SEMI-BLIND CHANNEL ESTIMATION IN MASSIVE MIMO SYSTEMS WITH DIFFERENT PRIORS ON DATA SYMBOLS

Elina Nayebi and Bhaskar D. Rao

Department of Electrical and Computer Engineering
University of California, San-Diego

ABSTRACT

This paper investigates semi-blind channel estimation in massive multiple-input multiple-output (MIMO) systems using different priors on data symbols. We derive two tractable expectation-maximization (EM) based channel estimation algorithms; one based on a Gaussian prior and the other one based on a Gaussian mixture model (GMM) for the unknown data symbols. The numerical results show that the semi-blind estimation schemes provide better channel estimates compared with the estimation based on training sequences only. The EM algorithm with a Gaussian prior provides superior channel estimates compared to the EM algorithm with a GMM prior in low signal-to-noise ratio (SNR) regime. However, the latter one outperforms the EM algorithm with Gaussian prior as the SNR or as the number antennas at the base station (BS) increases. Furthermore, the performance of the semi-blind estimators become closer to the genie-aided maximum likelihood estimator based on known data symbols as the number of antennas at the BS increases.

***Index Terms*—** Massive MIMO, channel Estimation, EM Algorithm, prior distribution

1. INTRODUCTION

Channel state information is an important factor in achieving the expected high capacity gains in multiple-input multiple-output (MIMO) systems, which in practical systems is determined by the accuracy of the channel estimates. Using training pilot sequences is a simple method to estimate the channel coefficients [1]. However, large number of training symbols are required to obtain a reliable channel estimate with this method, which reduces the achievable throughput of the system. One can improve the channel estimation quality by using the information in the unknown data symbols instead of only using the pilot sequences [2–6]. This approach provides more accurate channel estimates or results in utilizing smaller number of training pilots to estimate the channel coefficients with the same accuracy.

In time-division duplex (TDD) systems, where uplink and downlink physical channels are assumed to be reciprocal [7], better channel estimation not only leads to better uplink detection but it also helps the base station (BS) to form more accurate downlink precoders. Massive MIMO systems, in which the emphasis in transmission protocol is mostly on TDD rather than frequency-division duplex (FDD) (see [8] and [9]), the benefit from semi-blind channel estimation accrues in both uplink and downlink transmissions. This makes semi-blind estimation more attractive for the next generation wireless systems and motivates our re-examination of semi-blind channel estimation with an eye towards massive MIMO systems.

Semi-blind channel estimation has been investigated in several papers, e.g., [2–5] and references therein. In [2], the authors study the conditions under which the channel and the data signals are blindly and semi-blindly identifiable and obtain blind and semi-blind channel estimates based on an expectation-maximization (EM) algorithm in the frequency domain and utilize a discrete random variable model for the unknown data. In [3], two iterative channel estimators based on the EM algorithm are proposed. In [5], a semi-blind estimation technique for MIMO systems is introduced, which uses an iterative two-level optimization loop to jointly estimate channel coefficients and data symbols.

In our previous work [10], in contrast to other works, we considered a Gaussian distribution for the unknown data symbols in the EM algorithm, which enabled us to derive a closed form solution for the E-step of the EM algorithm. We further derived deterministic and stochastic Cramer-Rao bounds (CRBs) for semi-blind channel estimation and studied their behavior in massive MIMO systems with unlimited number of antennas at the BS. In this paper, in order to improve the performance of estimation for the case when data symbols are drawn from a discrete constellation such as quadrature phase-shift keying data (QPSK), we first consider a heuristic algorithm by demapping the conditional mean of the data symbols to the nearest constellation point in the EM algorithm. Numerical results illustrate that in the high signal-to-noise ratio (SNR) regime, this heuristic algorithm considerably outperforms the EM algorithm with Gaussian prior. Motivated by this observation, to provide support for the procedure, we

This research was supported by the National Science Foundation Grant No. CCF-1617365, the Center for Wireless Communications and a grant from the King Abdulaziz City for Science and Technology.

pursue deriving an analytically rigorous EM algorithm assuming a Gaussian mixture model (GMM) for data symbols which achieves a performance similar to the heuristic algorithm. Numerical results indicate that the semi-blind estimation schemes provide better channel estimates compared with channel estimation based on training sequences only. Furthermore, as the number of antennas in the massive MIMO systems increases, the mean squared error (MSE) of semi-blind schemes becomes closer to the genie-aided estimation when all data symbols are assumed to be known at the BS.

The paper is organized as follows. In section 2, we describe the system model and two ML estimators. In section 3, three semi-blind channel estimation schemes based on the EM algorithm are derived. We present numerical results in section 4 and conclude the paper in section 5.

Throughout the paper, we use superscript ^{H} to denote conjugate transpose, uppercase symbols to denote matrices, and bold symbols to denote vectors. $\mathbb{E}(.)$ and $\mathrm{Cov}(.)$ indicate the expectation and covariance matrix operators respectively.

2. SYSTEM MODEL

We use the same system model as in [10]. However, we recall it here for the sake of completeness. We consider a single cell with a BS equipped with M antennas and randomly located K single antenna users, where $M \geq K$. We study uplink transmission in a communication system with TDD protocol. However, similar estimation techniques can also be applied to a system with FDD protocol. We use a flat fading channel model for each orthogonal frequency-division multiplexing (OFDM) subcarrier. The OFDM subcarrier index is omitted for simplicity. The channel matrix between the BS and users is given by

$$G = HB^{1/2}, \tag{1}$$

where $H \in \mathbb{C}^{M \times K}$ is a matrix representing small scale fading and $B \in \mathbb{R}^{K \times K}$ is a diagonal matrix with $\beta_1, \cdots, \beta_K$ on its diagonal, where β_k is the large scale fading coefficient between BS and user k that accounts for the path loss and shadow fading. We assume columns of H are independent from B and are i.i.d circularly-symmetric complex normal vectors $\mathbf{h_k} \sim \mathcal{CN}(0, I_M)$ that stay constant during a block of N symbols and change to independent values at the next coherence block.

In uplink, transmitting users send L known pilot sequences followed by $(N-L)$ unknown data symbols. The uplink signal received by BS at time n is given by

$$\mathbf{y}[n] = G\mathbf{s}[n] + \mathbf{v}[n], \quad n = 0, \cdots, N-1, \tag{2}$$

where $\mathbf{s}[n] \in \mathbb{C}^{K \times 1}$ for $n = 0, \cdots, L-1$ are known pilot sequences and $\mathbf{s}[n] \in \mathbb{C}^{K \times 1}$ for $n = L, \cdots, N-1$ are the unknown data symbols with unit power $\mathbb{E}\left(\mathbf{s}[n]\mathbf{s}[n]^H\right) = I_K$ and $\mathbf{v}[n] \sim \mathcal{CN}(0, \sigma_v^2 I_M)$ is additive Gaussian noise. Let $S_p = [\mathbf{s}[0], \cdots, \mathbf{s}[L-1]]$ and $S_d = [\mathbf{s}[L], \cdots, \mathbf{s}[N-1]]$ denote, respectively, the known pilot sequences and data symbols in a channel coherence time. Similarly, let Y_p and Y_d represent the row stacked received training output and received data signals respectively. The complete transmit and received symbols are given by $S = [S_p\ S_d]$ and $Y = [Y_p\ Y_d]$ respectively.

Based on [4, Lemma 1], as the MSE of channel matrix becomes closer to the trace of CRB, then the error covariance matrix approaches the CRB. Thus, we will use *MSE*, i.e., trace of the error covariance matrix, in the numerical results to illustrate the accuracy of the channel estimates.

2.1. ML Estimators

2.1.1. Training Pilot Sequences

A simple channel estimation method is with the aid of known training sequences. The ML estimate of the channel matrix G based on the pilot sequences (S_p) is given by

$$\hat{G}_{\mathrm{ML}}^{\mathrm{tr}} = \left(Y_p S_p^H\right)\left(S_p S_p^H\right)^{-1}. \tag{3}$$

The training sequences that minimize the MSE subject to the total transmit power are orthogonal sequences, i.e., $S_p S_p^H = L I_K$, and the corresponding MSE is equal to $\mathbb{E}(\|G - \hat{G}_{\mathrm{ML}}^{\mathrm{tr}}\|_F^2) = {}^{MK\sigma_v^2}/_L$ [11].

2.1.2. Genie-Aided

The genie-aided channel estimation when all data symbols are known at the BS provides an upper bound on the performance of semi-blind channel estimation and is given by

$$\hat{G}_{\mathrm{ML}}^{\mathrm{full}} = \left(Y S^H\right)\left(S S^H\right)^{-1}. \tag{4}$$

This is an estimator that the semi-blind procedure aspires to imitate.

3. SEMI-BLIND CHANNEL ESTIMATION

The ML estimate of channel coefficients based on both received training and unknown data signals is given by

$$\hat{G}_{\mathrm{ML}} = \underset{G}{\mathrm{argmax}}\ \log p\left(Y|G\right), \tag{5}$$

for which obtaining a closed form solution is known to be hard [12]. Iterative algorithms have been proposed to solve the problem, e.g., see [4], [5].

3.1. EM Algorithm with Gaussian Prior

The problem in (5) can be solved using the EM algorithm, by iterating between an expectation evaluation (E-step) and a maximization (M-step) procedure [13].

The channel estimate in the EM algorithm ($\hat{G}_{\ell+1}$) is updated based on the old estimate ($\hat{G}_\ell$) in the following manner:

$$\hat{G}_{\ell+1} = \underset{G}{\operatorname{argmax}}\ \mathbb{E}_{p(S_d|Y,\hat{G}_\ell)} \left(\log p\left(Y, S_d|G\right)\right), \quad (6)$$

where (Y, S_d) is the complete data. By carrying out the maximization (M-step), one can show that the channel estimate at the $(\ell+1)$th iteration is given by [2]

$$\hat{G}_{\ell+1} = \left(Y_p S_p^H + \sum_{n=L}^{N-1} \mathbf{y}[n](\boldsymbol{\mu}_n^\ell)^H \right) \times \left(S_p S_p^H + \sum_{n=L}^{N-1} \left(\boldsymbol{\mu}_n^\ell (\boldsymbol{\mu}_n^\ell)^H + \Sigma^\ell \right) \right)^{-1}. \quad \text{(7a)}$$

where $\boldsymbol{\mu}_n^\ell \triangleq \mathbb{E}(\mathbf{s}[n]|\hat{G}_\ell, Y)$, and $\Sigma_n^\ell \triangleq \operatorname{Cov}(\mathbf{s}[n]|\hat{G}_\ell, Y)$. The details of the expectation computation needed to complete the E-step of the algorithm has not been mentioned explicitly in [2]. Using a discrete random variable model such as QPSK modulation for data symbols leads to an excessively complex E-step which grows exponentially with K. Thus, for tractability of the problem, we assume that the data symbols are Gaussian, i.e., $\mathbf{s}[n] \sim \mathcal{CN}\left(0, I_K\right),\ n = L, \cdots, N-1$. Given G, S_d and Y are jointly Gaussian. Thus, $\boldsymbol{\mu}_n^\ell$ and Σ_n^ℓ in (7a), can be computed from the conditional density of circularly symmetric Gaussian random vectors [14]. The E-step based on the estimates at the ℓth iteration is given by

$$\begin{aligned} \boldsymbol{\mu}_n^\ell &= \left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 I_K \right)^{-1} \hat{G}_\ell^H \mathbf{y}[n], \\ \Sigma^\ell &= \sigma_v^2 \left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 I_K \right)^{-1}. \end{aligned} \quad \text{(7b)}$$

Derivation of the E- and M-steps and the computational complexity of (7) are presented in [10].

3.2. Heuristic Semi-Blind Algorithm

Since data symbols $\mathbf{s}[n],\ n = L, \cdots, N-1$, are drawn from a discrete constellation, we now modify the EM algorithm to improve the estimation performance. A heuristic approach is to assign the conditional mean of data symbols $\mathbb{E}(\mathbf{s}[n]|Y, \hat{G}_\ell),\ n = L, \cdots, N-1$, to the closest constellation point, which results in the following E-step:

$$\begin{aligned} \boldsymbol{\mu}_n^\ell &= F\left(\left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 I_K \right)^{-1} \hat{G}_\ell^H \mathbf{y}[n] \right), \\ \Sigma^\ell &= \sigma_v^2 \left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 I_K \right)^{-1}, \end{aligned} \quad (8)$$

where $F(.)$ is the element-wise constellation demmaping function. Note that the M-step remains the same as (7a).

3.3. EM Algorithm with GMM Prior

Numerical results in Section 4 suggest that this modification of the EM algorithm improves the estimation performance for discrete constellations. To provide analytical support for this heuristic approach, we derive a mathematically rigorous algorithm in the following by assuming a GMM distribution for data symbols which has a similar flavor. This algorithm is also based on the EM algorithm and hence its convergence to a local maximum is assured. Suppose data symbols have GMM distribution, i.e.,

$$\mathbf{s}[n] \sim \mathcal{CN}\left(\mathbf{c}_n, \sigma_s^2 I_K\right), \quad n = L, \cdots, N-1, \quad (9)$$

where $\mathbf{c}[n]$ is the transmitted constellation vector at time n that will be treated as the unknown parameter in the EM algorithm. The hyperparameter σ_s^2 in (9) is the variance of each data symbol around the corresponding constellation point. As σ_s^2 becomes smaller, the GMM distribution in (9) becomes closer to the actual discrete distribution of the data symbols.

Let $\Theta = [G, \mathbf{c}[L], \cdots, \mathbf{c}[N-1]]$ denote the unknown variables in the EM algorithm. Given Θ, S_d and Y are jointly Gaussian. Similar to Section 3.1, $\boldsymbol{\mu}_n^\ell = \mathbb{E}(\mathbf{s}[n]|\hat{\Theta}_\ell, Y)$ and $\Sigma_n^\ell \triangleq \operatorname{Cov}(\mathbf{s}[n]|\hat{\Theta}_\ell, Y)$ in the E-step can be computed from the conditional density of circularly symmetric Gaussian random vectors [14] as follows

$$\begin{aligned} \boldsymbol{\mu}_n^\ell &= \left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 \left(\hat{\mathbf{c}}_n^\ell (\hat{\mathbf{c}}_n^\ell)^H + \sigma_s^2 I_K \right)^{-1} \right)^{-1} \hat{G}_\ell^H \mathbf{y}[n], \\ \Sigma_n^\ell &= \sigma_v^2 \left(\hat{G}_\ell^H \hat{G}_\ell + \sigma_v^2 \left(\hat{\mathbf{c}}_n^\ell (\hat{\mathbf{c}}_n^\ell)^H + \sigma_s^2 I_K \right)^{-1} \right)^{-1}. \end{aligned} \quad \text{(10a)}$$

Maximizing the log likelihood function yields in the following M-step:

$$\begin{aligned} \hat{G}_{\ell+1} &= \left(Y_p S_p^H + \sum_{n=L}^{N-1} \mathbf{y}[n](\boldsymbol{\mu}_n^\ell)^H \right) \times \left(S_p S_p^H + \sum_{n=L}^{N-1} \left(\boldsymbol{\mu}_n^\ell (\boldsymbol{\mu}_n^\ell)^H + \Sigma_n^\ell \right) \right)^{-1}, \\ \hat{\mathbf{c}}_n^{\ell+1} &= F\left(\boldsymbol{\mu}_n^\ell\right), \end{aligned} \quad \text{(10b)}$$

where $F(.)$ is the element-wise constellation demmaping function. Due to lack of space, the detailed derivation of the EM algorithm is omitted.

4. NUMERICAL RESULTS

We consider a BS located at the center of a single cell of radius 500m and uniformly distributed users. We use the same model as in [10] for the large scale fading coefficients. Signal-to-noise ratio is defined as $\text{SNR} = \frac{\mathbb{E}(\beta_k)}{\sigma_v^2}$. A QPSK constellation is used for pilot sequences (S_p) and data symbols (S_d). Pilot sequences are chosen to be orthogonal, i.e., $S_p S_p^H = L I_K$. We initialize all semi-blind algorithms using the ML training-based estimate in (3). In the EM algorithm with GMM prior, we set $\sigma_s = 0.001$.

Experiment 1: This experiment shows that more accurate channel estimates can be obtained using the semi-blind estimation schemes. Figure 1 illustrates the scaled MSE, i.e., $\mathbb{E}(\|G - \hat{G}\|_F^2)/\mathbb{E}(\beta_k)$, of the channel estimates versus SNR for the ML estimators in Subsection 2.1 and the semi-blind algorithms given in (7), (8), and (10) for $M = 8$, $K = 4$, $L = 16$, and $N = 512$. One can observe that all semi-blind algorithms provide better channel estimates compared with the ML training-based estimation. In low SNRs, EM algorithm with Gaussian prior outperforms the other semi-blind schemes. However, as SNR increases, the heuristic semi-blind estimation and EM algorithm with GMM prior provide better channel estimates and become closer to the genie-aided ML estimator. To explain this behavior, we point out that the constellation demapping in the heuristic algorithm and the EM algorithm with GMM prior adds to the estimation error when the estimates of $\boldsymbol{\mu}_n$ are uncertain in low SNRs. This phenomenon works in favor of these two algorithms in the high SNR regime by mapping $\boldsymbol{\mu}_n$ to its true value.

In Figure 2, we plot the scaled MSE of the semi-blind algorithms versus number of iterations for $M = 64$, $K = 8$, $L = 16$, $N = 512$, and SNR$=$ 20dB. One can see that all semi-blind algorithms converge after a few iterations and that the heuristic semi-blind algorithm and the EM algorithm with GMM prior show faster convergence compared to the EM algorithm with Gaussian prior.

Experiment 2: In this experiment, the effect of increasing the number of antennas at the BS is studied for $K = 8$, $L = 16$, $N = 512$, and SNR $= 25$dB. Figure 3 shows the scaled MSE of the ML estimators described in Subsection 2.1 and the semi-blind algorithms versus number of antennas (M). As the number of antennas increases, the performance of the semi-blind algorithms become closer to that of the genie-aided ML estimator. This property makes semi-blind channel estimation an attractive approach to alleviate the pilot contamination problem in massive MIMO systems, which originates from non-orthogonal pilot sequences or the reuse of pilot sequences for neighboring cells, which is our future topic of study.

5. CONCLUSION

We developed semi-blind channel estimation algorithms using Gaussian and GMM priors for data symbols and compared their performances with known ML estimators. Numerical results indicate that performances of the semi-blind estimation schemes become closer to the genie-aided ML estimator as the number of BS antennas increases. The EM algorithm with Gaussian prior has superior performance compared with the EM algorithm with GMM prior in the low SNR regime. However, as the SNR or as the number of antennas at BS increases, the performance of the EM algorithm with GMM prior improves compared to the EM algorithm with Gaussian prior and becomes closer to the genie-aided ML estimator.

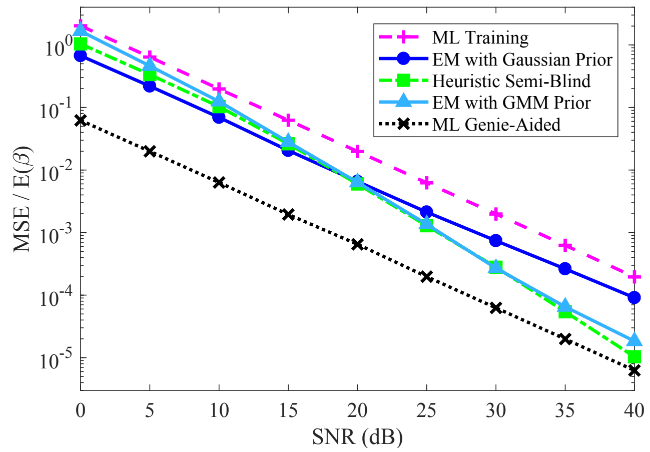

Fig. 1. Scaled MSE versus SNR with $M = 8$, $K = 4$, $L = 16$, and $N = 512$.

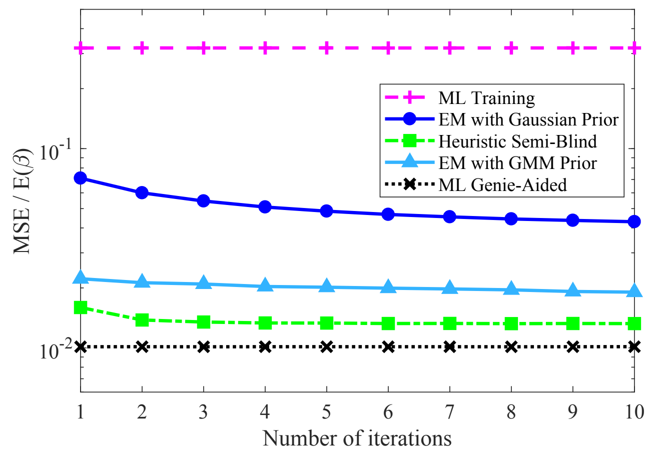

Fig. 2. Scaled MSE versus number of iterations with $M = 64$, $K = 8$, $L = 16$, $N = 512$, and SNR $= 20$dB.

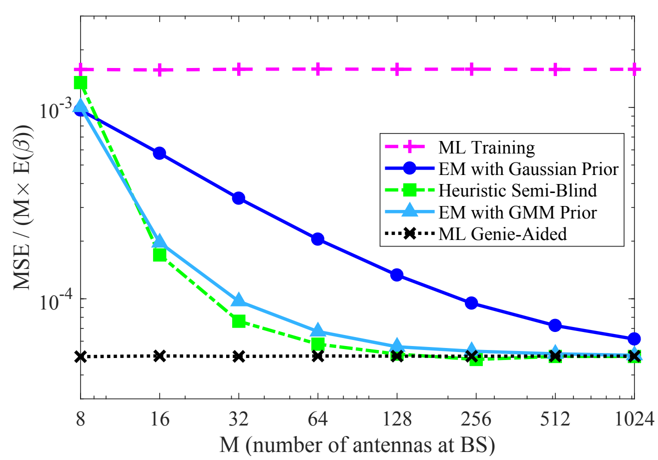

Fig. 3. $\frac{1}{M}$MSE versus M with $K = 8$, $L = 16$, $N = 512$, and SNR $= 25$dB.

6. REFERENCES

[1] T. L. Marzetta, “How much training is required for multiuser MIMO?,” in *Proc. Asilomar Conference on Signals, Systems and Computer*, Oct. 2006, pp. 359–363.

[2] C. H. Aldana, E. de Carvalho, and J. M. Cioffi, “Channel estimation for multicarrier multiple input single output systems using the EM algorithm,” *IEEE Trans. Signal Process.*, vol. 51, no. 12, pp. 3280–3292, Dec. 2003.

[3] X. Wautelet, C. Herzet, A. Dejonghe, J. Louveaux, and L. Vandendorpe, “Comparison of EM-based algorithms for MIMO channel estimation,” *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 216–226, 2007.

[4] A. K. Jagannatham and B. D. Rao, “Whitening-rotation-based semi-blind MIMO channel estimation,” *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 861–869, Mar. 2006.

[5] M. Abuthinien, S. Chen, and L. Hanzo, “Semi-blind joint maximum likelihood channel estimation and data detection for MIMO systems,” *IEEE Signal Process. Lett.*, vol. 15, pp. 202–205, Nov. 2008.

[6] J. Choi, *Adaptive and iterative signal processing in communications*, Cambridge, U.K.: Cambridge University Press, 2006.

[7] C. Shepard, H. Yu, N. Anand, L. E. Li, T. L. Marzetta, R. Yang, and L. Zhong, “Argos: Practical many-antenna base stations,” in *Proc. of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, pp. 53–64.

[8] T. L. Marzetta, “Noncooperative cellular wireless with unlimited numbers of base station antennas,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[9] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up MIMO: Opportunities and challenges with very large arrays,” *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[10] E. Nayebi and B. D. Rao, “Semi-blind channel estimation for multi-user massive MIMO systems,” *IEEE Trans. Signal Process.*, vol. 66, no. 2, pp. 540 – 553, Jan. 2018.

[11] S. D. Silverstein, “Application of orthogonal codes to the calibration of active phased array antennas for communication satellites,” *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 206–218, Jan. 1997.

[12] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *J. Roy. Statist. Soc. B*, vol. 39, no. 1, pp. 1–38, 1977.

[13] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[14] R. G. Gallager, *Stochastic processes: theory for applications*, Cambridge University Press, 2013.