

Analytical Derivation of the Inverse Moments of One-sided Correlated Gram Matrices with Applications

Khalil Elkhailil, *Student Member, IEEE*, Abla Kammoun, *Member, IEEE*,
Tareq Y. Al-Naffouri, *Member, IEEE*, and Mohamed-Slim Alouini, *Fellow, IEEE*

Abstract—This paper addresses the development of analytical tools for the computation of the inverse moments of random Gram matrices with one side correlation. Such a question is mainly driven by applications in signal processing and wireless communications wherein such matrices naturally arise. In particular, we derive closed-form expressions for the inverse moments and show that the obtained results can help approximate several performance metrics such as the average estimation error corresponding to the Best Linear Unbiased Estimator (BLUE) and the Linear Minimum Mean Square Error (LMMSE) estimator or also other loss functions used to measure the accuracy of covariance matrix estimates.

Index Terms—Gram matrices, One sided correlation, Inverse moments, Linear estimation, BLUE, LMMSE, Sample covariance matrix.

I. INTRODUCTION AND BASIC ASSUMPTIONS

THE study of the behavior of random matrices is a key question that appears in many disciplines such as wireless communication, signal processing and economics, to name a few. The main motivation behind this question comes from the fundamental role that random matrices play in modeling unknown and unpredictable physical quantities. In many situations, meaningful metrics expressed as scalar functionals of these random matrices arise naturally. The understanding of their behaviour is, however, a difficult task which might be out of reach especially when involved random models are considered. One approach to tackle this problem is represented by the moment method. It basically resorts to the Taylor expansion of differentiable functionals in order to turn this difficult question into that

of computing the moments of random matrices, where the moment of an $m \times m$ random matrix \mathbf{S} refers to the quantities $\frac{1}{m} \text{tr} \mathbb{E}(\mathbf{S}^r)$ for $r \in \mathbb{Z}$. Along this line, a large amount of works, mainly driven by the recent advances in spectral analysis of large dimensional random matrices, has considered the computation of the asymptotic moments¹. Among the existing works in this direction, we can cite for instance, the work in [1,2] where the computation of the asymptotic moments is used to infer the transmit power of multiple signal sources, that of [3] dealing with the asymptotic moments of random Vandermonde matrices and finally that of [4], where the authors studied the asymptotic behavior of the moments in order to allow for the design of a low complexity receiver with a comparable performance to the LMMSE detector. While working under the asymptotic regime has enabled the derivation of closed-form expressions for all kind of moments, it presents the drawback of being less accurate for finite dimensions. Alternatively, one might consider the exact approach, which relies on the already available expression of the marginal eigenvalues' density of Gram random matrices. Interestingly, this approach, despite its seeming simplicity, has mainly been limited to computing the moments of Wishart random matrices [5,6]. To the best of the authors' knowledge, the case of random Gram matrices has never been thoroughly investigated. This constitutes the principal motivation of the present work.

In this paper, we consider the derivation of the exact moments of random matrices of the form $\mathbf{S} = \mathbf{H}^* \mathbf{\Lambda} \mathbf{H}$, where \mathbf{H} is a $n \times m$ ($n > m$) matrix with independent and identically distributed (i.i.d.) zero-mean unit variance complex Gaussian random entries, $\mathbf{\Lambda}$ is a fixed $n \times n$ positive definite matrix and the superscript * denotes the Hermitian of a matrix. It is worth pointing out that matrix \mathbf{S} cannot be classified as a Wishart random matrix. How-

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org. K. Elkhailil, A. Kammoun, T. Y. Al-Naffouri and M.-S. Alouini are with the Electrical Engineering Program, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; e-mails: {khalil.elkhailil, abla.kammoun, tareq.alnaffouri, slim.alouini}@kaust.edu.sa.

¹The term "asymptotic" referring to the regime in which the dimensions of the underlying random matrix grow simultaneously large.

ever, its positive moments $\frac{1}{m}\text{tr}\mathbb{E}(\mathbf{S}^r), r \geq 0$ coincide with those of the Wishart random matrix $\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{H}\mathbf{H}^*\mathbf{\Lambda}^{\frac{1}{2}}$, and can thus be computed by using existing results on the moments of Wishart matrices. As far as inverse moments are considered ($r < 0$), the same artifice is of no help, mainly because the random matrix $\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{H}\mathbf{H}^*\mathbf{\Lambda}^{\frac{1}{2}}$ becomes singular and thus inverse moments cannot be defined. Besides, from a theoretical standpoint, computing the inverse moments using the Mellin transform derived in [7] is not an easy task. The crude use of the expression provided in [7] brings about singularity issues, as will be demonstrated in the course of the paper. Answering to the so-far unsolved question of computing the inverse moments of Gram random matrices constitutes the main contribution of this work. Additionally, based on the obtained closed-form expression of the exact moments, we revisit some problems in linear estimation. In particular, we provide closed-form expressions of the mean square error for the BLUE and for the LMMSE in both high and low SNR regimes. These expressions are quite useful for performance analysis in Compressed Sensing as it allows to evaluate the mean square error when the support of the estimated signal is known [8,9]. As a further application, we study the selection of the windowing factor used in exponentially weighted sample covariance matrices. In particular, we derive a closed-form expression for the average loss metric, which can be numerically optimized in order to select the optimal windowing factor.

The remainder of this paper is organized as follows. In Section II, we present the main result of this paper giving closed form expressions of the inverse moments. In section III, we provide some potential applications and discuss some performance metrics. We then conclude the paper in section IV. Mathematical details are provided in the appendices.

Notations: Throughout this paper, we use the following notation: $\mathbb{E}(\mathbf{X})$ stands for the expectation of a random quantity \mathbf{X} and $\mathbb{E}_{\mathbf{X}}(f)$ stands for the expected value of f with respect to \mathbf{X} . Matrices are denoted by bold capital letters, rows and columns of the matrices are referred with lower case bold letters (\mathbf{I}_n is the size- n identity matrix). Given a matrix \mathbf{A} , we use $[\mathbf{A}]_{i,j}$ to refer its (i, j) th entry and use \mathbf{A}^t and \mathbf{A}^* to respectively denote its transpose and Hermitian. When \mathbf{A} is a square matrix, we respectively denote by $\text{tr}(\mathbf{A})$, $\det(\mathbf{A})$ and $\|\mathbf{A}\|$ its trace, determinant and spectral norm. Finally, we denote by $\text{diag}(a_1, a_2, \dots, a_n)$ the diagonal matrix with diagonal elements, a_1, a_2, \dots, a_n .

II. EXACT CLOSED-FORM EXPRESSION FOR THE MOMENTS

Consider an $(n \times m)$ random matrix \mathbf{H} with i.i.d zero-mean unit variance complex Gaussian random entries with $m < n$. Let $\mathbf{\Lambda}$ be a deterministic $(n \times n)$ positive definite matrix with distinct eigenvalues $(\theta_1 < \dots < \theta_n)$ and define the Gram matrix \mathbf{S} as

$$\mathbf{S} = \mathbf{H}^*\mathbf{\Lambda}\mathbf{H}. \quad (1)$$

In this paper, we consider the computation of the moments $\mu_{\mathbf{\Lambda}}(r)$ defined as

$$\mu_{\mathbf{\Lambda}}(r) \triangleq \frac{1}{m}\text{tr}(\mathbb{E}_{\mathbf{H}}\{\mathbf{S}^r\}), \quad r \in \mathbb{Z}. \quad (2)$$

As we will see later, in contrast to the positive moments ($r > 0$) which can be directly obtained from the marginal eigenvalues' density, the derivation of the inverse moments ($r < 0$) is not immediate, requiring a careful analysis of the available existing results.

In the following, we will build on the exact approach in order to derive closed-form expressions for the moments $\mu_{\mathbf{\Lambda}}(r)$. The asymptotic moments will be dealt with subsequently in order to illustrate their inefficiency in evaluating the moments of the eigenvalues of Gram matrices with small dimensions.

A. Closed-form Expressions for the Exact Moments in Fixed Dimensions

The exact calculation of moments is mainly based on existing results on the marginal density of the eigenvalues of \mathbf{S} . These results characterize the Mellin transform of the marginal density, the definition of which is given in the following

Definition 1. Denote by $\xi \mapsto f_{\lambda}(\xi)$ the marginal density distribution of an unordered eigenvalue of \mathbf{S} . Then, the Mellin transform of $f_{\lambda}(\cdot)$ is defined as

$$\mathcal{M}_{f_{\lambda}}(s) \triangleq \int_0^{\infty} \xi^{s-1} f_{\lambda}(\xi) d\xi. \quad (3)$$

With the above definition at hand, we are now in position to recall the following Lemma that provides a closed-form expression for the Mellin Transform of the marginal density of \mathbf{S}

Lemma 1. [7, Theorem 2] Let \mathbf{S} be as in (1). Then,

$$\begin{aligned} \mathcal{M}_{f_{\lambda}}(s) = & L \sum_{j=1}^m \sum_{i=1}^m \mathcal{D}(i, j) \Gamma(s + j - 1) \left(\theta_{n-m+i}^{n-m+s+j-2} \right. \\ & \left. - \sum_{l=1}^{n-m} \sum_{k=1}^{n-m} [\Psi^{-1}]_{k,l} \theta_l^{n-m+s+j-2} \theta_{n-m+i}^{k-1} \right), \end{aligned} \quad (4)$$

with $L = \frac{\det(\Psi)}{m \prod_{k<l}^n (\theta_l - \theta_k) \prod_{l=1}^{m-1} l!}$, $\Gamma(\cdot)$ the Gamma function, Ψ the $(n-m) \times (n-m)$ Vandermonde matrix given by

$$\Psi = \begin{bmatrix} 1 & \theta_1 & \cdots & \theta_1^{n-m-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \theta_{n-m} & \cdots & \theta_{n-m}^{n-m-1} \end{bmatrix}$$

and $\mathcal{D}(i, j)$ the (i, j) -cofactor of the $(m \times m)$ matrix \mathcal{C} whose (l, k) -th entry is given by

$$[\mathcal{C}]_{l,k} = (k-1)! \left(\theta_{n-m+l}^{n-m+k-1} - \sum_{p=1}^{n-m} \sum_{q=1}^{n-m} [\Psi^{-1}]_{p,q} \times \theta_{n-m+l}^{p-1} \theta_q^{n-m+k-1} \right).$$

The exact moments $\mu_{\Lambda}(r)$ for $r \geq 0$ can be obtained as a direct consequence of Lemma 1 by replacing s by $r+1$ in (4), thereby yielding the following corollary

Corollary 1. For $r \geq 0$, the moments $\mu_{\Lambda}(r)$ are given by

$$\mu_{\Lambda}(r) = \mathcal{M}_{f_{\lambda}}(r+1),$$

where $\mathcal{M}_{f_{\lambda}}(r+1)$ is given by (4).

In sharp contrast to the case of positive moments ($r \geq 0$), the inverse moments can not be obtained by a crude substitution of s by $-r-1$ for $r \geq 0$. The problem essentially stems from the terms in the sum wherein the Gamma function is applied to negative integers on which it is not defined. This might give the impression that the inverse moments are infinite and cannot be thus computed. Such a quick conclusion goes, however, against the existing results on inverse moments available for wishart matrices, thus leading us to suspect the effect of the Gamma function to be cancelled out in one way or another. In order to study the expected compensation effect, it is natural to analyze the behavior of $\mathcal{M}_{f_{\lambda}}(s-r+1)$ for small values of s . If a limit exists as s goes to zero, one might expect it to coincide with the sought-for value of the r -th moment. Such an intuition is confirmed by theory under some conditions on r as it can be shown from the following Lemma.

Lemma 2. If $r \leq n-m$, then the limit $\lim_{s \rightarrow 0} \mathcal{M}_{f_{\lambda}}(s-r+1)$ exists and

$$\mu_{\Lambda}(-r) = \lim_{s \rightarrow 0} \mathcal{M}_{f_{\lambda}}(s-r+1).$$

Proof: Let $x \mapsto p(x)$ be the probability density function corresponding to the smallest eigenvalue of $\mathbf{H}\mathbf{H}^*$. Then, obviously,

$$\mathcal{M}_{f_{\lambda}}(s-r+1) \leq \frac{1}{\theta_1} \int_0^{\infty} x^{s-r} p(x) dx. \quad (5)$$

It ensues from the Monotone convergence theorem applied to the sequence of functions $(x \mapsto x^{s-r} p(x))_{s \geq 0}$ and $(x \mapsto x^{s-r} f(x))_{s \geq 0}$ that if limits for the both hand sides of (5) exist, they must be equal respectively to $\mu_{\Lambda}(-r)$ and $\frac{1}{\theta_1} \int_0^{\infty} x^{-r} p(x) dx$. From Theorem 5.4, page 62 in [10], we know that $x^{-r} p(x)$ is integrable provided that $r \leq n-m$. Therefore, for $r \leq n-m$, $\mu_{\Lambda}(-r)$ is finite and satisfies

$$\mu_{\Lambda}(-r) = \lim_{s \rightarrow 0} \mathcal{M}_{f_{\lambda}}(s-r+1).$$

■

From Lemma 2, we can see that the computation of the moments $\mu_{\Lambda}(-r)$ amounts to evaluating the limit of $\mathcal{M}_{f_{\lambda}}(s-r+1)$ as s goes to zero. A careful scrutiny of the expression of $\mathcal{M}_{f_{\lambda}}(s-r+1)$ reveals that the sum over index j creates two types of terms. The first type corresponds to those indices of j for which $-r+j-1$ is positive. The limits of these terms can be computed normally by setting s to 0 since $\Gamma(-r+j-1)$ is properly defined. The second type of terms is more difficult to analyze, since it corresponds to those indices of j for which $-r+j-1$ is negative. Obviously, these two types of terms cannot be handled similarly. In light of this observation, it is sensible to decompose $\mathcal{M}_{f_{\lambda}}(s-r+1)$ as the sum of two quantities depending on the value of j , depending on whether it is below or above $r+1$. This decomposition writes as

$$\mathcal{M}_{f_{\lambda}}(s-r+1) = \mathcal{M}_1(s-r+1) + \mathcal{M}_2(s-r+1), \quad (6)$$

where

$$\begin{aligned} \mathcal{M}_1(s) &= L \sum_{j=1}^r \sum_{i=1}^m \mathcal{D}(i, j) \Gamma(s+j-1) \left(\theta_{n-m+i}^{n-m+s+j-2} \right. \\ &\quad \left. - \sum_{l=1}^{n-m} \sum_{k=1}^{n-m} [\Psi^{-1}]_{k,l} \theta_l^{n-m+s+j-2} \theta_{n-m+i}^{k-1} \right). \\ \mathcal{M}_2(s) &= L \sum_{j=r+1}^m \sum_{i=1}^m \mathcal{D}(i, j) \Gamma(s+j-1) \left(\theta_{n-m+i}^{n-m+s+j-2} \right. \\ &\quad \left. - \sum_{l=1}^{n-m} \sum_{k=1}^{n-m} [\Psi^{-1}]_{k,l} \theta_l^{n-m+s+j-2} \theta_{n-m+i}^{k-1} \right). \end{aligned}$$

We will first handle the second term $\mathcal{M}_2(s-r+1)$, gathering indices j for which $-r+j-1$ is positive. Interestingly, we can prove that its limit is zero as $s \downarrow 0$,

which shows that it does not contribute in the expression of the final moment.

Proposition 1. *The term $\mathcal{M}_2(s-r+1)$ vanishes as s goes to zero i.e,*

$$\lim_{s \rightarrow 0} \mathcal{M}_2(s-r+1) = 0, \quad r = 1, \dots, m.$$

Proof: See Appendix A for a proof. ■

The expression of the moment is thus totally ruled out by the contribution of the first term $\mathcal{M}_1(s-r+1)$. Before providing the expression of its limit as $s \downarrow 0$, we introduce the following piece of notation.

$$\begin{aligned} \mathbf{a}_j &= \left[\theta_1^{n-m-r+j-1}, \theta_2^{n-m-r+j-1}, \dots, \theta_{n-m}^{n-m-r+j-1} \right]^t. \\ \mathbf{D}_i &= \text{diag} \left(\log \left(\frac{\theta_{n-m+i}}{\theta_1} \right), \log \left(\frac{\theta_{n-m+i}}{\theta_2} \right), \right. \\ &\quad \left. \dots, \log \left(\frac{\theta_{n-m+i}}{\theta_{n-m}} \right) \right). \\ \mathbf{b}_i &\triangleq [1, \theta_{n-m+i}, \dots, \theta_{n-m+i}^{n-m-1}]^t. \end{aligned}$$

We are now in position to state the following result

Proposition 2. *Let $p = \min(m, n-m)$, then for $1 \leq r \leq p$ we have*

$$\begin{aligned} &\lim_{s \rightarrow 0} \mathcal{M}_1(s-r+1) \\ &= L \sum_{j=1}^r \sum_{i=1}^m \mathcal{D}(i, j) \frac{(-1)^{r-j}}{(r-j)!} \mathbf{b}_i^t \Psi^{-1} \mathbf{D}_i \mathbf{a}_j. \end{aligned}$$

Proof: See Appendix B for a detailed proof. ■

Combining the findings of the above propositions, we finally obtain the following result

Theorem 1. *Let $p = \min(m, n-m)$, then for $1 \leq r \leq p$, we have*

$$\mu_{\Lambda}(-r) = L \sum_{j=1}^r \sum_{i=1}^m \mathcal{D}(i, j) \frac{(-1)^{r-j}}{(r-j)!} \mathbf{b}_i^t \Psi^{-1} \mathbf{D}_i \mathbf{a}_j.$$

Remark 1. *Without loss of generality, we can easily show that matrix Λ can be considered as diagonal with diagonal elements $\theta_1, \dots, \theta_n$. This can be seen from the eigendecomposition of Λ as follows*

$$\Lambda = \mathbf{U}^* \mathbf{D} \mathbf{U}, \quad (7)$$

where $\mathbf{D} = \text{diag}(\theta_1, \theta_2, \dots, \theta_n)$ and \mathbf{U} is a unitary matrix, i.e. $\mathbf{U}^* \mathbf{U} = \mathbf{U} \mathbf{U}^* = \mathbf{I}_n$. Then,

$$\begin{aligned} \mathbf{S} &= \mathbf{H}^* \mathbf{U}^* \mathbf{D} \mathbf{U} \mathbf{H} \\ &= (\mathbf{U} \mathbf{H})^* \mathbf{D} (\mathbf{U} \mathbf{H}) \\ &= \mathbf{G}^* \mathbf{D} \mathbf{G}, \end{aligned}$$

where $\mathbf{G} = \mathbf{U} \mathbf{H}$. Since the wishart distribution is unitarily invariant, \mathbf{G} has the same distribution as \mathbf{H} . Therefore,

$$\mu_{\Lambda}(r) = \frac{1}{m} \text{tr}(\mathbb{E}_{\mathbf{G}}\{(\mathbf{G}^* \mathbf{D} \mathbf{G})^r\}), \quad r \in \mathbb{Z}. \quad (8)$$

B. Asymptotic Inverse Moments

It is well-known from standard results on random matrix theory that moments of Gram random matrices, can be well-approximated, when properly scaled and for m and n large enough, by deterministic quantities. However, the derivation of these deterministic approximations differs from the exact approach in several respects, namely it does not rely on the same tool of the Mellin transform and does not necessarily yield simple closed-form expressions for any high order moment. As a matter of fact, it is shown in [4] that except the special case of Λ coinciding with the identity matrix, the computation of higher positive order moments has to be performed iteratively. This also holds for the case of asymptotic inverse moments, which can be derived using the same approach as in [4]. This represents the main goal of this section, for which details will be provided for sake of completeness. The obtained asymptotic moments will be compared later with the exact ones derived in the previous subsection.

Prior to stating the main algorithm leading to the asymptotic inverse moments, we shall first review the following results from random matrix theory.

Definition 2. (*Empirical Spectral Distribution*) Let $\mathbf{A} \in \mathbb{C}^{m \times m}$ be a Hermitian matrix with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_m$. The empirical spectral distribution $F^{\mathbf{A}}$ of \mathbf{A} is defined as

$$F^{\mathbf{A}}(x) = \frac{1}{m} \sum_{i=1}^m \mathbf{1}(\lambda \leq x). \quad (9)$$

Working directly on the empirical distribution function $F^{\mathbf{A}}$ is in general a tedious task. Instead, a characterization of its Stieltjes transform is often considered. The Stieltjes transform corresponding to the empirical distribution $F^{\mathbf{A}}$ is defined as:

Definition 3. (*Stieltjes Transform*) For a Hermitian matrix \mathbf{A} , the Stieltjes transform is defined as

$$\begin{aligned} \hat{m}_{\mathbf{A}}(z) &\triangleq \int \frac{1}{\lambda - z} dF^{\mathbf{A}}(\lambda) \\ &= \frac{1}{m} \text{tr}(\mathbf{A} - z \mathbf{I}_m)^{-1}. \end{aligned} \quad (10)$$

From this definition, it is easy to prove that

$$\left(\frac{\partial^k \hat{m}_{\mathbf{A}}(z)}{\partial z^k} \right)_{z=0} = \frac{k!}{m} \text{tr}(\mathbf{A}^{-(k+1)}). \quad (11)$$

Let \mathbf{D} be the diagonal matrix defined in (7). Then, the Stieltjes Transform (ST) of the empirical measure of $\frac{1}{m}\mathbf{S}$ converges to a deterministic measure whose ST $\underline{m}(z)$ is the solution of the following fixed point equation [11]:

$$\underline{m}(z) = \frac{1}{-z + \frac{1}{m} \sum_{k=1}^n \frac{[\mathbf{D}]_{k,k}}{1 + [\mathbf{D}]_{k,k} \underline{m}(z)}}. \quad (12)$$

Denote by $\underline{m}^{(r)}$ the r -th derivative of $\underline{m}(z)$ at $z = 0$. Along the same arguments as in [4], we can prove that $\underline{m}^{(r)}$ is a consistent estimate of $r!m^r \text{tr}(\mathbf{H}^* \mathbf{\Lambda} \mathbf{H})^{-(r+1)}$. This suggests in particular estimating the scaled inverse moments $\frac{1}{m} \text{tr}((\frac{1}{m}\mathbf{S})^{-r})$ by $\frac{1}{r!} \underline{m}^{(r)}$. Note that, in practice, the computation of the higher derivatives $\underline{m}^{(r)}$ requires the computation of the lower order derivatives $\underline{m}^{(k)}$ for $k < r$. This is illustrated by the following Theorem

Theorem 2. *Let $p \geq 1$ and $f_k(z) = -\frac{1}{1 + [\mathbf{D}]_{k,k} \underline{m}(z)}$. Denote by $f_k^{(p)}$ the p -th derivative of $f_k(z)$ at $z = 0$. Then, the following relations hold true:*

$$pm^{(p-1)} + \frac{m^{(p)}}{m} \sum_{k=1}^n \frac{[\mathbf{D}]_{k,k} f_k^{(0)}}{1 + [\mathbf{D}]_{k,k} m(0)} + \frac{1}{m} \sum_{k=1}^n \sum_{l=1}^{p-1} \binom{p}{l} \frac{[\mathbf{D}]_{k,k} m^{(l)} f_k^{(p-l)}}{1 + [\mathbf{D}]_{k,k} m(0)} = 0, \quad (13)$$

$$f_k^{(p)} + \frac{[\mathbf{D}]_{k,k} m^{(p)} f_k^{(0)}}{1 + [\mathbf{D}]_{k,k} m(0)} + \sum_{l=1}^{p-1} \binom{p}{l} \frac{[\mathbf{D}]_{k,k} m^{(l)} f_k^{(p-l)}}{1 + [\mathbf{D}]_{k,k} m(0)} = 0. \quad (14)$$

Proof: See Appendix C for detailed proof. ■

Based on the previous theorem, an algorithm can be provided to recursively compute the higher order derivatives, $\underline{m}^{(k)}$. These values will thus immediately serve to compute the deterministic approximations for the moments.

Algorithm 1 Asymptotic inverse moments computation

- 1: Compute $\underline{m}(0)$ using (12)
 - 2: Compute $f_k(0) = -\frac{1}{1 + [\mathbf{D}]_{k,k} \underline{m}(0)}$
 - 3: **for** $i = 1 \rightarrow p$ **do**
 - 4: compute $\underline{m}^{(i)}$ using (13)
 - 5: compute $f_k^{(i)}$ using (14)
 - 6: **end for**
-

C. Numerical Examples

We validate the theoretical result stated in Theorem 1 for different values of m and n . In particular we compare $\mu_{\mathbf{\Lambda}}(-r)$ with the normalized asymptotic moments $\frac{m^{(r)}}{r!m^{r+1}}$ and the empirical moments obtained by Monte Carlo simulations.

We start by verifying the result in Figure 1, in the case where $\mathbf{\Lambda}$ is a correlation matrix having the following structure

$$[\mathbf{\Lambda}]_{i,j} = J_0\left(\pi|i-j|^2\right), \quad (15)$$

where $J_0(\cdot)$ is the zero-order Bessel function of the first kind. This kind of matrices is used to model the correlation between transmit antennas in a dense scattering environment [7,12]. For simulations, we set $m = 4$ and vary n such that $n > m$.

In Figure 2, we compare the same quantities in the case where $\mathbf{\Lambda}$ is a random positive definite matrix generated as follows²

$$\mathbf{\Lambda} = \mathbf{I}_n + \mathbf{W}^* \mathbf{W}, \quad (16)$$

where \mathbf{W} is a $(n \times n)$ matrix with i.i.d zero-mean unit variance complex Gaussian random entries.

Figures 1 and 2 show a perfect match between the exact moments derived in Theorem 1 and the empirical moments obtained from Monte Carlo simulations. It is worth mentioning that as n increases further, the exact approach might present some numerical issues, mainly caused by the high instability of computing the inverse of matrix Ψ . On the other hand, we can see from Figures 1 and 2 that the asymptotic approach improves in accuracy with n getting higher values, but with a slower pace as the moment order r increases. In other words, for the same n , the asymptotic approach is more accurate for the approximation of the first order moment than it is for the higher order moments.

III. APPLICATIONS OF THE INVERSE MOMENTS

The computation of inverse moments of one-sided correlated Gram matrices is paramount to many applications of signal processing. For sake of illustration, we will discuss in the sequel applications of our results to the fields of linear estimation and covariance matrix estimation.

A. Linear Estimation

The problem of estimating an unknown signal from a sequence of observations has been widely studied in the

²We use such matrices to make sure that the obtained results are applicable for broad class of positive definite matrices.

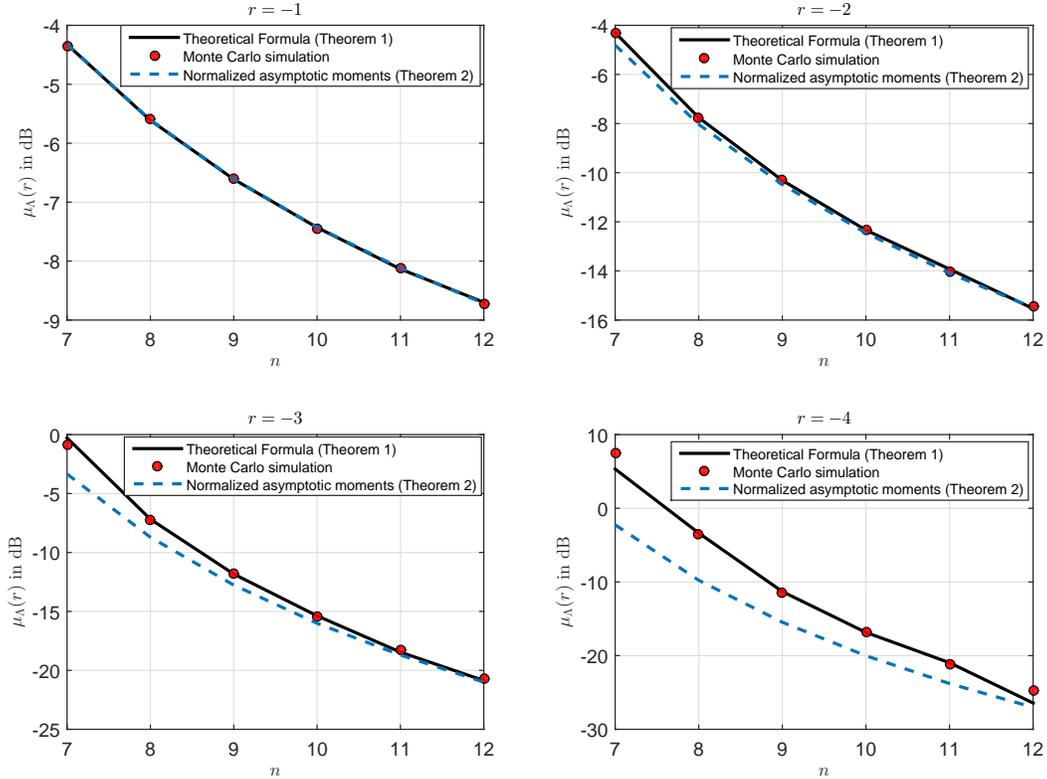


Figure 1. Inverse moments for Λ defined as in (15): A comparison between theoretical result (Theorem 1), normalized asymptotic moments (Theorem 2) and Monte Carlo simulations (10^5 realizations).

literature [13]–[15] and can be solved if joint statistics³ or cross correlations of the unknown signal and the observations vector are available. Here, linear models are a special case where, the input and the output are linearly related according to

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}, \quad (17)$$

where $\mathbf{y} \in \mathbb{C}^{n \times 1}$ is the observations vector, $\mathbf{H} \in \mathbb{C}^{n \times m}$ the channel matrix, $\mathbf{x} \in \mathbb{C}^{m \times 1}$ the unknown signal with covariance matrix Σ_x and $\mathbf{z} \in \mathbb{C}^{n \times 1}$ the noise vector with covariance matrix Σ_z . We consider the case of a block fading model in which the channel matrix \mathbf{H} is taken to be constant over a given time interval, called block, and assumed to change independently from each block to the next. As stated earlier, in order to recover \mathbf{x} , joint statistics are required. However, acquiring joint statistics is generally a difficult task either because of the unknown nature of the signal or simply because of the unavailability of the statistics. To overcome this issue, linear estimators can be viewed as a good alternative. They are merely based on applying a linear

transformation to the observation vector. Obviously, this is a sub-optimal strategy in regards of the minimization of the *mean square error*, but it is more tractable and permits to explicitly analyze performances, depending on the information available about the signal and the noise statistics.

In what follows, we make the following assumptions:

- \mathbf{H} is a $(n \times m)$ matrix with *i.i.d* complex zero mean unit variance Gaussian random entries,
- \mathbf{z} is a $(n \times 1)$ zero mean additive Gaussian noise with covariance matrix $\Sigma_z = \mathbb{E}\{\mathbf{z}\mathbf{z}^*\}$, i.e. $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}_n, \Sigma_z)$.

Let $\hat{\mathbf{x}}$ be an estimate of \mathbf{x} obtained using some channel estimation method. For a given channel condition, the performance of this method is assessed using the covariance of the error vector $\Sigma_e \triangleq \mathbb{E}[(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x}^H - \hat{\mathbf{x}}^H)]$ where the expectation is taken over \mathbf{z} and \mathbf{x} . In Table I, we recall the expressions of the covariance error of some common estimation techniques. It is worth mentioning that the noise and the signal vectors vary on a different scale from that of \mathbf{H} . While vectors \mathbf{x} and \mathbf{z} take different values for each observation, the channel matrix \mathbf{H} is taken to be constant over each block and assumed to

³We mean by joint statistics the joint PDF that relates the observed signal to the unknown signal which is out of reach in general.

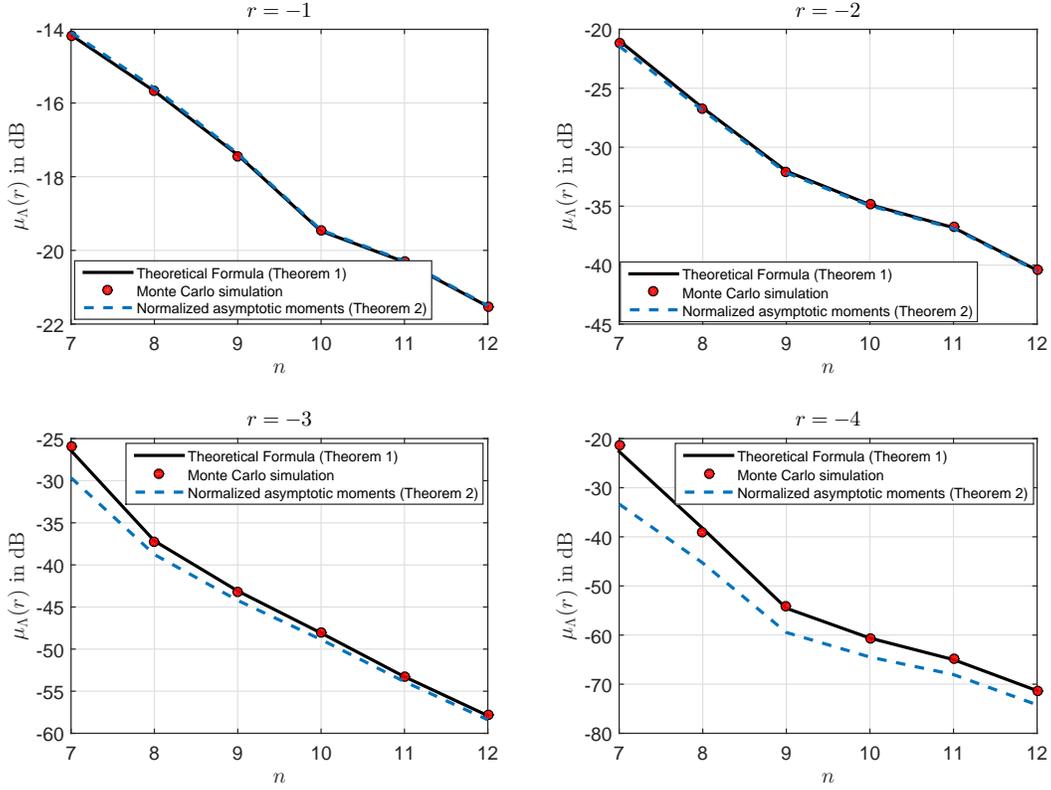


Figure 2. Inverse moments for Λ defined as in (16): A comparison between theoretical result (Theorem 1), normalized asymptotic moments (Theorem 2) and Monte Carlo simulations (10^5 realizations).

Table I
LINEAR ESTIMATION TECHNIQUES DEPENDING ON THE AVAILABLE STATISTICS

Linear estimator	Required Statistics	Estimated signal, $\hat{\mathbf{x}}$	Error covariance matrix, $\mathbb{E}(\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^*$
LS	\emptyset	$(\mathbf{H}^* \mathbf{H})^{-1} \mathbf{H}^* \mathbf{y}$	$(\mathbf{H}^* \mathbf{H})^{-1}$
BLUE	Σ_z	$(\mathbf{H}^* \Sigma_z^{-1} \mathbf{H})^{-1} \mathbf{H}^* \Sigma_z^{-1} \mathbf{y}$	$(\mathbf{H}^* \Sigma_z^{-1} \mathbf{H})^{-1}$
LMMSE	Σ_z, Σ_x	$(\Sigma_x^{-1} + \mathbf{H}^* \Sigma_z^{-1} \mathbf{H})^{-1} \mathbf{H}^* \Sigma_z^{-1} \mathbf{y}$	$(\Sigma_x^{-1} + \mathbf{H}^* \Sigma_z^{-1} \mathbf{H})^{-1}$

change independently from each block to the next [16]. Thus, it is most informative in these circumstances to focus on average performances by taking the expectation over \mathbf{H} . This amounts to considering the expectation over \mathbf{H} of Σ_e . In the sequel, we will derive closed-form expressions for the trace of $\mathbb{E}(\Sigma_e)$ when BLUE and LMMSE estimators (See Table I) are used.

1) *An Exact expression for The BLUE Average Estimation Error:* Let $n > m$ and consider the same linear system as in (17). With the the noise covariance matrix Σ_z at hand, the best linear unbiased estimator (BLUE)

[14] recovers \mathbf{x} as:

$$\begin{aligned} \hat{\mathbf{x}}_{blue} &= (\mathbf{H}^* \Sigma_z^{-1} \mathbf{H})^{-1} \mathbf{H}^* \Sigma_z^{-1} \mathbf{y} \\ &= \mathbf{x} + (\mathbf{H}^* \Sigma_z^{-1} \mathbf{H})^{-1} \mathbf{H}^* \Sigma_z^{-1} \mathbf{z} \\ &= \mathbf{x} + \mathbf{e}_{blue}, \end{aligned} \quad (18)$$

where $\mathbf{e}_{blue} = (\mathbf{H}^* \Sigma_z^{-1} \mathbf{H})^{-1} \mathbf{H}^* \Sigma_z^{-1} \mathbf{z}$ is the residual error after applying the BLUE. We denote by $\Sigma_{e,blue} = \mathbb{E}\{\mathbf{e}_{blue} \mathbf{e}_{blue}^*\}$ the covariance matrix of \mathbf{e}_{blue} , then $\Sigma_{e,blue} = (\mathbf{H}^* \Sigma_z^{-1} \mathbf{H})^{-1}$. Note that the BLUE can also be used as a refinement tool after compressed sensing (CS) recovery as demonstrated in [8,9,17,18].

Lemma 3. *Based on the result of Theorem 1, the BLUE average estimation error can be expressed as a function*

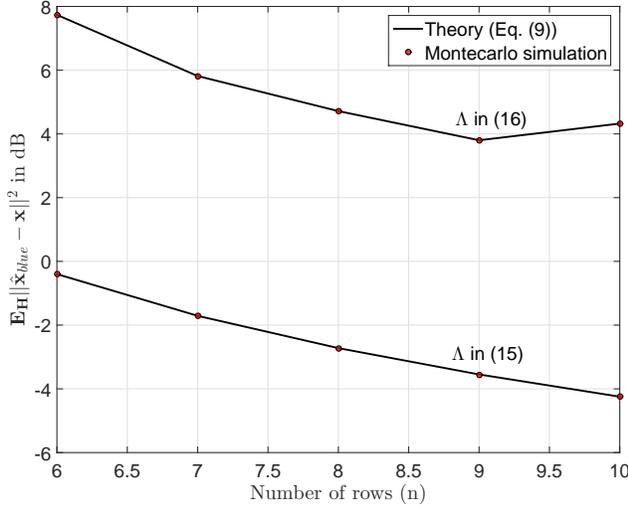


Figure 3. BLUE average estimation error for $m = 3$: Monte Carlo simulation (10^4 realizations) versus theory (Theorem 1)

of the inverse moments of one sided correlated Gram matrix as follows

$$\begin{aligned} \mathbb{E}_{\mathbf{H}}\{\|\hat{\mathbf{x}}_{blue} - \mathbf{x}\|^2\} &= \mathbb{E}_{\mathbf{H}}\text{tr}(\boldsymbol{\Sigma}_{e,blue}) \\ &= \mathbb{E}_{\mathbf{H}}\text{tr}\left(\left(\mathbf{H}^* \boldsymbol{\Sigma}_z^{-1} \mathbf{H}\right)^{-1}\right) \quad (19) \\ &= m \mu_{\Lambda}(-1), \end{aligned}$$

where $\Lambda = \boldsymbol{\Sigma}_z^{-1}$.

For simulation purposes, we set $m = 3$, and consider Λ as in (15) and (16). Then, we compare the empirical average estimation error using Monte Carlo simulation for different values of n with the theoretical result derived in Theorem 1. As shown in Figure 3, the theoretical performance exactly matches the actual performance of the BLUE in terms of average estimation error for both models of Λ in (15) and (16).

2) *Approximation of the LMMSE average estimation error:* Consider the linear system as in (17), where we assume additionally that the covariance matrix of the unknown signal \mathbf{x} is known and given by $\boldsymbol{\Sigma}_x$. The LMMSE estimate of \mathbf{x} is thus given by:

$$\hat{\mathbf{x}}_{lmmse} = \left(\boldsymbol{\Sigma}_x^{-1} + \mathbf{H}^* \boldsymbol{\Sigma}_z^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^* \boldsymbol{\Sigma}_z^{-1} \mathbf{y}. \quad (20)$$

Consequently, the estimation error can be calculated as

$$\mathbf{e}_{lmmse} = \left(\boldsymbol{\Sigma}_x^{-1} + \mathbf{H}^* \boldsymbol{\Sigma}_z^{-1} \mathbf{H}\right)^{-1} \mathbf{H}^* \boldsymbol{\Sigma}_z^{-1} \mathbf{z}. \quad (21)$$

By standard computations and based on the result derived in [14], the error covariance matrix is given by

$$\boldsymbol{\Sigma}_{e,lmmse} = \left(\boldsymbol{\Sigma}_x^{-1} + \mathbf{H}^* \boldsymbol{\Sigma}_z^{-1} \mathbf{H}\right)^{-1}. \quad (22)$$

Therefore, the average estimation error for the LMMSE estimator is given by

$$\begin{aligned} \mathbb{E}_{\mathbf{H}}\{\|\hat{\mathbf{x}}_{lmmse} - \mathbf{x}\|^2\} &= \mathbb{E}_{\mathbf{H}}\text{tr}(\boldsymbol{\Sigma}_{e,lmmse}) \\ &= \mathbb{E}_{\mathbf{H}}\text{tr}\left(\left(\boldsymbol{\Sigma}_x^{-1} + \mathbf{H}^* \boldsymbol{\Sigma}_z^{-1} \mathbf{H}\right)^{-1}\right). \end{aligned} \quad (23)$$

Evaluating the LMMSE average estimation error is in general very difficult, and upper bounds on the mean square error are usually derived [19]. However, for the simple case where $\boldsymbol{\Sigma}_x = \sigma_x^2 \mathbf{I}_n$, it is possible to obtain an approximation depending on the value of σ_x^2 . In the following theorem, we provide tight approximations for the LMMSE average estimation error for the cases: $\sigma_x^2 \gg 1$ (High SNR regime) and $\sigma_x^2 \ll 1$ (Low SNR regime).

Theorem 3. Let $\Lambda = \boldsymbol{\Sigma}_z^{-1}$. Then, the LMMSE average estimation error at both the high SNR regime ($\sigma_x^2 \gg 1$) and the low SNR regime ($\sigma_x^2 \ll 1$) is given by

1) *High SNR regime:*

$$\begin{aligned} \mathbb{E}_{\mathbf{H}}\{\|\hat{\mathbf{x}}_{lmmse} - \mathbf{x}\|^2\} &= m \sum_{k=0}^l \frac{(-1)^k}{\sigma_x^{2k}} \mu_{\Lambda}(-k-1) + o(\sigma_x^{-2r}), \end{aligned} \quad (24)$$

where $l \leq p - 1$ with $p = \min(m, n - m)$.

2) *Low SNR regime:*

$$\mathbb{E}_{\mathbf{H}}\{\|\hat{\mathbf{x}}_{lmmse} - \mathbf{x}\|^2\} = m \sum_{k=0}^{\infty} (-1)^k \sigma_x^{2k+2} \mu_{\Lambda}(k). \quad (25)$$

Proof: See Appendix D for a Proof. ■

To validate the approximations derived in Theorem 3, we set $m = 3$ and $n = 10$ and apply the obtained results for both correlation models in (15) and (16). For that in Figures 4 and 5, we compare the mean square error of the LMMSE using Monte Carlo simulations with the approximations derived in Theorem 3. As shown in Figures 4 and 5, the approximation is quite tight in both high and low SNR regimes and covers almost the whole SNR range.

B. Sample Correlation Matrix (SCM)

Estimation of covariance matrices is of fundamental importance to several adaptive processing applications. Assume that the measurements are arranged into an input vector $\mathbf{u} \in \mathbb{C}^{m \times 1}$ also called the observation vector. If the input process is zero-mean, its covariance matrix is given by

$$\mathbf{R} \triangleq \mathbb{E}\{\mathbf{u}\mathbf{u}^*\}, \quad (26)$$

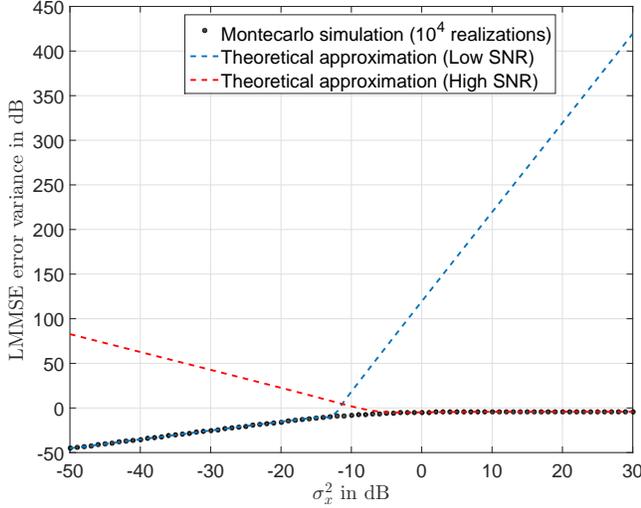


Figure 4. LMMSE mean square error with Σ_z as in (15): Monte Carlo simulation versus theoretical approximation for the low and high SNR regimes.

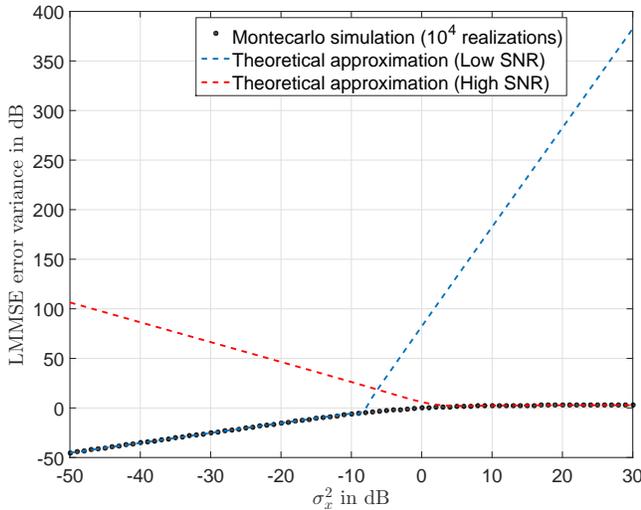


Figure 5. LMMSE mean square error with Σ_z as in (16): Monte Carlo simulation versus theoretical approximation for the low and high SNR regimes.

where the expectation is taken over all realization of the input. The covariance matrix \mathbf{R} is usually unknown, and thus has to be estimated. Assuming the input process to be ergodic, the covariance matrix can be estimated via time averaging. A well-known estimator is the sample correlation matrix (SCM) which is given by [20, eq. (1.2), page 16]

$$\hat{\mathbf{R}}(n) = \frac{1}{n} \sum_{k=1}^n \mathbf{u}(k) \mathbf{u}^*(k), \quad (27)$$

This is called rectangularly windowed SCM, where $\mathbf{u}(k)$ is the input vector at discrete time k and n is the length of the observation window. When the observations are Gaussian distributed, the SCM is the maximum likelihood (ML) estimator of the correlation matrix [21, eq. (7.9), page 715]. Moreover, for a fixed and finite input size m , as the window size $n \rightarrow \infty$, the SCM converges to the input correlation matrix [22, eq. (1.1), page 2], in the sense that

$$\left\| \mathbf{R} - \hat{\mathbf{R}}(n) \right\| \rightarrow 0, \quad a.s. \quad (28)$$

However, the number of measurement is usually finite for practical applications. Thus, it is for a practical interest to evaluate the performance of the SCM when the window size is finite. In order to measure the accuracy of the estimator, we define the average *loss* as the average distance between the input correlation matrix and its estimated version using SCM for a given window size n [23]:

$$Loss(n) \triangleq \mathbb{E} \left\| \mathbf{R}^{\frac{1}{2}} \hat{\mathbf{R}}^{-1}(n) \mathbf{R}^{\frac{1}{2}} - \mathbf{I}_m \right\|_F^2, \quad (29)$$

where $\mathbf{R}^{\frac{1}{2}}$ is a positive semi-definite square root of \mathbf{R} and $\|\cdot\|_F$ is the Forbenius norm of a matrix.

In order to emphasize some measurements relevant for the estimation of the correlation matrix, an exponentially weighted SCM can be used and it is given by [20, eq. (2.62), page 56]

$$\hat{\mathbf{R}}(n) = (1 - \lambda) \sum_{k=1}^n \lambda^{n-k} \mathbf{u}(k) \mathbf{u}^*(k), \quad (30)$$

where λ^{n-k} is the weight associated with the measurement vector at time instant k , the coefficient λ being the forgetting factor. In the case where $\mathbf{u}(k)$ is modeled as a colored process $\mathbf{u}(k) = \mathbf{R}^{\frac{1}{2}} \mathbf{x}(k)$, where $\mathbf{x}(k) \in \mathbb{C}^{m \times 1}$ is a vector of *i.i.d* Gaussian zero mean, unit variance entries, the SCM can be written in a matrix form as

$$\hat{\mathbf{R}}(n) = \mathbf{R}^{\frac{1}{2}} \mathbf{X} \mathbf{\Lambda}(n) \mathbf{X}^* \mathbf{R}^{\frac{1}{2}}, \quad (31)$$

where \mathbf{X} is $m \times n$ matrix whose k th column is $\mathbf{x}(k)$ and $\mathbf{\Lambda}(n) = (1 - \lambda) \text{diag}(\lambda^{n-1}, \lambda^{n-2}, \dots, 1)$. In the following lemma, we show that we can derive a closed-form expression for the loss function defined in (29).

Lemma 4. *Let $\mathbf{S}_n = \mathbf{X} \mathbf{\Lambda}(n) \mathbf{X}^*$. Then, the loss can be expressed as a function of inverse moments of \mathbf{S}_n as*

$$Loss(n) = m \left(1 + \mu_{\mathbf{\Lambda}(n)}(-2) - 2\mu_{\mathbf{\Lambda}(n)}(-1) \right). \quad (32)$$

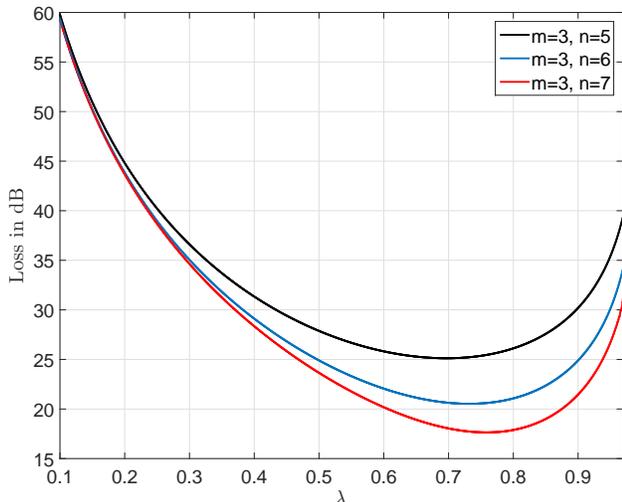


Figure 6. The estimation loss as a function of λ : Theoretical expression in (33).

Proof:

$$\begin{aligned}
Loss(n) &= \mathbb{E} \left\| \mathbf{S}_n^{-1} - \mathbf{I}_m \right\|_F^2 \\
&= \mathbb{E} \text{tr} \left[\mathbf{S}_n^{-1} - \mathbf{I}_m \right]^* \left[\mathbf{S}_n^{-1} - \mathbf{I}_m \right] \\
&= \mathbb{E} \text{tr} \left[\mathbf{S}_n^{-2} - 2\mathbf{S}_n^{-1} + \mathbf{I}_m \right] \\
&= m + \mathbb{E} \text{tr} \left[\mathbf{S}_n^{-2} \right] - 2\mathbb{E} \text{tr} \left[\mathbf{S}_n^{-1} \right] \\
&= m + m\mu_{\Lambda(n)}(-2) - 2m\mu_{\Lambda(n)}(-1) \\
&= m \left(1 + \mu_{\Lambda(n)}(-2) - 2\mu_{\Lambda(n)}(-1) \right).
\end{aligned} \tag{33}$$

One interesting problem is to find the optimal $\lambda \in (0, 1)$ denoted by λ^* that minimizes the loss function. This can be performed by evaluating the loss function with respect to $\lambda \in (0, 1)$ and then picking the λ that gives the lowest loss. To evaluate the loss function, one can resort to Monte Carlo simulations. This would involve high complexity since they should be repeated for each value of λ . The use of the provided closed-form expression represent thus a valuable alternative being at the same time accurate and easier to implement. For $m = 3$ and $n = 10$, we plot the estimation loss as a function of λ using the theoretical expression derived in (33). As shown in Figure 6, for all cases a minimum exist and thus the performance can be optimized accordingly. To obtain the optimal λ , one can resort to a simple one dimensional optimization using some built-in functions in Matlab such as *fminbnd*.

IV. CONCLUSION

In this paper, we derived closed form expressions for the inverse order moments of general Gram matrices with

one sided correlation. Based on this formula, the exact average estimation error of the BLUE estimator has been derived and an accurate approximation for the LMMSE average estimation error was proposed in both high and low SNR regimes. Additionally, we have shown that our results can be used to evaluate the accuracy of covariance matrix estimates.

APPENDIX A

PROOF OF PROPOSITION 1

As stated in section II, $\mathcal{M}_2(s)$ is given by:

$$\begin{aligned}
\mathcal{M}_2(s) &= L \sum_{j=r+1}^m \sum_{i=1}^m \mathcal{D}(i, j) \Gamma(s + j - 1) \left(\theta_{n-m+i}^{n-m+s+j-2} \right. \\
&\quad \left. - \sum_{l=1}^{n-m} \sum_{k=1}^{n-m} [\Psi^{-1}]_{k,l} \theta_l^{n-m+s+j-2} \theta_{n-m+i}^{k-1} \right).
\end{aligned}$$

The handling of $\mathcal{M}_2(s)$ does not pose any difficulty, since the Gamma function being applied to non-negative arguments has finite values. Interestingly, we can show that this term turns out to be equal to zero as s goes to zero. To this end, note that

$$\begin{aligned}
&\lim_{s \rightarrow 0} \mathcal{M}_2(s - r + 1) \\
&= L \sum_{j=r+1}^m \sum_{i=1}^m \mathcal{D}(i, j) \Gamma(-r + j) \\
&\quad \times \left(\theta_{n-m+i}^{n-m-r+j-1} - \sum_{l=1}^{n-m} \sum_{k=1}^{n-m} [\Psi^{-1}]_{k,l} \theta_l^{n-m-r+j-1} \theta_{n-m+i}^{k-1} \right) \\
&= L \sum_{j=r+1}^m \sum_{i=1}^m [\mathcal{D}]_{i,j} [\mathcal{C}]_{i,j-r} \\
&= L \sum_{j=r+1}^m [\mathcal{D}^t \mathcal{C}]_{j,j-r},
\end{aligned}$$

where \mathcal{D} and \mathcal{C} are as defined in Lemma 1. Since \mathcal{D} is the cofactor of \mathcal{C} , then $\mathcal{D}^t \mathcal{C} = \det(\mathcal{C}) \mathbf{I}_m$. Therefore, $[\mathcal{D}^t \mathcal{C}]_{j,j-r} = 0$ for $j = r + 1, \dots, m$.

APPENDIX B

PROOF OF PROPOSITION 2

The handling of $\mathcal{M}_1(s - r + 1)$ is delicate because it involves evaluation of the Gamma function at negative integers. Hopefully, a compensation effect occurs due to the multiplicative term in front of the Gamma function. The proof relies on a divide and conquer strategy that consists of decomposing $\mathcal{M}_1(s - r + 1)$ into a sum of

terms and then evaluating each term separately. To this end, we need to introduce the following notation.

$$\begin{aligned}
\mathbf{\Psi}_s &\triangleq \begin{bmatrix} \theta_1^s & \theta_1^{1+s} & \cdots & \theta_1^{n-m+s-1} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{n-m}^s & \theta_{n-m}^{1+s} & \cdots & \theta_{n-m}^{n-m+s-1} \end{bmatrix}. \\
\mathbf{a}_{s,j} &\triangleq \begin{bmatrix} \theta_1^{n-m+s-r+j-1}, \theta_2^{n-m+s-r+j-1}, \\ \cdots, \theta_{n-m}^{n-m+s-r+j-1} \end{bmatrix}^t. \\
\mathbf{b}_{s,i} &\triangleq [\theta_{n-m+i}^s, \theta_{n-m+i}^{1+s}, \cdots, \theta_{n-m+i}^{n-m+s-1}]^t. \\
\mathbf{b}_i &\triangleq [1, \theta_{n-m+i}, \cdots, \theta_{n-m+i}^{n-m-1}]^t. \\
\mathbf{e}_k &\triangleq [\mathbf{0}_{1 \times (n-m-k-1)}, 1, \mathbf{0}_{1 \times k}]^t, \\
k &= 0, \cdots, n-m-1,
\end{aligned} \tag{34}$$

where $\mathbf{0}_{p \times k}$ is the zero matrix or null matrix of dimension $p \times k$. Using the previously defined variables, we can rewrite $\mathcal{M}_1(s-r+1)$ as in (35) (on top of the next page).

The first term in equation (35) is equal to zero. This can be seen by noticing that $\mathbf{\Psi}_s \mathbf{e}_{r-j} = \mathbf{a}_{s,j}$ and $\mathbf{b}_{s,i}^t \mathbf{e}_{r-j} = \theta_{n-m+i}^{n-m+s-r+j-1}$. Thus, $\mathbf{\Psi}_s^{-1} \mathbf{a}_{s,j} = \mathbf{e}_{r-j}$ and consequently $\mathbf{b}_{s,i}^t \mathbf{\Psi}_s^{-1} \mathbf{a}_{s,j} = \theta_{n-m+i}^{n-m+s-r+j-1}$. It remains thus to deal with the last two terms. Using a Taylor approximation of $\mathbf{b}_{s,i}$ as s approaching 0, we have

$$\begin{aligned}
\mathbf{b}_{s,i} - \mathbf{b}_i &= s \begin{bmatrix} \log(\theta_{n-m+i}), \theta_{n-m+i} \log(\theta_{n-m+i}), \\ \cdots, \theta_{n-m+i}^{n-m-1} \log(\theta_{n-m+i}) \end{bmatrix}^t + o(s) \\
&= s \log(\theta_{n-m+i}) \mathbf{b}_i + o(s),
\end{aligned} \tag{36}$$

where the relation $f(s) = o(g(s))$ is equivalent to

$$\lim_{s \rightarrow 0} \frac{f(s)}{g(s)} = 0. \tag{37}$$

To deal with the Gamma function evaluated at non positive integers, we rely on the result of the following lemma.

Lemma 5. [24] *For non positive arguments $-k$, $k = 0, 1, 2, \cdots$, the Gamma function can be evaluated as*

$$\lim_{s \rightarrow 0} \frac{\Gamma(s-k)}{\Gamma(s)} = \frac{(-1)^k}{k!}, \tag{38}$$

where $\Gamma(s) = \frac{1}{s} + o(s)$ as s approaches 0.

Thus, $\Gamma(s-r+j) \underset{s \rightarrow 0}{=} \frac{(-1)^{r-j}}{s(r-j)!} + o(s)$. Therefore, as s approaches 0, we have

$$\begin{aligned}
&\Gamma(s-r+j) (\mathbf{b}_{s,i}^t - \mathbf{b}_i^t) \mathbf{\Psi}_s^{-1} \mathbf{a}_{s,j} \\
&= \frac{(-1)^{r-j} \log(\theta_{n-m+i})}{(r-j)!} \mathbf{b}_i^t \mathbf{\Psi}^{-1} \mathbf{a}_j + o(s),
\end{aligned}$$

Finally, to deal with the last term, we use the following resolvent identity :

$$\mathbf{\Psi}_s^{-1} - \mathbf{\Psi}^{-1} = \mathbf{\Psi}_s^{-1} (\mathbf{\Psi} - \mathbf{\Psi}_s) \mathbf{\Psi}^{-1}.$$

We also make use of the fact that as s approaches 0

$$(\mathbf{\Psi} - \mathbf{\Psi}_s) \underset{s \rightarrow 0}{=} -s \tilde{\mathbf{\Psi}} + o(s),$$

where

$$\tilde{\mathbf{\Psi}} = \mathbf{\Phi} \mathbf{\Psi},$$

with $\mathbf{\Phi} = \text{diag}(\log(\theta_1), \log(\theta_2), \cdots, \log(\theta_{n-m}))$. Thus, as s approaches 0, we have

$$\begin{aligned}
&\Gamma(s-r+j) \mathbf{b}_i^t (\mathbf{\Psi}_s^{-1} - \mathbf{\Psi}^{-1}) \mathbf{a}_{s,j} \\
&= \frac{(-1)^{r+1-j}}{(r-j)!} \mathbf{b}_i^t \mathbf{\Psi}^{-1} \tilde{\mathbf{\Psi}} \mathbf{\Psi}^{-1} \mathbf{a}_j + o(s).
\end{aligned}$$

Finally, we have the following limit

$$\begin{aligned}
&\lim_{s \rightarrow 0} \mathcal{M}_1(s-r+1) \\
&= L \sum_{j=1}^r \sum_{i=1}^m \mathcal{D}(i,j) \left[\frac{(-1)^{r-j} \log(\theta_{n-m+i})}{(r-j)!} \mathbf{b}_i^t \mathbf{\Psi}^{-1} \mathbf{a}_j \right. \\
&\quad \left. + \frac{(-1)^{r+1-j}}{(r-j)!} \mathbf{b}_i^t \mathbf{\Psi}^{-1} \tilde{\mathbf{\Psi}} \mathbf{\Psi}^{-1} \mathbf{a}_j \right].
\end{aligned}$$

This expression can be further simplified by noticing that $\tilde{\mathbf{\Psi}} \mathbf{\Psi}^{-1} = \mathbf{\Phi}$. Finally, we have

$$\begin{aligned}
&\lim_{s \rightarrow 0} \mathcal{M}_1(s-r+1) \\
&= L \sum_{j=1}^r \sum_{i=1}^m \mathcal{D}(i,j) \frac{(-1)^{r-j}}{(r-j)!} \mathbf{b}_i^t \mathbf{\Psi}^{-1} \left[\log(\theta_{n-m+i}) \mathbf{I}_{n-m} \right. \\
&\quad \left. - \mathbf{\Phi} \right] \mathbf{a}_j \\
&= L \sum_{j=1}^r \sum_{i=1}^m \mathcal{D}(i,j) \frac{(-1)^{r-j}}{(r-j)!} \mathbf{b}_i^t \mathbf{\Psi}^{-1} \mathbf{D}_i \mathbf{a}_j.
\end{aligned}$$

Thereby ending up the proof of the proposition.

$$\begin{aligned}
\mathcal{M}_1(s-r+1) &= L \sum_{j=1}^r \sum_{i=1}^m \mathcal{D}(i,j) \Gamma(s-r+j) \left(\theta_{n-m+i}^{n-m+s-r+j-1} - \sum_{l=1}^{n-m} \sum_{k=1}^{n-m} [\Psi^{-1}]_{k,l} \theta_l^{n-m+s-r+j-1} \theta_{n-m+i}^{k-1} \right) \\
&= L \sum_{j=1}^r \sum_{i=1}^m \mathcal{D}(i,j) \Gamma(s-r+j) \left(\theta_{n-m+i}^{n-m+s-r+j-1} - \mathbf{b}_i^t \Psi^{-1} \mathbf{a}_{s,j} \right) \\
&= L \sum_{j=1}^r \sum_{i=1}^m \mathcal{D}(i,j) \Gamma(s-r+j) \left(\theta_{n-m+i}^{n-m+s-r+j-1} - \mathbf{b}_i^t \Psi_s^{-1} \mathbf{a}_{s,j} \right) + L \sum_{j=1}^r \sum_{i=1}^m \mathcal{D}(i,j) \Gamma(s-r+j) \\
&\quad \times \mathbf{b}_i^t (\Psi_s^{-1} - \Psi^{-1}) \mathbf{a}_{s,j} \\
&= L \sum_{j=1}^r \sum_{i=1}^m \mathcal{D}(i,j) \Gamma(s-r+j) \left(\theta_{n-m+i}^{n-m+s-r+j-1} - \mathbf{b}_{s,i}^t \Psi_s^{-1} \mathbf{a}_{s,j} \right) + L \sum_{j=1}^r \sum_{i=1}^m \mathcal{D}(i,j) \Gamma(s-r+j) \\
&\quad \times (\mathbf{b}_{s,i}^t - \mathbf{b}_i^t) \Psi_s^{-1} \mathbf{a}_{s,j} + L \sum_{j=1}^r \sum_{i=1}^m \mathcal{D}(i,j) \Gamma(s-r+j) \mathbf{b}_i^t (\Psi_s^{-1} - \Psi^{-1}) \mathbf{a}_{s,j}
\end{aligned} \tag{35}$$

APPENDIX C

PROOF OF THEOREM 2

The existence and uniqueness of the solution to the fixed point equation (12) for $z < 0$ is a well known result of random matrix theory [11]. In the following, we will extend this result to the case of $z = 0$. From (12), $\underline{m}(z)$ satisfies

$$\underline{m}(z) = \frac{1}{-z + \frac{1}{m} \sum_{k=1}^n \frac{[\mathbf{D}]_{k,k}}{1 + [\mathbf{D}]_{k,k} \underline{m}(z)}}.$$

Substituting z by zero in the above equation, we get that $\underline{m}(0)$ satisfies:

$$\frac{1}{n} \sum_{k=1}^m \frac{[\mathbf{D}]_{k,k} \underline{m}(0)}{1 + [\mathbf{D}]_{k,k}} = 1,$$

or equivalently:

$$\sum_{k=1}^n \frac{1}{1 + [\mathbf{D}]_{k,k} \underline{m}(0)} = n - m.$$

The existence and uniqueness of $\underline{m}(0)$ follows then from a careful analysis of function g defined as:

$$g : \mathbb{R}^+ \rightarrow \mathbb{R} \\ x \mapsto g(x) = \sum_{k=1}^n \frac{1}{1 + [\mathbf{D}]_{k,k} x} - (n - m).$$

As a matter of fact, as $[\mathbf{D}]_{k,k} > 0, k = 1, \dots, n$, the derivative of g given by:

$$g'(x) = - \sum_{k=1}^n \frac{[\mathbf{D}]_{k,k}}{(1 + [\mathbf{D}]_{k,k} x)^2},$$

is negative. Thus, g is monotonically decreasing. It only remains now to evaluate g at $x = 0$ and its limit when $x \rightarrow \infty$:

$$\begin{aligned} g(0) &= m > 0. \\ g(\infty) &= -(n - m) < 0. \end{aligned}$$

Since g is continuous over its support, we conclude that the equation $g(x) = 0$ has a unique solution $\underline{m}(0)$.

We now proceed to the proof of Theorem 2. Let $f_k(z) = -\frac{1}{1 + [\mathbf{D}]_{k,k} \underline{m}(z)}$. Then, the above equation becomes:

$$-z \underline{m}(z) + \frac{n}{m} + \frac{1}{m} \sum_{k=1}^n f_k(z) = 1. \tag{39}$$

Taking the $p-1$ derivative of (39) and setting z to 0, we have:

$$p \underline{m}^{(p-1)}(0) = \frac{1}{m} \sum_{k=1}^n f_k^{(p)}(0). \tag{40}$$

On the other hand, functions $f_k(z)$ satisfy:

$$-f_k(z) - [\mathbf{D}]_{k,k} f_k(z) \underline{m}(z) = 1. \tag{41}$$

Taking the p -th derivative of equation (41), we get:

$$-f_k^{(p)} - [\mathbf{D}]_{k,k} \sum_{l=0}^p \binom{p}{l} f_k^{(p-l)} \underline{m}^{(l)}.$$

or equivalently,

$$f_k^{(p)} + \sum_{l=1}^p \binom{p}{l} \frac{[\mathbf{D}]_{k,k} \underline{m}^{(l)} f_k^{(p-l)}}{1 + [\mathbf{D}]_{k,k} \underline{m}(0)} = 0. \tag{42}$$

Hence, by separating the first term of the sum in (42), we obtain

$$f_k^{(p)} + \frac{[\mathbf{D}]_{k,k} \underline{m}^{(p)} f_k^{(0)}}{1 + [\mathbf{D}]_{k,k} \underline{m}^{(0)}} + \sum_{l=1}^{p-1} \binom{p}{l} \frac{[\mathbf{D}]_{k,k} \underline{m}^{(l)} f_k^{(p-l)}}{1 + [\mathbf{D}]_{k,k} \underline{m}^{(0)}} = 0.$$

Combining (40) and (42) and taking the sum over k of the above equation, we get:

$$\begin{aligned} p \underline{m}^{(p-1)} + \frac{\underline{m}^{(p)}}{m} \sum_{k=1}^n \frac{[\mathbf{D}]_{k,k} f_k^{(0)}}{1 + [\mathbf{D}]_{k,k} \underline{m}^{(0)}} \\ + \frac{1}{m} \sum_{k=1}^n \sum_{l=1}^{p-1} \binom{p}{l} \frac{[\mathbf{D}]_{k,k} \underline{m}^{(l)} f_k^{(p-l)}}{1 + [\mathbf{D}]_{k,k} \underline{m}^{(0)}} = 0, \end{aligned} \quad (43)$$

where $\underline{m}^{(0)}$ is the unique solution to the fixed point equation:

$$\underline{m}^{(0)} = \frac{1}{\frac{1}{m} \sum_{k=1}^n \frac{[\mathbf{D}]_{k,k}}{1 + [\mathbf{D}]_{k,k} \underline{m}^{(0)}}}. \quad (44)$$

This ends up the proof of the Theorem.

APPENDIX D

PROOF OF THEOREM 4

The proof of this theorem is based on a Taylor approximation of the LMMSE average error.

1) High SNR regime ($\sigma_x^2 \gg 1$):

As stated in equation (22)

$$\Sigma_{e,lmmse} = \left(\frac{1}{\sigma_x^2} \mathbf{I}_n + \mathbf{H}^* \Sigma_z^{-1} \mathbf{H} \right)^{-1}.$$

By setting $\Sigma_z^{-1} = \mathbf{\Lambda}$, we have

$$\Sigma_{e,lmmse} = \left(\frac{1}{\sigma_x^2} \mathbf{I}_n + \mathbf{S} \right)^{-1}.$$

Since $\frac{1}{\sigma_x^2} \ll 1$, then by Taylor expansion around $\mathbf{0}$, the trace of $\Sigma_{e,lmmse}$ can be expressed as

$$\begin{aligned} \text{tr}(\Sigma_{e,lmmse}) &= \sum_{k=0}^{\infty} \frac{(-1)^k}{\sigma_x^{2k}} \text{tr}(\mathbf{S}^{-k-1}) \\ &= \sum_{k=0}^l \frac{(-1)^k}{\sigma_x^{2k}} \text{tr}(\mathbf{S}^{-k-1}) + o\left(\frac{1}{\sigma_x^{2l}}\right). \end{aligned} \quad (45)$$

$$(46)$$

where $l \leq p-1$ is the order at which we truncate the Taylor expansion. Note that we impose the condition $r \leq p-1$ since the moments $\mu_{\mathbf{\Lambda}}(-k-1)$

are only defined for $1 \leq k \leq p-1$. Upon applying the expectation, we get:

$$\begin{aligned} \mathbb{E}_{\mathbf{H}}\{\|\hat{\mathbf{x}}_{lmmse} - \mathbf{x}\|^2\} &= \sum_{k=0}^l \frac{(-1)^k}{\sigma_x^{2k}} \mu_{\mathbf{\Lambda}}(-k-1) \\ &\quad + o\left(\sigma_x^{-2l}\right). \end{aligned} \quad (47)$$

2) Low SNR regime ($\sigma_x^2 \ll 1$):

We proceed similarly as the high SNR regime. For that, we make some manipulations on $\Sigma_{e,lmmse}$ as follows

$$\begin{aligned} \Sigma_{e,lmmse} &= \left(\frac{1}{\sigma_x^2} \mathbf{I}_n + \mathbf{S} \right)^{-1} \\ &= \sigma_x^2 (\mathbf{I}_n + \sigma_x^2 \mathbf{S})^{-1} \\ &= \sigma_x^2 (\sigma_x^2 \mathbf{I}_n + \mathbf{S}^{-1})^{-1} \mathbf{S}^{-1}. \end{aligned}$$

Using the same approach for proving Theorem 3.1), we get

$$\mathbb{E}_{\mathbf{H}}\{\|\hat{\mathbf{x}}_{lmmse} - \mathbf{x}\|^2\} = m \sum_{k=0}^{\infty} (-1)^k \sigma_x^{2k+2} \mu_{\mathbf{\Lambda}}(k). \quad (48)$$

As seen in the previous equation, we retain all the positive moments since they exist and can be computed by means of Theorem 1.

This completes the proof of Theorem 3

REFERENCES

- [1] R. Couillet and M. Debbah, "Free Deconvolution for OFDM Multicell SNR Detection," *19th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Cannes, France*, pp. 1–5, Sept 2008.
- [2] J. Yao, A. Kammoun, and J. Najim, "Eigenvalue estimation of parameterized covariance matrices of large dimensional data," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5893–5905, nov. 2012.
- [3] O. Ryan and M. Debbah, "Asymptotic behavior of random vandermonde matrices with entries on the unit circle," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3115–3147, July 2009.
- [4] J. Hoydis, M. Debbah, and M. Kobayashi, "Asymptotic moments for interference mitigation in correlated fading channels," in *Information Theory Proceedings (ISIT)*, July 2011, pp. 2796–2800.
- [5] D. Maiwald and D. Kraus, "Calculation of Moments of Complex Wishart and Complex Inverse Wishart Distributed Matrices," *IEE Proceeding-Radar Sonar and Navigation*, vol. 147, pp. 162–168, Aug. 2000.
- [6] G. Letac and H. Massam, "All Invariant Moments of the Wishart Distribution," *Scandinavian Journal of Statistics*, vol. 31, no. 2, pp. 295–318, Jun. 2004.
- [7] G. Alfano, A. M. Tulino, A. Lozano, and S. Verdú, "Capacity of MIMO Channels with One-sided Correlation," *IEEE Eighth International Symposium on Spread Spectrum Techniques and Applications, Sydney*, pp. 515–519, August 2004.

- [8] M. Elgaily, K. Elkhilil, H. Bahrami, and T. Al Naffouri, "Opportunistic Relay Selection with Limited Feedback," *IEEE Trans. on Communications*, vol. PP, no. 99, pp. 1–1, 2015.
- [9] K. Elkhilil, M. Elgaily, A. Kammoun, T. Y. Al Naffouri, and H. Bahrami, "On the Feedback Reduction of Relay Aided Multiuser Networks using Compressive Sensing," *To appear in IEEE Trans. on Communications*.
- [10] A. Edelman, "Eigenvalues and Condition Numbers of Random Matrices," Ph.D. dissertation, Massachusetts Institute of Technology, May 1989.
- [11] J. W. Silverstein and Z. D. Bai, "On the empirical distribution of Eigenvalues of a Class of Large Dimensional Random Matrices," *Journal of Multivariate Analysis*, vol. 54, pp. 175–192, May 2002.
- [12] X. Chen, "Antenna Correlation and Its Impact On Multi-Antenna System," *Progress In Electromagnetics Research B*, vol. 62, pp. 241–253, 2015.
- [13] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.
- [14] A. Sayed, *Fundamentals of Adaptive Filtering*. John Wiley & Sons, 2003.
- [15] H. V. Poor, *An Introduction to Signal Detection and Estimation*. Springer-Verlag, 1988.
- [16] B. Hassibi and B. Hochwald, "How Much Training is Needed in Multiple-Antenna Wireless Links?" *Information Theory, IEEE Transactions on*, vol. 49, no. 4, pp. 951–963, April 2003.
- [17] K. Elkhilil, M. Eltayeb, H. Dahrouj, and T. Al-Naffouri, "Distributed User Selection in Network MIMO Systems with Limited Feedback," *to appear in the 2015 IEEE 82nd Vehicular Technology Conference, Sept 2015, Boston, USA*.
- [18] M. Eltayeb, K. Elkhilil, A. Mas'ud, and T. Al-Naffouri, "Relay Selection with Limited and Noisy Feedback," *to appear in the 2015 IEEE 82nd Vehicular Technology Conference, Sept 2015, Boston, USA*.
- [19] K. Elkhilil, M. Eltayeb, H. Shibli, H. Bahrami, and T. Al-Naffouri, "Opportunistic Relay Selection in Multicast Relay Networks using Compressive Sensing," in *Global Communications Conference (GLOBECOM), 2014 IEEE*, Dec 2014, pp. 3126–3131.
- [20] Milutin Pajovic, "The Development and Application of Random Matrix Theory in Adaptive Signal Processing in the Sample Deficient Regime," Ph.D. dissertation, Massachusetts Institute of Technology, 2014.
- [21] H. L. Van Trees., *Optimum Array Processing: Detection, Estimation and Modulation Theory*. John Wiley and Sons, 2002.
- [22] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge University Press, 2011.
- [23] R. Yang and J. O. Berger, "Estimation of a covariance matrix using the Reference Prior," *Annals of Statistics*, vol. 22, no. 3, pp. 1195–1211, 1994.
- [24] M. A. Chaudhry and S. M. Zubair, *On a Class of Incomplete Gamma Function with Applications*. Boca Raton-London-Ney York-Washington, D.C.: Chapman & Hall/CRC, 2002.



Khalil Elkhilil was born in Gabes, Tunisia. He received the B.Eng. degree in Telecommunication Engineering (with first-class honors) from the Higher School of Communications of Tunis (Sup'Com), El-Ghazala, Tunisia, in 2012. He joined King Abdullah University of Science and Technology (KAUST) in August 2013 where he is currently a MS/PhD student in the Electrical Engineering program.

His research interests include signal processing for communication, compressive sensing, random matrix theory and its applications to wireless communications and radar systems.

Abla Kammoun was born in Sfax, Tunisia. She received the engineering degree in signal and systems from the Tunisia Polytechnic School, La Marsa, and the Master's degree and the Ph.D. degree in digital communications from Telecom Paris Tech [then Ecole Nationale Supérieure des Télécommunications (ENST)]. From June 2010 to April 2012, she has been a Postdoctoral Researcher in the TSI Department, Telecom Paris Tech. Then she has been at Suplec at the Alcatel-Lucent Chair on Flexible Radio until December 2013. Currently, she is a research scientist at KAUST university. Her research interests include performance analysis, random matrix theory, and semi-blind channel estimation. She is the recipient of the SAM 2014 second Prize Best paper award.



Tareq Y. Al-Naffouri received the B.S. degrees in mathematics and electrical engineering (with first honors) from King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, in 1994, the M.S. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, in 1998, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 2004. He was a

visiting scholar at California Institute of Technology, Pasadena, CA, from January to August 2005 and during summer 2006. He was a Fulbright scholar at the University of Southern California from February to September 2008. He has held internship positions at NEC Research Labs, Tokyo, Japan, in 1998, Adaptive Systems Lab, University of California at Los Angeles in 1999, National Semiconductor, Santa Clara, CA, in 2001 and 2002, and Beceem Communications Santa Clara, CA, in 2004. He is currently an Associate at the Electrical Engineering Department, King Abdullah University of Science and Technology (KAUST). His research interests lie in the areas of adaptive and statistical signal processing and their applications to wireless communications, seismic signal processing, and in multiuser information theory. He has recently been interested in compressive sensing and random matrix theory and their applications. He has over 130 publications in journal and conference proceedings, 9 standard contributions, 10 issued patents, and 4 pending.



Mohamed-Slim Alouini (S'94, M'98, SM'03, F09) was born in Tunis, Tunisia. He received the Ph.D. degree in Electrical Engineering from the California Institute of Technology (Caltech), Pasadena, CA, USA, in 1998. He served as a faculty member in the University of Minnesota, Minneapolis, MN, USA, then in the Texas A&M University at Qatar, Education City, Doha, Qatar before joining King

Abdullah University of Science and Technology (KAUST), Thuwal, Makkah Province, Saudi Arabia as a Professor of Electrical Engineering in 2009. His current research interests include the modeling, design, and performance analysis of wireless communication systems.