Low-complex Bayesian estimator for imperfect channels in massive muti-input multi-output system

Ahmed Hussein Shatti, Ehab AbdulRazzaq Hussein
Department of Electrical Engineering, Faculty of Engineering, University of Babylon, Babylon, Iraq

ABSTRACT

Motivated by the fact that the complexity of the computations is one of the main challenges in large multiple input multiple output systems, known as massive multiple-input multiple-output (MIMO) systems, this article proposes a low-complex minimum mean squared error (MMSE) Bayesian channel estimator for uplink channels of such systems. First, we have discussed the necessity of the covariance information for the MMSE estimator and how their imperfection knowledge can affect its accuracy. Then, two reduction phases in dimension and floating-point operations have been suggested to reduce its complexity: in phase 1, eigenstructure reduction for channel covariance matrices is implemented based on some truncation rules, while in phase 2, arithmetic operations reduction for matrix multiplications in the MMSE equation is followed. The proposed procedure has significantly reduced the complexity of the MMSE estimator to the first order $O(M)$, which is less than that required for the conventional MMSE with $O(M^3)$ in terms of matrix dimension. It has been shown that the estimated channels using our proposed procedure are asymptotically aligned and serve the same quality as the full-rank estimated channels. Our results are validated by averaging the normalized mean squared error (NMSE) over a length of 500 sample realizations through a Monte Carlo simulation using MATLAB R2020a.

Corresponding Author:

Ahmed Hussein Shatti
Department of Electrical Engineering, Faculty of Engineering, University of Babylon
Babylon, Iraq
Email: eng.ahmed.hussein@uob.edu.iq, dr.ehab@uob.edu.iq

1. INTRODUCTION

Massive multiple-input multiple-output (MIMO) systems are appreciated as an enabling technology for 5G and future generations [1]–[8]. The technology improved on traditional systems by employing large-scale antennas at the base station (BS). This allows tens of terminal users (TUs) to be spatially multiplexed at the BS. For this purpose, the base station needs to pick up (to estimate) the uplink channels from all TUs in the cell. However, corresponding to the channel propagation issues, the channel of each TU should be repeatedly estimated during each coherence block. In practice, the individual elements of each channel vector are correlated at the BS, which implies that the practical channel is spatially correlated [9]. On the other hand, a careful design for such a large antenna system, also defined as a massive MIMO system, should be considered to improve antenna efficiency and performance on both the BS and the receiving side [10]–[13].

The channel covariance information describes the spatial characteristics of the channel, and its knowledge plays a crucial role in investigating an optimal minimum mean squared error (MMSE) channel estimate at the BS. The knowledge of statistical information of the channel is also relevant to the pilot contamination suppression [14]–[16] and source allocation [17]. In almost all literature on massive MIMO,

covariance matrices are assumed to be perfectly known [18], which is debatable because the dimensions of the covariance matrices change (growing) over time [19], [20]. However, in practice, the BS has incomplete knowledge about the statistics of the covariance matrices of the channels and needs to be estimated. The common approach to capturing such matrices is by evaluating sample covariance matrices for them, where the BS needs to see many pilot signal observations from each TU in the network. However, the estimated channels produced by this method are full-rank covariance matrices with large dimensions. Hence, promising approaches have been suggested in [21]–[24], to estimate such practical large-dimensional matrices, based on regularizing the diagonal and off-diagonal elements.

Since these large matrices have to be manipulated every millisecond, their computational complexity will be a bottleneck problem. The underlying linear algebra operations such as matrix inversions and matrix-matrix multiplications have well-known defined structures that can be efficiently implemented in hardware and make the implementation of such large matrices possible. In this paper, two main phases were applied depending on the power of linear algebra and its possibilities in solving problems of large matrices. In the first one, a proposed truncation approach was applied for the large matrices that are estimated in [23] to reduce their dimensions/eigenstructure. We have done this based on the truncated singular value decomposition (TSVD) with a new proposed threshold (denoted by successive empirical threshold) and the singular value hard thresholding (SVHT) with the optimal threshold [25]. In the second phase, a factor-solve procedure [26] was implemented to reduce the arithmetic operation or what is called floating-point operations (FLOPS) of the MMSE computations. Our suggested procedure shows that the complex computations of our proposed MMSE estimator are significantly reduced to the order of $O(M)$ instead of $O(M^3)$ corresponding to the conventional MMSE estimator. On the other hand, the quality of the estimated channels of our proposed approach has been validated in terms of the normalized MSE measurement. Due to the large reduction in complex computations, the results of our work will have important implications for the channel estimation using the MMSE estimator.

The paper is organized: In section 2, two different covariance matrix estimation models are presented depending on the knowledge of the statistical information of the channel. In section 3, a proposed approach is suggested for the full-rank matrices developed in section 2. Evaluation results are discussed, analyzed, and validated in section 4. Finally, the paper concluded in section 5.

2. CHANNEL ESTIMATION MODELS

The main purpose of the estimation process is to find an approximate value for the unknown variable based on some observations. In this paper, we are particularly concerned with the Bayesian estimator, where the unknown variable is the channel between the TU and the BS and is a realization of a random variable with known or partially known statistical information at the BS. On the other hand, the observation signals are the received pilot signals at the BS. Therefore, in motivation with the knowledge of channel statistics, two models of channel estimation will be discussed here. The first one depends on the local scattering model [18], which has completely known statistical information at the BS, while the second model deals with the practical case, that is, when the BS has imperfect covariance information. We will use the following corollaries from [18], and [27] to define the local scattering model and the Bayesian estimator, respectively:

Corollary1: The local scattering model: In this model, rich scattering objects are assumed to be surrounded the terminal user (TU), while the BS is elevated in a way that makes it free from scatters as shown in Figure 1, which is a non-line-of-sight (NLoS) propagation channel model. In almost all the literature on massive MIMO systems, the local scattering model is used to show how the channels are strongly spatially correlated, that is, which spatial direction is more likely than others. If $h_{\ell k}^{j} \in \mathbb{C}^{M_{j}}$ denotes the uplink channel vector from a single-antenna TU k in cell ℓ to a BS j equipped with a uniform linear array (ULA) antennas, then the complex Gaussian distribution of the channel will be denoted by $h_{\ell k}^{j} \sim \mathbb{C}_{\mathcal{N}}\left(0_{M_{j}}, R_{\ell k}^{j}\right)$, where $R_{\ell k}^{j} \in \mathbb{C}^{M_{j} \times M_{j}}$, is a positive semi-definite covariance matrix. The $(l, m)^{th}$ element of $R_{\ell k}^{j}$ is described:

$$\left[R_{\ell k}^{j}\right]_{l,m} = \beta_{\ell k}^{j} e^{2\pi j d_H (l-m) \sin(\varphi)} e^{-\frac{\sigma_{\varphi}^2}{2}(2\pi d_H (l-m) \cos(\varphi))^2} \qquad (1)$$

where β is the average channel gain, φ denotes the nominal angle of the reached planner array from the TU, and σ_{φ} refers to the angular standard deviation from the nominal angle, assumed to be Gaussian distributed.

Corollary2: The Bayesian MMSE estimator: the MMSE estimator is defined as the mean of the posterior PDF of the random variable h given the observation y i.e., $\hat{h}(y) = \mathbb{E}\{h|y\}$. Also, it is defined as the

estimator that minimizes the expectation of the squared difference of the loss function $\ell(h - \hat{h}(y))$, that is it will minimize the MSE $\mathbb{E}\{\|h - \hat{h}(y)\|^2)\}$. Let $\hat{h}(y)$ denotes the Bayesian estimator of the unknown variable $h \sim \mathbb{C}_{\mathcal{N}}(0_M, R)$ based on the observation signal $y = hq + n$, where $q \in \mathbb{C}$ and n is the additive noise with $n \sim \mathbb{C}_{\mathcal{N}}(0_M, S)$. If R and S are positive semi-definite covariance matrices and they are independent, then the Bayesian MMSE estimator can be written as: $\hat{h}(y) = q^* R(|q|^2 R + S)^{-1} y$ and the MSE is given by $MSE = tr(R - |q|^2 R(|q|^2 R + S)^{-1} R)$. More details are in [27].
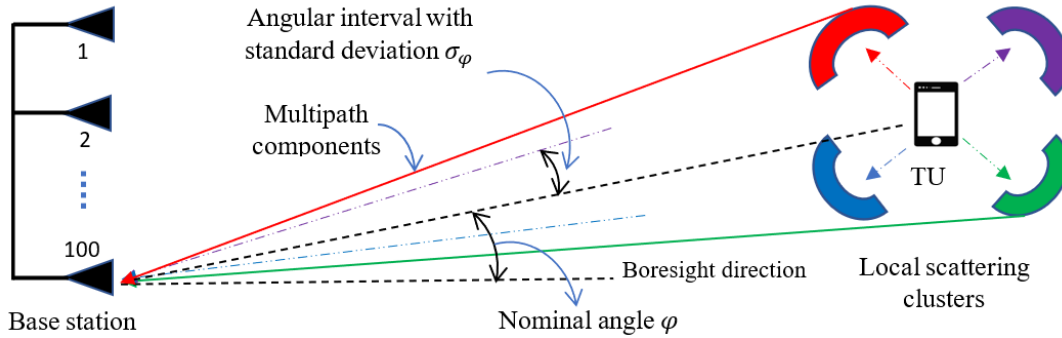

Figure 1. NLoS, local scattering model, four of many path components are shown with key parameters: $d_H = 1/2$ wavelength spacing antenna, $\varphi =$ nominal angle of TU, and $\sigma_\varphi =$ angular standard deviation

2.1. Channel estimation of the completely known covariance matrix

In general, to estimate the channel h_{jk}^{j} at the BS j in cell j from a particular TU k, first, the BS observes the sending pilot signals (Y_j^P) by all TUs in the network, each with τ_p pilot sequence length, then, correlates its observation with the pilot sequence $\phi_{jk} \in \mathbb{C}^{\tau_p}$ of the desired kth TU. The received pilot signal Y_j^P and its processing signal y_{jjk}^P are defined in (2) and (3), respectively:

$$Y_j^P = \underbrace{\sum_{k=1}^{K_j} \sqrt{p_{jk}} \, h_{jk}^{j} \, \phi_{jk}^T}_{intra-cell\ TUs} + \underbrace{\sum_{\ell=1}^{L} \sum_{i=1}^{K_i} \sqrt{p_{\ell i}} \, h_{\ell i}^{j} \, \phi_{\ell i}^T}_{inter-cell\ TUs\ (\ell \neq j)} + N_j^P \tag{2}$$

$$y_{jjk}^P = Y_j^P \phi_{jk}^* = \sum_{\ell'=1}^{L} \sum_{i'=1}^{K_{i'}} \sqrt{p_{\ell' i'}} \, h_{\ell' i'}^{j} \phi_{\ell' i'}^T \, \phi_{jk}^* + N_j^P \phi_{jk}^* \tag{3}$$

where $Y_j^P \in \mathbb{C}^{M_j \times \tau_p}$, $N_j^P \in \mathbb{C}^{M_j \times \tau_p} \sim \mathbb{C}_{\mathcal{N}}(0_{M_j}, \sigma_{UL}^2)$ is the additive noise at the ULA of the BS, $\phi_{\ell' i'}$ is the pilot sequence associated with each ith TU in the network, and $p_{\ell' i'}$ is the sending signal power during the pilot transmission period. However, (3) can be rewritten in general form:

$$y_{jjk}^P = \sqrt{p_{jk}} \, \tau_p h_{jk}^{j} + \sum_{(\ell', i') \in \mathcal{P}_{jk}} \sqrt{p_{\ell' i'}} \, \tau_p h_{\ell' i'}^{j} + N_j^P \phi_{jk}^* \tag{4}$$

where $\tau_p = \phi_{\ell' i'}^T \phi_{jk}^*$, and the set $(\ell', i) \in \mathcal{P}_{jk}$ is defined for all TUs that have a similar pilot sequence as the kth desired TU. If the last equation is normalized by $s_{jk}^{tr} = \sqrt{p_{jk}} \, \tau_p$, the processed received signal will be:

$$y_{jjk}^P = h_{jk}^{j} + \sum_{\ell=1, \ell \neq j}^{L} h_{\ell i}^{j} + \frac{1}{\sqrt{s_{jk}^{tr}}} N_j^P \phi_{jk}^* \tag{5}$$

Now, based on the predesigned pilots at the BS and the normalized processed received signal y_{jjk}^P, the MMSE estimate of the channel h_{jk}^{j} [18], [23] will has the formula:

$$\hat{h}_{jk}^{j} = R_{jk}^{j} (\Psi_{jk}^{j})^{-1} y_{jjk}^P \tag{6}$$

where $\Psi_{jk}^{j} = \mathbb{E}\left\{ y_{jjk}^P (y_{jjk}^P)^H \right\} / \tau_p$, refers to the sum of correlation matrices of all interfering TUs, given by (7).

$$\Psi_{jk}^{j} = \sum_{\ell=1}^{L} R_{\ell k}^{j} + \frac{1}{s_{jk}^{tr}} I_M \tag{7}$$

However, the estimator in (6) is a channel vector that is used to minimize the second-order moment of estimation error $\mathbb{E}\{\left\|\tilde{h}_{jk}^{j}\right\|^2 = \left\|h_{jk}^{j} - \hat{h}_{jk}^{j}\right\|^2\}$. The vectors $\hat{h}_{jk}^{j}$ and $\tilde{h}_{jk}^{j}$ are both independent random vectors, their distributions described as $\hat{h}_{jk}^{j} \sim \mathbb{C}_{\mathcal{N}}(0_{M_j}, \widehat{R}_{jk}^{j})$ and $\tilde{h}_{jk}^{j} \sim \mathbb{C}_{\mathcal{N}}(0_{M_j}, R_{jk}^{j} - \widehat{R}_{jk}^{j})$, where $\widehat{R}_{jk}^{j} = \mathbb{E}\{\hat{h}_{jk}^{j}(\hat{h}_{jk}^{j})^H\}$ is the statistical information of the estimated channel using the MMSE estimator [24].

$$\widehat{R}_{jk}^{j} = R_{jk}^{j}(\Psi_{jk}^{j})^{-1}R_{jk}^{j} \tag{8}$$

Note that the covariance matrix of the channel estimate $\hat{h}_{jk}^{j}$ is mainly dependent on the statistics of all channels in the network. It means that the BS can determine the MMSE estimator $\hat{h}_{jk}^{j}$ only when it knows the covariance matrices R_{jk}^{j} and Ψ_{jk}^{j}. However, in practice, the BS has no prior knowledge about these statistics. In the following part, we will explain how the BS can estimate the channel covariance matrices practically.

2.2. Channel estimation of the practical covariance matrix

In practice, to estimate the channel vector, the BS needs to see many observations from the sending pilot signal by all TUs in the network. The channel statistics are assumed constant during the entire system bandwidth (over several coherence blocks). Let's start with the estimation of Ψ_{jk}^{j}.

2.2.1. Estimation of Ψ_{jk}^{j}

Suppose the BS j has N_Ψ observations from y_{jjk}^{P} denoted by $y_{jjk}^{P}[1], \dots, y_{jjk}^{P}[N_\Psi]$. Hence, it can be shaped the *sample* covariance matrix as (9).

$$\widehat{\Psi}_{jk}^{j\,(sample)} = \frac{1}{N_\Psi} \sum_{n=1}^{N_\Psi} y_{jjk}^{P}[n](y_{jjk}^{P}[n])^H \tag{9}$$

Corresponding to the ergodic characteristics of the channel and large numbers law, the particular variance of any antenna, indexed by m, converges to the real variance when N_Ψ goes to infinity.

$$\left[\frac{1}{N_\Psi} \sum_{n=1}^{N_\Psi} y_{jjk}^{P}[n](y_{jjk}^{P}[n])^H\right]_{m,m} \xrightarrow{a.s} \left[\Psi_{jk}^{j}\right]_{m,m} \tag{10}$$

Note that the standard deviation in (10) decays as $1/\sqrt{N_\Psi}$. Hence, the BS needs only a few numbers from the observation y_{jjk}^{P} to achieve a near-optimal channel estimate. The same approach is utilized for estimating the $M \times M$ covariance matrix $\widehat{\Psi}_{jk}^{j\,(sample)}$. Similarly, every $(l,m)^{th}$ element in the sample covariance matrix will be close to the symmetrical location in the true covariance matrix Ψ_{jk}^{j}. However, due to the estimation error in the M^2 elements, there is a more challenge to get on a sample covariance matrix whose eigenstructure is well aligned with those of the true Ψ_{jk}^{j} matrix. This may have a crucial impact on the channel estimate in massive MIMO systems since the MMSE estimator exploits such eigenstructure in its major formula in (6). However, since the diagonal elements of the covariance matrices define the sample variances of all antennas at the BS; hence, one can exploit only the diagonal elements in $\widehat{\Psi}_{jk}^{j\,(sample)}$ matrix to formulate a diagonal matrix as $\widehat{\Psi}_{jk}^{j\,(diagonal)}$. However, convex combination approaches that use both the sample and diagonal matrices have been suggested in [21]–[23] to obtain a robust estimation:

$$\widehat{\Psi}_{jk}^{j}(c) = c\,\widehat{\Psi}_{jk}^{j\,(sample)} + (1-c)\widehat{\Psi}_{jk}^{j\,(diagonal)} \tag{11}$$

where $\widehat{\Psi}_{jk}^{j}(c)$ refers to the regularized sample covariance matrix, and c denotes the regularization factor, it takes random values between 0 and 1. Hence, one can optimize this parameter to obtain a robust estimation for the $\widehat{\Psi}_{jk}^{j}$. However, due to the regularization method in these convex approaches, they produce full-rank

covariance matrices from $\widehat{\Psi}_{jk}^{j}$ even the number of observations is less than the number of antennas at the base station.

2.2.2. Estimation of R_{jk}^{j}

In this part, the same approach is followed to estimate the individual covariance matrix R_{jk}^{j}, but with extra pilots for the interfering TUs [23]. A procedure for two-stage of estimations has been suggested here: in the first stage, the BS will estimate $\widehat{\Psi}_{jk,}^{j}{}^{(sample)}$, the sum of sample covariance matrices of all interfering TUs including desired TU (with extra pilots assigned for each TU), while, in the second stage, the BS uses the extra pilot observations N_{extra} of interfering users to estimate their covariance matrices as $\widehat{\Psi}_{jk,-k}^{j}{}^{(sample)}$, without including desired TU. Then, the desired sample covariance matrix of the TU is determined:

$$\widehat{R}_{jk}^{j}{}^{(sample)} = \widehat{\Psi}_{jk}^{j}{}^{(sample)} - \widehat{\Psi}_{jk,-k}^{j}{}^{(sample)} \tag{12}$$

Similarly, using a regularizing factor, the regularized individual matrix is given by (13):

$$\widehat{R}_{jk}^{j}(\alpha) = \alpha\widehat{R}_{jk}^{j}{}^{(sample)} + (1-\alpha)\widehat{R}_{jk}^{j}{}^{(diagonal)} \tag{13}$$

2.2.3. Estimation of h_{jk}^{j}

Depending on the estimated covariance matrices in the previous parts, it can be now computing the approximate MMSE channel estimate:

$$\widehat{h}_{jk}^{j} = A_{jk}^{j}(c,\alpha)\, y_{jjk}^{P} \tag{14}$$

where $A_{jk}^{j}(c,\alpha) = \widehat{R}_{jk}^{j}(\alpha)\,(\widehat{\Psi}_{jk}^{j}(c))^{-1}$, and is mainly dependent on the observation y_{jjk}^{P}. However, the estimation quality can be measured by the normalized MSE in terms of the matrix $A_{jk}^{j}(c,\alpha)$ [18]:

$$\text{NMSE}\left(A_{jk}^{j}\right) = 1 - \frac{2\sqrt{p\tau_p}\,\Re\left(\text{tr}\left(\widehat{R}_{jk}^{j}A_{jk}^{j}\right)\right) - \tau_p\,\text{tr}\left(A_{jk}^{j}\widehat{\Psi}_{jk}^{j}{}^{-1}A_{jk}^{j}{}^{H}\right)}{\text{tr}\left(\widehat{R}_{jk}^{j}\right)} \tag{15}$$

where $\widehat{R}_{jk}^{j}$ and $\widehat{\Psi}_{jk}^{j}{}^{-1} = (p\tau_p\widehat{R}_{jk}^{j} + \sigma_{UL}^{2}I_M)$ are the regularized covariance matrices.

3.   PROPOSED APPROACH

In this section, two phases of complexity reduction have been suggested for getting a lower computational estimator. In phase 1, the covariance matrices of the spatial channels were reduced in dimensions exploiting their eigenstructure properties. In phase 2, an arithmetic operation reduction procedure was followed to minimize matrix multiplications of the MMSE in (14).

3.1. Phase 1: eigenstructure reduction of the covariance matrices

In this phase, we will describe how the dimensions of the channel's covariance matrices are reduced with the aid of their eigenstructure properties. It has used the singular value decomposition (SVD) as the basic math of Lemmas (1) and (2). A greater dimension's reduction for the covariance matrices happens in Lemma (3) with a new proposed threshold the "Successive empirical TSVD threshold (SETSVD)".

3.1.1. Lemma 1. Eigenstructure of R

Let $h \sim \mathbb{C}_{\mathcal{N}}(0_M, R)$ be an arbitrary channel of any TU at the BS. The eigenstructure of the channel covariance can be written as $R = U\Sigma U^H$, where $U \in \mathbb{C}^{M\times M}$ is a matrix that contains the eigendirections and $\Sigma = \text{diag}(\lambda_1, ..., \lambda_M)$ contains the corresponding eigenvalues. If $\tilde{h}_{jk}^{j} \sim \mathbb{C}_{\mathcal{N}}(0_{M_j}, R_{jk}^{j} - \widehat{R}_{jk}^{j})$ denotes the estimation error; then, it can be defined its correlation matrix [18]:

$$C = R - p\tau_p R(p\tau_p R + \sigma_{UL}^{2}I_M)^{-1}R = U\left(\Sigma - p\tau_p\Sigma(p\tau_p\Sigma + \sigma_{UL}^{2}I_M)^{-1}\Sigma\right)U^H$$

Substitute Σ by the individual eigenvalue λ_i, then:

$$C = U \, diag(\lambda_1 - \frac{p\tau_p\lambda_1{}^2}{p\tau_p\lambda_1+\sigma_{UL}^2}, \dots\dots\dots, \lambda_M - \frac{p\tau_p\lambda_M{}^2}{p\tau_p\lambda_M+\sigma_{UL}^2}) \, U^H \tag{16}$$

Also, it can be rewritten (16):

$$C = U \, diag(\frac{\lambda_1}{SNR_{jk}^{eff}\frac{\lambda_1}{\beta_{jk}^j}+1}, \dots\dots\dots, \frac{\lambda_M}{SNR_{jk}^{eff}\frac{\lambda_M}{\beta_{jk}^j}+1}) \, U^H \tag{17}$$

where $SNR_{jk}^{eff} = p_{jk}\tau_p\beta_{jk}^j/\sigma_{UL}^2$ denotes the effective SNR, and τ_p refers to the channel processing gain. Note that the eigenvectors of matrix C are the same as the eigenvectors of matrix R. Only the eigenvalues are different and take a fractional form as described in (17). However, each eigenvalue in C represents the variance of the estimation error in each eigendirection, and it approaches zero when $SNR_{jk}^{eff} \to \infty$. Each eigendirection in R with a large eigenvalue will have a smaller normalized error variance. Hence, killing the unnecessary eigenvalues and keeping only the large ones, will have a major impact on reducing the errors in the dimensions of the covariance matrices, and the complex computations as well.

3.1.2. Lemma 2. Reduced eigenstructure of R

Let $R = U\Sigma U^H$ be the SVD structure of an arbitrary large matrix $R \in \mathbb{C}^{M \times M}$. Hence, the reduced eigenstructure of the matrix depending on its rank r will be $\tilde{R} \approx \tilde{U}\tilde{\Sigma}\tilde{U}^H$. The matrix $\tilde{U}$ contains the first $r \ll M$ columns of U and $\tilde{\Sigma}$ contains the first $r \times r$ upper left block of the Σ.

$$R = U\Sigma U^H = [U_r|U_{r+1 \to M}]\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}[U_r|U_{r+1 \to M}]^H =$$
$$[U_r|U_{r+1 \to M}]\begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} U_r^H \\ U_{r+1 \to M}^H \end{bmatrix} = [U_r][\Sigma_r][U_r^H]$$

The last term is the approximation of R and is sometimes denoted as the reduced SVD of R or $\tilde{R} \approx \tilde{U}\tilde{\Sigma}\tilde{U}^H$.

3.1.3. Lemma 3. Successive-empirical TSVD threshold vs. optimal SVHT threshold

If R is a large $M \times M$ matrix, two well-known truncation approaches have been used to reduce such large matrices. The first one is called the truncated singular value decomposition (TSVD) which is given by (18):

$$R_{TSVD} = \sum_{i=1}^{r} \lambda_i u_i v_i' \tag{18}$$

where r refers to the rank of matrix R, λ_i denotes the individual eigenvalue, $u_i v_i'$ are left and right eigenvectors, respectively. The equation (18) represents a standard truncation rule since it uses the rank of the matrix to find its approximation. However, to get more reduction in the eigenstructure, we have suggested successive threshold values [28], [29] (SETSVD) that are extracted empirically from the data of the covariance matrices and applied in (18). The second type of truncation approach utilizes the SVHT rule [25], which has the formula:

$$R_{SVHT} = \sum_{i=1}^{M} \mathcal{T}_H(\lambda_i; \tau) u_i v_i' \tag{19}$$

where $\mathcal{T}_H(\lambda_i; \tau) = \{\lambda_i : \lambda_i \geq \tau\}$ denotes the hard thresholding rule. The optimal SVHT threshold for $R^{M \times M}$ is given by: $\tau = 2.858 \, y_{med}$, where $y_{med} =$ median of the singular values.

3.2. Phase 2: floating-point operations reduction of the MMSE estimator

Often, the computational complexity is measured by counting the FLOPS of the arithmetic operations and expressed as a polynomial (function of the matrices dimensions). Each flop point is defined as one addition, multiplication, subtraction, or division of two floating-point numbers [26]. To simplify the cost of the total FLOPS, we often ignore the lower terms and keep the leading terms (the terms with the largest exponents). For example, to evaluate the number of FLOPS of the inner product of two vectors x and y both are $\in \mathbb{R}^M$, the term $x^T y$ will multiply all $x_i y_i$ elements and then add them with M multiplication and $M - 1$ addition i.e., $2M - 1$ FLOPS are required, which can be then simplified to $2M$ FLOPS by keeping only the

leading term. Hence, we can say that the inner product requires $2M$ FLOPS only. Another example is matrix-vector multiplication y, where $\Psi \in \mathbb{R}^{M \times M}$ and $y \in \mathbb{R}^M$ requires $2M^2$ FLOPS which represents the calculation of the inner product of the M components of y with each row of the matrix. The standard complexity of computations of the MMSE estimator in (6) is given in [18]: i) correlating $y_{\acute{j}\acute{j}k}^P$ with predesigned pilot sequences at the BS, given by $M_{\acute{j}}\tau_p$, ii) multiplication $y_{\acute{j}\acute{j}k}^P$ with $R_{\acute{j}k}^{\acute{j}}$ and $\Psi_{\acute{j}k}^{\acute{j}}$ matrices, given by $(4M_{\acute{j}}^3 - M_{\acute{j}})/3$, and iii) per additional TU computations, given by $M_{\acute{j}}^2$.

The problems of matrix inversions and matrix-vector multiplications are often arising when we try to solve linear equations such as the MMSE estimator in (6). Notably, the generic method for solving these linear equations requires a computational complexity that grows approximately as M^3. However, to solve such equations more efficiently with computational savings, the factor-solve method [26] with two essential steps, matrix factorization and backward solving provides an effective approach that can be followed. In this paper, we have used the factor-solve method based on SVD matrix factorization with and without reducing the eigenstructure of the covariance matrices.

3.2.1. Factor-solve method without reducing eigenstructure of R

From corollary2 we have derived $\hat{h}(y) = q^* R(|q|^2 R + S)^{-1}y$, which can be expressed in terms of the eigenstructure of R described in Lemma (1): $\hat{h}(y) = q^* U\Sigma U^H(|q|^2 U\Sigma U^H + S)^{-1}y$. The factor-solve procedure will be: For SVD factorization, the cost will be $2mn^2 + 11n^3$ for an $m \times n$ matrix [30] without any reduction in the eigenstructure, that is if $m = n = M$, the cost will be $13M^3$. The factorization and solving steps can be summarized: i) factorization: factor R as $R = U\Sigma U^H$ without reducing eigenstructure this has a cost ($13M^3$) FLOPS; ii) backward solving: we start with inverse matrix computations.

– Solving $\Sigma U^H = (M \times M) * (M \times M) = M \times M$ matrix, costs ($2M^2$) FLOPS since Σ is diagonal.
– Solving $U(\Sigma U^H) = (M \times M) * (M \times M) = M \times M$ matrix, costs ($2M^3$) FLOPS.
– Multiplication ($M \times M$) matrix by $|q|^2$ costs (M^2) FLOPS.
– Addition ($M \times M$) matrix with S is also ($M \times M$) matrix, costs (M^2) FLOPS.

Now, let $|q|^2 U\Sigma U^H + S = \Psi$, then $\Psi^{-1}y = w \Rightarrow \Psi w = y$ is a set of M linear equations with M variables that costs $(2/3)M^3$ [26]. Hence, the total number of FLOPS for solving the inverse matrix is $2M^3 + 4M^2 + (2/3)M^3$ FLOPS. The rest of the computations of the multiplication $E^{-1}y$ by $q^* U\Sigma U^H$ using the backward solving procedure will cost $4M^2 + 3M$. Thus, the total number of FLOPS will be the sum of the cost at both the factorization and backward solving steps, that is $2M^3 + (2/3)M^3 + 8M^2 + 3M$. Keeping only the leading terms, the total cost of computations without reducing the eigenstructure of R will be $(8/3)M^3$ FLOPS, which has $O(M^3)$ order of complexity.

3.2.1. factor-solve method with reducing eigenstructure of R

Depending on the reduced eigenstructure in Lemma (2), the MMSE estimator in corollary2 can be rewritten as $\hat{h} = \tilde{U}_R \tilde{\Sigma}_R U_R^H (\tilde{U}_{\Psi^{-1}} \tilde{\Sigma}_{\Psi^{-1}} \tilde{U}_{\Psi^{-1}}^H)y$ using a small trick here, by equating $(|q|^2 U\Sigma U^H + S)^{-1} = \Psi^{-1}$ and making another factorization for the resultant matrix. Although this procedure costs us another factorization cost for Ψ^{-1} matrix but it contributes to reducing the computation to a lower order. However, following the same procedure above, the total number of FLOPS is calculated: first, the multiplication $\tilde{U}_{\Psi^{-1}}^H y$ is an $(r \times M).(M \times 1)$ matrix-vector multiplication that yields $(r \times 1)$ vector and results in $2Mr$ FLOPS. Second, the next multiplication $\tilde{\Sigma}_{\Psi^{-1}}(\tilde{U}_{\Psi^{-1}}^H y)$ is another matrix-vector multiplication that produces $2r$ FLOPS, since $\tilde{\Sigma}_{\Psi^{-1}}$ is a diagonal matrix that multiplies (r) diagonal elements from $(r \times r)$ matrix by $(r \times 1)$ vector, resulting in another $(r \times 1)$ vector. The last multiplication $\tilde{U}_{\Psi^{-1}}(\tilde{\Sigma}_{\Psi^{-1}}(\tilde{U}_{\Psi^{-1}}^H y))$ is multiplied $(r \times M).(M \times 1)$ and results in $2Mr$ FLOPS. Thus, the number of FLOPS in solving Ψ^{-1} matrix is given by $4Mr + 2r$. The same procedure was followed for the R matrix, which in turn results in the number of FLOPS $4Mr + 2r$. Hence, for solving steps, the cost is $2(4Mr + 2r)$ FLOPS. By adding the cost of factorization step for the reduced eigenstructure $(4Mr^2 - (4/3)r^3)$ [30], the total number of FLOPS for the MMSE estimator will be $2(4Mr + 2r) + 2(4Mr^2 - (4/3)r^3)$ FLOPS. Since $\ll M$, hence, the last polynomial can be more simplified by keeping only the leading terms $8Mr^2$ which has a first-order $O(M)$ of complexity. This result represents the major contribution of our work; hence, our proposed MMSE estimator requires only $O(M)$ instead of $O(M^3)$ of complex computations of the conventional MMSE estimator. The complexity of our proposed procedure and the state of the art in [22] are compared in Table 1.

Moreover, since we have assumed constant statistics during the coherence time. So, R and Ψ matrices will be identical for all subscribers in the coherence block. Hence, the multiplication by these matrices is precomputed at one time and only for one TU then accounts for all users in the same coherence time.

Table 1. A comparison of complex computations in MMSE estimator

Algorithm	Computational complexity	Total number of (flops)
Factor-solve algorithm (without reducing eigenstructure)	$O(M^3)$	$2M^3 + (2/3) M^3 + 8M^2 + 3M \approx (8/3) M^3$
Factor-solve algorithm (our proposed) (with reducing eigenstructure)	$O(M)$	$2(4Mr + 2r) + 2(4Mr^2 - (4/3) r^3) \approx 8Mr^2$
Polynomial expansion channel (PEACH) algorithm [22]	$O(M^2)$	$k_c[(8L + 4)M^2 - (4L + 2)] + k_s[2M^2 - M]$ Where L is a polynomial degree, k_c is the precomputations per channel realization, and k_s is the precomputations per coherence time.

4.   RESULTS ANALYSIS

To evaluate the performance of our proposed approach, we have considered the scenario of the local scattering model shown in Figure 1, with the following parameters: the number of antennas at the BS $M = 100$, the wavelength antenna spacing $d_H = 1/2$, and the angular standard deviation $\sigma_\varphi = 10^o$, assuming Gaussian distribution with an effective signal-to-noise ratio of 10 dB. Our proposed low-complex MMSE estimator is implemented using the NMSE in (14) with different ranks from the eigenstructure of covariance matrices and averaged over different sample realizations (500 samples) using a Monte Carlo simulation in Figure 2. It is found that the curve of the NMSE of the truncated matrices of our proposed approach has aligned with the curve of the full-rank covariance matrices as shown in Figure 2.
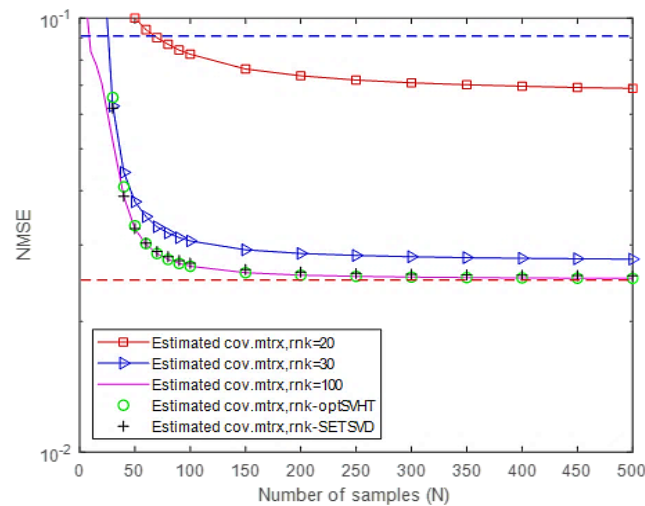

Figure 2. NMSE of the estimated spatially correlated channel for intra-cell TU averaged over a range of samples under imperfect knowledge covariance matrices and different ranks, the number of observations=50

Also, it has been shown that the curve of the NMSE decays asymptotically and close to the lower bound (the case of closing to minimum MSE) with about (100 samples) range, which is the case for the number of antennas at the BS. On the other hand, the complex calculations including reduced and full-rank eigenstructure of covariance matrices are concluded in Table 2. Behind the large reduction in the eigenstructure of the covariance matrices with SETSVD and (optimal SVHT) thresholds; the computational complexity of the estimated channels has been significantly reduced, especially at the ranks $r = 35$ and $r = 48$. However, we have obtained undesirable estimated channels when reducing the eigenstructure of the covariance matrices to ranks $r = 20$ and $r = 30$ as shown in Figure 2 (blue and red curves). As mentioned in (17), the error in each eigendirection is inversely proportional to the SNR_{jk}^{eff}, this fact can be clearly seen in Figure 3, the NMSE curve decays in two orders of magnitude when the effective signal-to-noise ratio becomes as high as 20 dB. Our results in Figure 3 show that the estimation error with the truncated matrices is even better than the completely known covariance matrix of the local scattering model since the error in the last one is larger. Figure 4 shows the keeping and killing eigenvalues of the channel covariance matrices, however, it shows how we can ultimately reduce the eigenstructure of the covariance matrices when using the proposed empirical threshold and the optimal as well.

Table 2. Complex computations of the low-complex MMSE estimator based on reduced eigenstructure

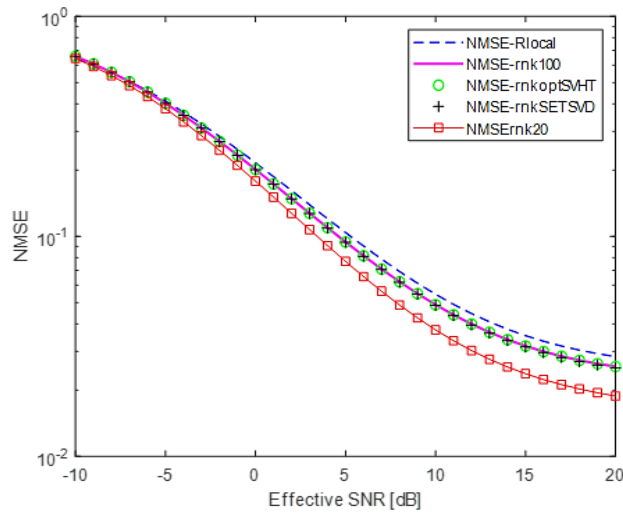| Rank	Flops Polynomial	Flops Computations
$r1 = 20$	$8Mr^2$	32 e+04
$r2 = 30$	$8Mr^2$	72 e+04
$r3 = 35$, ESTSVD	$8Mr^2$	98 e+04
$r4 = 48$, optSVHT	$8Mr^2$	184.32 e+04
$r5 = 100$	$(8/3)M^3$	2.666666+06

Figure 3. NMSE of the estimated channels using the truncated covariances averaged over different values from the effective SNR and different ranks, the number of observations=50
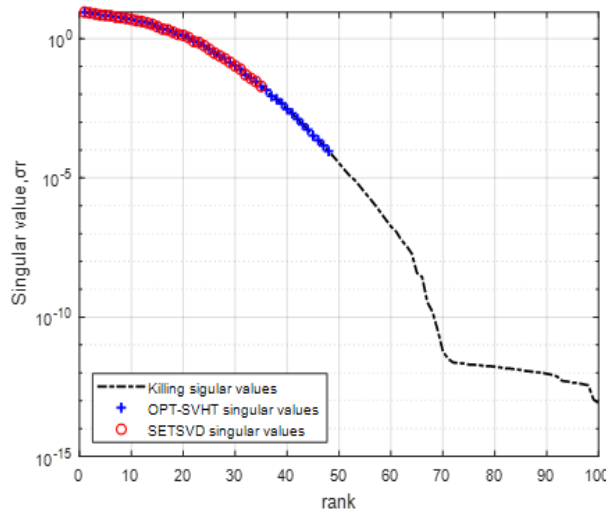
Figure 4. truncation singular values using optimal SVHT threshold=48 and the proposed SETSVD threshold=35

5.   CONCLUSION

In this paper, we have proposed a low-complex MMSE channel estimator for the uplink (UL) channels in the massive MIMO system. First, we have discussed the necessity of the covariance information for the MMSE channel estimator in such a large MIMO system and how can they affect the accuracy of the estimator. Then, we have used two phases of a procedure to reduce the computational complexity in the conventional MMSE estimator. A truncation approach has been used in phase 1 to reduce the eigenstructure of the estimated channels based on the well-known matrix truncation methods TSVD and SVHT, while the factor-solving procedure has been followed in phase 2 to reduce the total number of FLOPS of the MMSE

estimator. The main advantage of this article is that the MMSE estimator has been reduced to the order of $O(M)$ complexity which is less than that required for the conventional MMSE estimator. It is noticed that the estimated channels with reduced eigenstructure have served the same quality as the full-rank channels while keeping lower computational complexity. Also, it is found that the estimation error with truncated matrices is less than the error when using the completely known covariance matrix of the local scattering model. The precomputations of the MMSE estimator at the BS including covariance matrices multiplications are significantly reduced when we have applied the proposed SETSVD. The last one outperforms the optimal-SVHT threshold in reducing the computations listed in Table 2, in particular when using the rank $r = 35$, which is the best empirical rank for approximating the covariance matrices without estimation error, that is the estimated channels at this rank keep a minimum MSE for the Bayesian estimator.

REFERENCES

[1]   T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010, doi: 10.1109/TWC.2010.092810.091092.

[2]   J. Hoydis, S. Ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, Feb. 2013, doi: 10.1109/JSAC.2013.130205.

[3]   E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, Feb. 2014, doi: 10.1109/MCOM.2014.6736761.

[4]   E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, Feb. 2016, doi: 10.1109/MCOM.2016.7402270.

[5]   T. Van Chien, E. Bjornson, and E. G. Larsson, "Joint power allocation and user association optimization for massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6384–6399, Sep. 2016, doi: 10.1109/TWC.2016.2583436.

[6]   L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE Journal on Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, Oct. 2014, doi: 10.1109/JSTSP.2014.2317671.

[7]   T. L. Marzetta, "Massive MIMO: an introduction," *Bell Labs Technical Journal*, vol. 20, pp. 11–22, 2015, doi: 10.15325/BLTJ.2015.2407793.

[8]   F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, and F. Tufvesson, "Scaling up MIMO: opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan. 2013, doi: 10.1109/MSP.2011.2178495.

[9]   X. Gao, O. Edfors, F. Rusek, and F. Tufvesson, "Massive MIMO performance evaluation based on measured propagation data," *IEEE Transactions on Wireless Communications*, vol. 14, no. 7, pp. 3899–3911, Jul. 2015, doi: 10.1109/TWC.2015.2414413.

[10] M. Alibakhshikenari *et al.*, "Interaction between closely packed array antenna elements using meta-surface for applications such as MIMO systems and synthetic aperture radars," *Radio Science*, vol. 53, no. 11, pp. 1368–1381, Nov. 2018, doi: 10.1029/2018RS006533.

[11] M. Alibakhshikenari *et al.*, "High-gain on-chip antenna design on silicon layer with aperture excitation for terahertz applications," *IEEE Antennas and Wireless Propagation Letters*, vol. 19, no. 9, pp. 1576–1580, Sep. 2020, doi: 10.1109/LAWP.2020.3010865.

[12] M. Alibakhshikenari, B. S. Virdee, C. H. See, R. A. Abd-Alhameed, F. Falcone, and E. Limiti, "Surface wave reduction in antenna arrays using metasurface inclusion for MIMO and SAR systems," *Radio Science*, vol. 54, no. 11, pp. 1067–1075, Nov. 2019, doi: 10.1029/2019RS006871.

[13] M. Alibakhshikenari *et al.*, "A comprehensive survey on 'various decoupling mechanisms with focus on metamaterial and metasurface principles applicable to SAR and MIMO antenna systems,'" *IEEE Access*, vol. 8, pp. 192965–193004, 2020, doi: 10.1109/ACCESS.2020.3032826.

[14] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 264–273, Feb. 2013, doi: 10.1109/JSAC.2013.130214.

[15] E. Bjornson, J. Hoydis, and L. Sanguinetti, "Pilot contamination is not a fundamental asymptotic limitation in massive MIMO," in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–6., doi: 10.1109/ICC.2017.7996674.

[16] H. Yin, D. Gesbert, M. C. Filippou, and Y. Liu, "Decontaminating pilots in massive MIMO systems," in *2013 IEEE International Conference on Communications (ICC)*, Jun. 2013, pp. 3170–3175., doi: 10.1109/ICC.2013.6655031.

[17] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing-the large-scale array regime," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013, doi: 10.1109/TIT.2013.2269476.

[18] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends in Signal Processing*, vol. 11, no. 3–4, pp. 154–655, 2017, doi: 10.1561/2000000093.

[19] V. C. Rodrigues, J. C. Marinello, and T. Abrão, "Exponential spatial correlation with large-scale fading variations in massive MIMO channel estimation," *Transactions on Emerging Telecommunications Technologies*, vol. 30, no. 5, May 2019, doi: 10.1002/ett.3563.

[20] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 5113–5129, Nov. 2008, doi: 10.1109/TIT.2008.929938.

[21] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, no. 2, pp. 365–411, Feb. 2004, doi: 10.1016/S0047-259X(03)00096-4.

[22] N. Shariati, E. Bjornson, M. Bengtsson, and M. Debbah, "Low-complexity polynomial channel estimation in large-scale MIMO with arbitrary statistics," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 815–830, Oct. 2014, doi: 10.1109/JSTSP.2014.2316063.

[23] E. Bjornson, L. Sanguinetti, and M. Debbah, "Massive MIMO with imperfect channel covariance information," in *2016 50th Asilomar Conference on Signals, Systems and Computers*, Nov. 2016, pp. 974–978., doi: 10.1109/ACSSC.2016.7869195.

[24] A. K. Kocharlakota, K. Upadhya, and S. A. Vorobyov, "On achievable rates for massive mimo system with imperfect channel covariance information," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 4504–4508., doi: 10.1109/ICASSP.2019.8682163.

[25] M. Gavish and D. L. Donoho, "The optimal hard threshold for singular values is 4/√3," *IEEE Transactions on Information Theory*, vol. 60, no. 8, pp. 5040–5053, Aug. 2014, doi: 10.1109/TIT.2014.2323359.

[26] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004., doi: 10.1017/CBO9780511804441.

[27] S. K. Sengupta and S. M. Kay, "Fundamentals of statistical signal processing: estimation theory," *Technometrics*, vol. 37, no. 4, Nov. 1995, doi: 10.2307/1269750.

[28] A. H. S. Al-isawi, "Improve wavelet threshold selection based on PSO algorithm and successive-approximation method in image compression applications," *Journal of University of Babylon*, vol. 24, no. 3, 2016

[29] S. A. Kadhim and A. H. Shatti, "A new signal de-noising method usingadaptive wavelet threshold based on PSO algorithm and kurtosis measuring for residual noise," *Journal of Babylon University/Engineering Sciences*, vol. 25, no. 1, pp. 8–19, 2017

[30] L. N. Trefethen and D. Bau, *Numerical linear algebra*. 3600 Market Street, 6th Floor Philadelphia, PA 19104-2688: SIAM, 1997., doi: 10.1137/1.9780898719574.

BIOGRAPHIES OF AUTHORS

Ahmed Hussein Shatti ⓘ 🔗 SC ⟳ was born in Babylon-Iraq, in 1982. He received a B.Sc. degree in electrical engineering from the University of Babylon in 2006 and an MSc degree in communications engineering from the international University of KAI for technology and scientific research in 2013, Russia. He is working as a lecturer at the electrical department/college of engineering at the University of Babylon. He is currently working on his Ph.D. degree in the communication field in the same department at the University of Babylon. His research interests include signal processing, cloud computing, IoT, and massive MIMO. He can be contacted at email: eng.ahmed.hussein@uob.edu.iq.

Ehab AbdulRazzaq Hussein ⓘ 🔗 SC ⟳ PhD. MSc. Electrical Engineering was born in Babylon on January 1, 1976. He obtained his BSc degree (1997) in Electrical Engineering at the Faculty of Engineering, University of Babylon, and his MSc degree (2000), in electrical engineering at the Department of Electrical Engineering, University of Technology, and his Ph.D. Degree from the Department of Electrical Engineering at the Faculty of Engineering, University of Basrah. Currently, he works as an assistant professor in the Electrical Department at the Faculty of Engineering, University of Babylon. His main interest is signal processing, analysis, information transition, sensors, and control system analysis. He can be contacted at email: dr.ehab@uob.edu.iq.