{d}\mathbf{v}. \tag{27}$$

• **Maximization step (M-STEP)**: Maximize the average log-likelihood of the complete data:

$$\widehat{\boldsymbol{\theta}}_{k+1} \;=\; \underset{\boldsymbol{\theta}}{\mathrm{argmax}}\; Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_k). \tag{28}$$

Now, using the fact that $\mathbf{v} \triangleq [\mathbf{y}, \mathbf{x}]^T$, it follows that:

$$p_{\boldsymbol{\mathcal{V}}}(\mathbf{v}; \boldsymbol{\theta}) \;=\; p_{\boldsymbol{\mathcal{Y}},\boldsymbol{\mathcal{X}}}(\mathbf{y}, \mathbf{x}; \boldsymbol{\theta}) \;=\; p_{\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{X}}}(\mathbf{y}|\mathbf{x}; \sigma_w^2) p_{\boldsymbol{\mathcal{X}}}(\mathbf{x}; b),$$

which is used back in (27) to yield:

$$\begin{aligned} Q(\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}}_k) \;=\; & \int_{\mathbf{x}} \ln p_{\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{X}}}(\mathbf{y}|\mathbf{x}; \sigma_w^2) p_{\boldsymbol{\mathcal{X}}|\boldsymbol{\mathcal{Y}}}\big(\mathbf{x}|\mathbf{y}; \widehat{\boldsymbol{\theta}}_k\big)\, \mathrm{d}\mathbf{x} \\ & + \int_{\mathbf{x}} \ln p_{\boldsymbol{\mathcal{X}}}(\mathbf{x}; b) p_{\boldsymbol{\mathcal{X}}|\boldsymbol{\mathcal{Y}}}\big(\mathbf{x}|\mathbf{y}; \widehat{\boldsymbol{\theta}}_k\big)\, \mathrm{d}\mathbf{x}. \end{aligned} \tag{29}$$

Consequently, the k^{tk} EM updates, $\widehat{\boldsymbol{\theta}}_{k+1} = [\widehat{\sigma}^2_{w,k+1}, \widehat{b}_{k+1}]$, given in (28) are obtained as follows:

$$\widehat{b}_{k+1} \;=\; \underset{b>0}{\mathrm{argmax}}\, \mathbb{E}_{\boldsymbol{\mathcal{X}}|\boldsymbol{\mathcal{Y}}}\big\{ \ln p_{\boldsymbol{\mathcal{X}}}(\mathbf{x}; b) \big| \mathbf{y}, \widehat{\boldsymbol{\theta}}_k \big\}. \tag{30}$$

$$\widehat{\sigma}^2_{w,k+1} \;=\; \underset{\sigma_w>0}{\mathrm{arg\,max}}\, \mathbb{E}_{\boldsymbol{\mathcal{X}}|\boldsymbol{\mathcal{Y}}}\big\{ \ln p_{\boldsymbol{\mathcal{Y}}|\boldsymbol{\mathcal{X}}}(\mathbf{y}|\mathbf{x}; \sigma_w^2) \big| \mathbf{y}, \widehat{\boldsymbol{\theta}}_k \big\}, \tag{31}$$

Furthermore, since the components of $\mathbf{x}$ are assumed to be i.i.d. with $x_n \sim p_{\mathcal{X}}(x_n; b)$, it follows from (30) that:

$$\widehat{b}_{k+1} \;=\; \underset{b>0}{\mathrm{argmax}} \textstyle\sum_{n=1}^{N} \mathbb{E}_{\boldsymbol{\mathcal{X}}|\boldsymbol{\mathcal{Y}}}\big\{ \ln p_{\mathcal{X}}(x_n; b) \big| \mathbf{y}; \widehat{\boldsymbol{\theta}}_k \big\}, \tag{32}$$

The expectation in (32) is taken with respect to the posterior density $p_{\boldsymbol{\mathcal{X}}|\boldsymbol{\mathcal{Y}}}\big(\mathbf{x}|\mathbf{y}; \widehat{\boldsymbol{\theta}}_k\big)$ which we further approximate by:

$$p_{\boldsymbol{\mathcal{X}}|\boldsymbol{\mathcal{Y}}}\big(\mathbf{x}|\mathbf{y}; \widehat{\boldsymbol{\theta}}_k\big) \;=\; \prod_{n=1}^{N} p_{\mathcal{X}_n|\boldsymbol{\mathcal{Y}}}\big(x_n|\mathbf{y}; \widehat{\boldsymbol{\theta}}_k\big). \tag{33}$$

Using (33) in (32), it follows that:

$$\widehat{b}_{k+1} \;=\; \underset{b>0}{\mathrm{argmax}} \underbrace{\textstyle\sum_{n=1}^{N} \mathbb{E}_{\mathcal{X}_n|\boldsymbol{\mathcal{Y}}}\Big\{ \ln p_{\mathcal{X}}(x_n; b) \Big| \mathbf{y}; \widehat{\boldsymbol{\theta}}_k \Big\}}_{\triangleq\, g(b|\mathbf{y}, \widehat{\boldsymbol{\theta}}_k)}. \tag{34}$$

Clearly, $\widehat{b}_{k+1}$ is the value of b that zeros:

$$\frac{\mathrm{d}}{\mathrm{d}b} g(b|\mathbf{y}, \widehat{\boldsymbol{\theta}}_k) = \sum_{n=1}^{N} \mathbb{E}_{\mathcal{X}_n|\boldsymbol{\mathcal{Y}}} \left\{ \frac{1}{p_{\mathcal{X}}(x_n; b)} \frac{\mathrm{d}}{\mathrm{d}b} p_{\mathcal{X}}(x_n; b) \Big| \mathbf{y}; \widehat{\boldsymbol{\theta}}_k \right\}. \tag{35}$$

By using the Laplacian distribution given in (9) into (35), setting the result to zero, then solving for b, we obtain:

$$\widehat{b}_{k+1} \;=\; \tfrac{1}{N} \textstyle\sum_{n=1}^{N} \mathbb{E}_{\mathcal{X}_n|\boldsymbol{\mathcal{Y}}}\big\{ |x_n| \,\big|\, \mathbf{y}; \widehat{\boldsymbol{\theta}}_k \big\} \tag{36}$$

The expectations in (36) are computed with respect to the posterior distributions $p_{\mathcal{X}_n|\boldsymbol{\mathcal{Y}}}\big(x_n|\mathbf{y}; \widehat{\boldsymbol{\theta}}_k\big)$ which are readily obtained from the auxiliary outputs of GAMP according to (14) in which the true parameter vector $\boldsymbol{\theta}$ is replaced by its previous EM update $\widehat{\boldsymbol{\theta}}_k$. By doing so and resorting to somewhat tedious algebraic manipulations, we establish the EM update for the scale parameter, b, as follows:

$$\widehat{b}_{k+1} \;=\; \frac{2\mu_n^r}{\sqrt{2\pi\mu_n^r}} e^{-\frac{\widehat{r}_n^2}{2\mu_n^r}} \left(\frac{1}{N} \sum_{n=1}^{N} \frac{1}{\xi_{n,k}} \right) + \frac{1}{N} \sum_{n=1}^{N} \frac{\xi'_{n,k}}{\xi_{n,k}},$$

in which $\xi_{n,k}$ and $\xi'_{n,k}$ are explicitly given by:

$$\begin{aligned} \xi_{n,k} \;=\; & e^{-\alpha^+_{n,k}} \mathsf{Q}\left(\frac{-\gamma^+_{n,k}}{\sqrt{\mu_n^r}} \right) \;+\; e^{-\alpha^-_{n,k}} \mathsf{Q}\left(\frac{\gamma^-_{n,k}}{\sqrt{\mu_n^r}} \right), \\ \xi'_{n,k} \;=\; & \gamma^+_{n,k} e^{-\alpha^+_{n,k}} \mathsf{Q}\left(\frac{-\gamma^+_{n,k}}{\sqrt{\mu_n^r}} \right) \;-\; \gamma^-_{n,k} e^{-\alpha^-_{n,k}} \mathsf{Q}\left(\frac{\gamma^-_{n,k}}{\sqrt{\mu_n^r}} \right). \end{aligned}$$

Using equivalent manipulations, it can also be shown that the EM update for the noise variance is given by:

$$\widehat{\sigma}_{w,k+1} \;=\; \tfrac{1}{M} \textstyle\sum_{m=1}^{M} \big(|y_m - \widehat{z}_m|^2 \;+\; \mu_m^z \big), \tag{37}$$

Finally, it is worth mentioning that the results disclosed in this paper can be readily applied to the mmWave frequency-selective channel estimation problem since the latter can also be recast as a CS recovery problem under linear mixing as shown in [7]. In this case, the beam-domain coefficients of each channel tap can be modelled by a Laplacian prior and the GAMP updates for their posterior means and variances are computed using the same expressions established in this paper.

V. Simulation Results

This section assesses the performance of the proposed massive MIMO mmWave channel estimator using Monte-Carlo simulations. The normalized mean-square error (NMSE), $\mathrm{NMSE} \triangleq \mathbb{E}\big\{ \|\mathbf{x} - \mathbf{x}\|^2 / \|\mathbf{x}\|^2 \big\}$, is used as a performance measure. The received powers are normalized to one such that the SNR of the system is simply given by $\mathrm{SNR} = 1/(2\sigma_w^2)$.

As baselines, we consider the least-squares (LS) estimator $\widehat{\mathbf{x}}_{\mathrm{LS}} = (\mathbf{A}^{\mathsf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathsf{T}}\mathbf{y}$, as well as, the GAMP-based estimator under the GM prior investigated recently in [7] and referred to here as GAMP-GM. As a representative example, we consider a system with $M_t = M_r = 64$ antennas in an environment with $P = 4$ path clusters each of which consisting of $Q_p = 10$ sub-paths with angular spread of 3.5 degrees. The results reported here are obtained using 100 Monte-Carlo trials. Further, we set the maximum number of iterations $T_{\mathrm{max}} = 50$ and the precision tolerance $\epsilon = 10^{-6}$ for GAMP (cf. Algorithm 1).

Fig. 1 depicts the NMSE performance of the different methods against SNR for two different sizes of the training sequence, namely $K = 64$ and $K = 128$. First, Fig. 1(a) indicates that the performance of the LS estimator is very poor since $K = 64$ corresponds to the case where the number of unknowns $N = M_t M_r$ is equal to the number of observations $M = KM_r$. In fact, it is widely known that LS estimation requires more observations than unknowns. This is confirmed by Fig 1(b) in which we double the number of observations. In the latter case, the advantage of the two Bayesian approaches over the non-Bayesian LS estimator in low-to-moderate SNRs is due to their ability to exploit the prior information about the

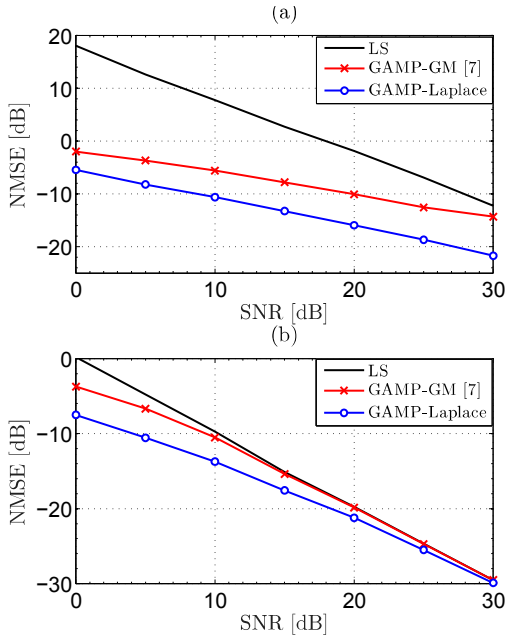

Fig. 1: Estimation NMSE for GAMP-Laplace, GAMP-GM, and ML versus SNR: (a) $K = 64$ and (b) $K = 128$.

unknown channel instead of simply assuming it to be unknown but deterministic. It is also seen that in both cases GAMP-Laplace offers remarkable performance gains over GAMP-GM, e.g., at target NMSE $= -10$ dB, the gains in terms of SNR are as high as 10 dB and 5 dB for $K = 64$ and $K = 128$, respectively. We also verified via computer simulations that these NMSE improvements lead to substantial improvements in terms of achievable rates but the results were not included here due to space limitations.

Fig. 2 plots the average number of iterations required by GAMP (until the condition in Line 18 of Algorithm 1 is satisfied) under both the GM and Laplace priors. There, it is seen that GAMP-Laplace converges much faster than GAMP-GM. For instance, at SNR $= 10$ dB, GAMP-Laplace converges in almost 25 iterations when $K = 64$ as opposed to 35 iterations for GAMP-GM thereby leading to tremendous computational savings in practice. Recall here that GAMP performs four matrix/vector multiplications at each iteration (cf. Algorithm 1 which is actually a scalarized version of GAMP). Hence, GAMP-Laplace saves on average 40 matrix/vector multiplications over GAMP-GM. This is to be added to the computational savings stemming from the fact that GAMP-Laplace needs to learn only one parameter, b, under each iteration, as opposed to $3L$ different parameters for GAMP-GM where L is the order of the Gaussian mixture.

VI. Conclusion

This paper proposes a new channel estimator for massive MIMO mmWave systems that leverages the inherent sparsity of the channel in the angular domain. The proposed estimator belongs to the family of Bayesian estimators and builds upon the GAMP algorithm. As compared to prior works, our key observation is that the angular-domain channel coefficients should be modeled by a Laplacian prior distribution. This paper also proposes an EM-based approach to systematically learn the unknown scale parameter of the underlying Laplace distribution

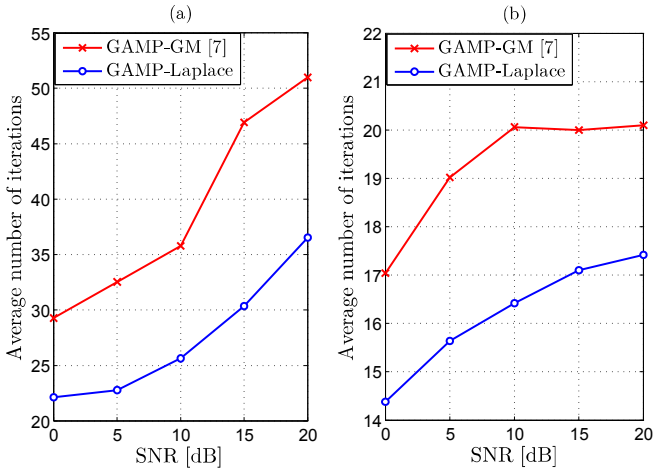

Fig. 2: Average number of iterations until convergence for GAMP-GM and GAMP-Laplace: (a) $K = 64$ and (b) $K = 128$.

and the noise variance. As compared to the Gaussian mixture model advocated in the recent literature, it is seen that the Laplacian prior leads to remarkable performance improvements in terms of channel estimation accuracy on the top of speeding up the convergence of GAMP.

References

[1] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, June 2014.

[2] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74–80, Feb. 2014.

[3] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, "Millimeter-wave massive MIMO: The next wireless revolution?" *IEEE Commun. Mag.*, vol. 52, no. 9, pp. 56–62, Sept. 2014.

[4] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[5] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wide-band millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sept. 2015.

[6] T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband millimeter-wave propagation measurements and models using adaptive-beam antennas for outdoor urban cellular communications," *IEEE Trans. Antennas Propag.*, vol. 61, no. 4, pp. 1850–1859, Apr. 2013.

[7] J. Mo, P. Schniter, and R. W. Heath, "Channel estimation in broadband millimeter wave MIMO systems with few-bit ADCs," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1141–1154, Mar. 2018.

[8] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, St. Petersburg, Russia, July 2011, pp. 2168–2172.

[9] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, Sept. 2009.

[10] C. K. Wen, S. Jin, K. K. Wong, J. C. Chen, and P. Ting, "Channel estimation for massive MIMO using Gaussian-mixture bayesian learning," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1356–1368, Mar. 2015.

[11] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000.

[12] A. M. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[13] J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.