Blind Measurement Selection: A Random Matrix Theory Approach

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

Abstract—This paper considers the problem of selecting a set of k measurements from n available sensor observations. The selected measurements should minimize a certain error function assessing the error in estimating a certain m dimensional parameter vector. The exhaustive search inspecting each of the $\binom{n}{k}$ possible choices would require a very high computational complexity and as such is not practical for large n and k. Alternative methods with low complexity have recently been investigated but their main drawbacks are that 1) they require perfect knowledge of the measurement matrix and 2) they need to be applied at the pace of change of the measurement matrix. To overcome these issues, we consider the asymptotic regime in which k, n and m grow large at the same pace. Tools from random matrix theory are then used to approximate in closed-form the most important error measures that are commonly used. The asymptotic approximations are then leveraged to select properly k measurements exhibiting low values for the asymptotic error measures. Two heuristic algorithms are proposed: the first one merely consists in applying the convex optimization artifice to the asymptotic error measure. The second algorithm is a low-complexity greedy algorithm that attempts to look for a sufficiently good solution for the original minimization problem. The greedy algorithm can be applied to both the exact and the asymptotic error measures and can be thus implemented in blind and channel-aware fashions. We present two potential applications where the proposed algorithms can be used, namely antenna selection for uplink transmissions in large scale multi-user systems and sensor selection for wireless sensor networks. Numerical results are also presented and sustain the efficiency of the proposed blind methods in reaching the performances of channel-aware algorithms.

Index Terms—Measurement selection, blind selection, random matrix theory, Gram random matrices, massive MIMO, wireless sensor networks.

I. Introduction

MEASUREMENT selection is an old concept that finds its roots in many applications such as robotics, wireless sensor networks and wireless communications to name a few [1]–[3]. It aims to reduce the complexity of the estimation problem in linear models where the n-dimensional response vector is linearly related to the unknown m-dimensional vector. The reduction in computational complexity is achieved by using only the k measurements that minimize a certain given error function assessing the quality of the selected measurements. One naive approach for solving the measurement selection problem is to go through all $\binom{n}{k}$ possible selections

K. Elkhalil, 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.elkhalil, abla.kammoun, tareq.alnaffouri, slim.alouini}@kaust.edu.sa.

and select the ones that present the lowest achievable value of a given error measure. This procedure, though being optimal, is not practical especially when high dimensional observations are considered. In general, it seems that looking for an optimal solution is expected to call for solving an NP-hard problem as asserted by [1], which gives little hope of determining the optimal solution using a polynomial complexity algorithm. As a result, attention has turned to sub-optimal alternatives to solve the measurement selection problem. In this vein, genetic algorithms using some local search methods have been proposed in [4]. Local optimization techniques have been also proposed in [5,6]. Although these methods can have good performance with modest complexities, they do not guarantee any theoretically achievable bound on the performance. The first achievable bound has been derived in [1] where the authors in [1] resorted to convex relaxation artifice. This has led to a convex problem that can be solved with a complexity growing as $\mathcal{O}\left(n^3\right)$.

The aforementioned algorithms allow good performances coupled with a lower complexity as compared to exhaustive search. However, in order to accurately evaluate the overall complexity, it is important to consider how often the selection procedure should be repeated. For fast-fading varying linear models, the underlying measurement matrix, capturing the linear dependence between the input and output vectors, changes at a rapid pace. Hence, the overall complexity should be scaled by the number of times the measurement matrix changes over a given time window, which can result in a prohibitively high computational complexity. To overcome these issues, we propose in this work blind selection algorithms that leverage the statistics of the measurement matrix rather than its instantaneous realization. This can be for instance useful when for some practical concerns, it is not possible to acquire the measurement matrix. The main idea behind the proposed blind methods lies in the observation that the considered error measure depending on the random measurement matrix can be approximated by some deterministic quantities depending solely on the statistics of the measurement matrix. This fact is supported by results from random matrix theory, confirming the accuracy of the approximation in large dimensional settings. Using this theory, we show that the most used error measures can be approximated in closed-form by deterministic quantities depending soleley on the correlation between the columns of the measurement matrix. Interestingly, it turns out that, as far as the asymptotic regime is con, it concerned the correlation matrix that holds all the information about the best set of measurements to be selected. Particularly, it

is shown that if the columns of the measurement matrix are uncorrelated, any randomly selected subset of k measurements would asymptotically exhibit the same performances. It thus unfolds that the benefit from optimizing over the set of measurements to be selected is more significant in case of high correlation between the columns of the measurement matrix.

Based on the obtained asymptotic approximations of the considered error measures, we propose two different blind approaches. The first one is mostly inspired from the work of [1] and consists in applying the convex artifice to the asymptotic error measures. However, the optimization of the resulting problem might not be tractable, as there is no guarantee of the convexity of the asymptotic error measure. We, therefore, establish in this paper the convexity of the asymptotic error measure, which opens up the possibility of using standard convex optimization tools. The second algorithm is a greedy algorithm that attempts to get close to the optimal solution within a few iterations. Interestingly, the greedy algorithm can be also applied to the exact error measures, and can be thus implemented in both blind and channel-aware scenarios. It is shown that not only the greedy algorithm presents lower complexity but it also achieves higher performances than the convex-relaxation based algorithm when implemented in either blind or channel-aware modes.

The proposed algorithms can be used in many applications. We select in this paper two potential applications where the measurement selection problem arise. The first one concerns the design of low complexity linear receivers for uplink large-scale multi-user MIMO systems, better known as Massive MIMO systems. Such systems are gaining an increasing interest and constitute promising candidates for future wireless systems, primarily due to their abilities of achieving remarkable performance enhancements in terms of capacity, radiated energy efficiency and link reliability [7]–[9]. However, while the use of multiple antennas allows to significantly improve the spatial diversity, it comes inevitably at the cost of a higher computational complexity, which might call into question the feasibility of such systems [10]. Besides, it is not even clear whether the performance enhancement is worthy of using all antenna resources. It might happen in many scenarios that some antennas undergo severe fading, and as such, selecting a subset of antennas will result in substantial saving in complexity without sacrificing performance. Selecting these antennas is the crucial achievement of antenna selection algorithms. The use of these algorithms has already been advocated in [11,12] as an efficient solution to reduce the number of RF chains in conventional MIMO systems, leading to a significant reduction in complexity and costs while preserving most of the potential of full MIMO systems. The second application concerns the problem of sensor selection in wireless sensor networks (WSN), already extensively studied in the literature of signal processing. A question of interest in this field, is how to find the optimal placement of k sensors, assuming n available sensors in different locations [13].

To sum up, the major contributions of this paper are summarized as follows:

1) We provide accurate asymptotic approximations for the three most used error measures, namely the Mean Square Error (MSE), the Log Volume of the Confidence Ellipsoid (LCE) and the Worst Case Error Variance (WEV). These approximations depend solely on the statistics of the measurement matrix and are shown to be convex.
2) Based on the provided approximations, we propose two blind algorithms to perform antenna selection without knowledge of the instantaneous measurement matrix. The first one is based on the concept of convex relaxation while the second one is a greedy iterative algorithm that attempts to get close to the optimal solution.
3) We study the complexity of both algorithms and show that the greedy algorithm achieves quadratic complexity.
4) We select two applications in which the problem of sensor selection arise, namely antenna selection in massive MIMO systems and sensor selection in WSN. We show how the proposed algorithms can be used in both applications and compare their performances with channel aware algorithms which assume perfect knowledge of the measurement matrix. We show that as long as the correlation between the columns of the measurement matrix is high, the average performance is close to that of channel aware algorithms.

The remainder of the paper is organized as follows. In section II, we state the measurement selection problem and present some related works. In section III, we provide asymptotic approximations for the asymptotic error measures, based on which we propose two different blind algorithms to perform measurement selection. Finally, and prior to concluding the paper in section V, we discuss in section IV two potential applications for the proposed blind approach.

Notations: Throughout the paper, we use the following notations: Vectors are denoted by lower case bold letters and matrices are denoted by bold capital letters ($\mathbf{I}_n$ is the identity matrix of size n). For a given matrix $\mathbf{A}$, we refer by $[\mathbf{A}]_{i,j}$ its (i,j)th entry, and use $\mathbf{A}^T$ and $\mathbf{A}^H$ to denote its transpose and Hermitian respectively. We respectively denote by $\|.\|$, $\det(.)$ and $\mathrm{tr}(.)$, the spectral norm, the determinant and the trace of a matrix. Finally, we denote by $\mathrm{diag}(\mathbf{a})$, the diagonal matrix with diagonal elements, the entries of $\mathbf{a}$.

II. PROBLEM FORMULATION AND RELATED WORKS

We consider the problem of estimating an $m-$dimensional vector $\mathbf{x} \in \mathbb{C}^{m \times 1}$ from measurements when the observed vector $\mathbf{y} \in \mathbb{C}^{n \times 1}$ is related to $\mathbf{x}$ by the following relation:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}, \tag{1}$$

Herein $\mathbf{H} = \{h_{i,j}\} \in \mathbb{C}^{n \times m}$ denotes the measurement matrix and is assumed to be Gaussian with one-side correlation $\mathbf{R}$, i.e., $\mathbf{H} = \mathbf{R}^{\frac{1}{2}}\mathbf{W}$, where $\mathbf{W} \in \mathbb{C}^{n \times m}$ is a matrix with i.i.d zero mean unit variance entries and $\mathbf{v}$ is the additive noise vector with independent, zero mean, unit variance, circularly symmetric complex Gaussian entries, i.e. $\mathbf{v} \sim \mathcal{CN}(\mathbf{0}_{n \times 1}, \mathbf{I}_n)$. Assuming that the number of measurements n exceeds the signal dimension m, i.e., $n > m$, the estimate of $\mathbf{x}$ denoted by $\widehat{\mathbf{x}}$ can be recovered by using the least square (LS) estimator

as [14]

$$\widehat{\mathbf{x}} = \left(\mathbf{H}^H\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{y}$$
$$= \left(\mathbf{W}^H\mathbf{RW}\right)^{-1}\mathbf{W}^H\mathbf{R}^{\frac{1}{2}}\mathbf{y} \qquad (2)$$
$$= \mathbf{x} + \left(\mathbf{W}^H\mathbf{RW}\right)^{-1}\mathbf{W}^H\mathbf{R}^{\frac{1}{2}}\mathbf{v}.$$

Clearly, the estimation error, $\mathbf{x} - \widehat{\mathbf{x}}$ is zero mean with covariance:

$$\mathbf{\Sigma} = \left(\mathbf{W}^H\mathbf{RW}\right)^{-1}. \qquad (3)$$

A. Measurement selection

Measurement selection intends to select the k *best* measurements that are the most representative in the sense that they constitute the set of k measurements minimizing a certain given error. Once selected, these measurements will be used in place of the whole available data vector. By reducing the available number of measurements, measurement selection inevitably induces a performance loss, but allows a reduction in the computational complexity. This becomes all the more important in several applications such as large-scale MIMO systems, where the treatment of the whole number of measurements received by the large antenna array might be not possible. To mathematically formulate measurement selection, we define the selection matrix $\mathbf{S} \in \mathbb{R}^{k \times n}$ as the matrix that permits to extract k measurements from $\mathbf{y}$. Let $\mathcal{S}$ denote the set of indexes of cardinality k containing the indexes of measurements to be selected. The selected measurement vector is thus given by

$$\mathbf{y}_{\mathcal{S}} = \mathbf{S}\mathbf{y}, \qquad (4)$$

where $\mathbf{S}$ is defined as follows

$$[\mathbf{S}]_{i,j} = \begin{cases} 1 & j = \mathcal{S}[i] \\ 0 & \text{otherwise} \end{cases}, i = 1, \cdots, k. \qquad (5)$$

where $\mathcal{S}[i]$ denotes the i-th element in set $\mathcal{S}$. Based on the structure of $\mathbf{S}$ in (5), we have the following properties

- $\mathbf{S}\mathbf{S}^T = \mathbf{I}_k$.
- $\mathbf{S}^T\mathbf{S} = \text{diag}(\mathbf{s})$.

where $\mathbf{s} = \{s_i\}_{i=1,\cdots,n}$ is a n-dimensional vector with entries equal to 1 at the locations given by $\mathcal{S}$ and zeros elsewhere. The LS estimator $\widehat{\mathbf{x}}_{\mathcal{S}}$ obtained from using the selected measurement vector $\mathbf{y}_{\mathcal{S}}$ is :

$$\widehat{\mathbf{x}}_{\mathcal{S}} = \left(\mathbf{H}^H\text{diag}(\mathbf{s})\mathbf{H}\right)^{-1}\mathbf{H}^H\mathbf{S}^T\mathbf{y}_S$$

Upon applying the operator defined by $\mathbf{S}$, the resulting error covariance matrix, which we denote by $\mathbf{\Sigma}_{\mathcal{S}}$ easily writes as

$$\mathbf{\Sigma}(\mathbf{s}) = \left(\mathbf{H}^H\text{diag}(\mathbf{s})\mathbf{H}\right)^{-1}$$
$$= \left(\mathbf{W}^H\mathbf{R}^{\frac{1}{2}}\text{diag}(\mathbf{s})\mathbf{R}^{\frac{1}{2}}\mathbf{W}\right)^{-1}. \qquad (6)$$

It can be seen from (6) that $\mathbf{\Sigma}_{\mathcal{S}}$ is a Gram matrix with one side correlation given by the matrix $\mathbf{R}^{\frac{1}{2}}\text{diag}(\mathbf{s})\mathbf{R}^{\frac{1}{2}}$.
Measurement selection consists in selecting the optimal set $\mathcal{S}^*$ with cardinality k, or equivalently vector $\mathbf{s}$ with only k non-zero elements equal to 1, that minimizes a certain error measure assessing the estimation quality. In other words, the optimal set can be obtained as the solution of the following problem:

$$\mathbf{s}^* = \underset{\mathbf{s} \in \mathbb{R}^n}{\text{argmin}} \quad f(\mathbf{\Sigma}(\mathbf{s}))$$
$$\text{s.t.} \quad s_i \in \{0, 1\} \qquad (7)$$
$$\mathbf{1}^T\mathbf{s} = k$$

where here f denotes the considered measure function. There have been various measures proposed in the literature [15] to assess the estimation quality of the LS. All of them heavily depend on the covariance matrix of the error . In this paper, we will focus on the following ones:

1) The Mean Square Error (MSE): The mean square error is defined as the average euclidian distance between the estimated vector and $\mathbf{x}$. When only a set of k measurements is employed, the MSE writes as:

$$\text{MSE}(\mathbf{s}) = \text{tr}\left(\mathbf{W}^H\mathbf{R}^{\frac{1}{2}}\text{diag}(\mathbf{s})\mathbf{R}^{\frac{1}{2}}\mathbf{W}\right)^{-1} \qquad (8)$$

2) The Log Volume of the confidence Ellipsoid (LCE) For a Gaussian random vector $\widehat{\mathbf{x}}$ in $\mathbb{C}^m$ with mean $\mathbf{x}$ and covariance $\mathbf{\Sigma}$, the $\eta-$ confidence ellipsoid is a multi-dimensional generalization of the $\eta-$ confidence interval. It corresponds to the minimum volume ellipsoid that contains $\widehat{\mathbf{x}} - \mathbf{x}$ with propability η and is given by:

$$\mathcal{E} = \left\{\mathbf{z} : \mathbf{z}^H\mathbf{\Sigma}^{-1}\mathbf{z} \leq \alpha\right\}, \qquad (9)$$

where $\alpha = F_{\chi^2_{2m}}^{-1}(\eta)$, $F_{\chi^2_{2m}}$ being the cumulative distribution function of a chi-squared random variable with $2m$ degrees of freedom. The volume of the $\eta-$confidence ellipsoid defined in (9) is (see Serfling [15])

$$\text{vol}(\mathcal{E}_\alpha) = \frac{(\alpha\pi)^m}{\Gamma(m+1)}\det(\mathbf{\Sigma}), \qquad (10)$$

where $\Gamma(.)$ is the Gamma function. It appears from 10 that the determinant of $\mathbf{\Sigma}$ plays the role played by the variance in one dimension hence its name *generalized variance*. When vector $\widehat{\mathbf{x}}$ represents an estimate of a given parameter vector, the lowest is the generalized variance, the highest is the estimation quality. It might be more convenient in practice to work with the log of the volume of the $\eta-$ ellipsoid. A good estimate is thus characterized by a small value of the log-epsilloid volume, which will be confused, from now on, with $\frac{1}{m}\log\det(\frac{1}{n}\mathbf{\Sigma})$. We define thus the log volume of the confidence ellipsoid associated with the vector of selected measurement $\widehat{\mathbf{x}}_{\mathcal{S}}$ as:

$$\text{LCE}(\mathbf{s}) = -\frac{1}{m}\log\left(\det\left(\frac{1}{n}\mathbf{H}^H\text{diag}(\mathbf{s})\mathbf{H}\right)\right)$$
$$= -\frac{1}{m}\log\left(\det\left(\frac{1}{n}\mathbf{W}^H\mathbf{R}^{\frac{1}{2}}\text{diag}(\mathbf{s})\mathbf{R}^{\frac{1}{2}}\mathbf{W}\right)\right)$$

3) Worst Case Error Variance (WEV): The worst case error variance (WEV) quantifies the maximum variance of error over all directions. It corresponds to the maximum eigenvalue of the error covariance matrix. The WEV

associated with the vector of selected measurement $\widehat{\mathbf{x}}_{\mathcal{S}}$ is thus defined as [1]:

$$\text{WEV}(\mathbf{s}) = \frac{1}{\lambda_{min}\left(\frac{1}{m}\mathbf{W}^H\mathbf{R}^{\frac{1}{2}}\text{diag}(\mathbf{s})\mathbf{R}^{\frac{1}{2}}\mathbf{W}\right)}$$

B. Related Works

The main literature related to the present paper is represented by the work in [1], The main idea in this work relies on the observation that it is the non-convex nature of the constraints, requiring the selection vector $\mathbf{s}$ to possess elements in $\{0, 1\}$ that makes problem (7) intractable. To overcome this issue, [1] solves instead a convex related problem obtained by substituting the Boolean constraints $s_i \in \{0, 1\}$ by the convex constraints $0 \le s_i \le 1$:

$$\widehat{\mathbf{s}} = \arg\min_{\mathbf{s} \in \mathbb{R}^n} f(\mathbf{\Sigma}(\mathbf{s}))$$
$$\text{s.t.} \quad \mathbf{1}^T\mathbf{s} = k \qquad (11)$$
$$0 \le s_i \le 1, \quad i = 1, \cdots, n.$$

It is worth mentioning that the output of the optimization in (11) yields a higher value than the maximum objective function in (7), and as such can be viewed as a global upper bound on the performance. Moreover, the optimal vector $\widehat{\mathbf{s}}$ can contain real values not necessarily zeros and ones. In order to obtain the indexes of the selected measurements, one should order the entries of $\widehat{\mathbf{s}}$ and then assign ones to the k greatest values and zeros to the remaining entries. This results in a feasible solution to the selection problem in (7) which yields a lower bound on the objective function.

To solve the problem in (11), one can resort to interior-point methods which require few tens of iterations to converge where each iteration is performed with a complexity of $\mathcal{O}\left(n^3\right)$ computations. For more details on convex relaxation, the readers are referred to [1] and references therein.

III. BLIND MEASUREMENT SELECTION

Previous works dealing with sensor selection have essentially been based on the assumption of perfect knowledge of the measurement matrix $\mathbf{H}$. This not only might not be satisfied in practice but also can make the application of the previously proposed algorithms more difficult as they should be applied at every change in the measurement matrix $\mathbf{H}$.

To overcome this issue, we propose in this work blind methods that leverage the knowldege of the channel statistics to perform the selection. These methods rely heavily on advanced results of random matrix theory.

Main Idea: The idea behind the proposed methods hinges on the fact that as the dimensions of the channel matrix grow large, quantities depending on the measurement matrix become more predictable in that they can be well-approximated by deterministic quantities depending only on the channel statistics. Such deterministic quantities can be characterized by resorting to tools from random matrix theory. In light of this observation, we propose in this work to compute, in closed

form, accurate approximations of the three error measures, namely the MSE, LCE and WEV. As we will see later, the asymptotic analysis can be leveraged to blindly select good sets of measurements.

A. Asymptotic analysis of the error measures

In this section, we determine, in closed form, accurate approximations for the MSE, LCE and WEV. For technical purposes, we shall consider the following growth regime:

Assumption 1. *We assume both that n and m grow large while their ratio $\frac{n}{m}$ satisfies:*

$$\frac{n}{m} \to c \in (1, \infty)$$

We also assume that k grows large with:

$$0 < \liminf \frac{k}{n} < \limsup \frac{k}{n} < 1.$$

and

$$\liminf \frac{k}{m} > 1.$$

The channel matrix $\mathbf{H}$ is assumed to follow the following statistical model:

Assumption 2. $\mathbf{H} \in \mathbb{C}^{n \times m}$ *is complex Gaussian matrix with one-side correlation $\mathbf{R}$, i.e,*

$$\mathbf{H} = \mathbf{R}^{\frac{1}{2}}\mathbf{W}$$

where $\mathbf{W}$ is a matrix with i.i.d. normally distributed entries having zero-mean and unit variance. Moreover, the matrix $\mathbf{R}$ satisfies the following conditions:

1) $\mathbf{R}$ *has a bounded spectral norm, i.e,*

$$\sup_n \|\mathbf{R}\| < \infty$$

2) *The normalized trace of $\mathbf{R}$ satisfies:*

$$\inf_n \frac{1}{n}\text{tr}\mathbf{R} > 0.$$

With these assumptions at hand, we are ready to analyze the asymptotic behavior for the MSE and LCE.

Lemma 1. *[16] Let δ be the unique solution to the following equation*

$$\delta = m\left(\text{tr}\left[\mathbf{R}\text{diag}(s)\left(\mathbf{I}_n + \delta\mathbf{R}\text{diag}(\mathbf{s})\right)^{-1}\right]\right)^{-1} \qquad (12)$$

Define $\overline{\text{MSE}}(\mathbf{s})$ as

$$\overline{\text{MSE}}(\mathbf{s}) = \delta$$

Then, under assumptions 1 and 2, $\text{MSE}(\mathbf{s})$ satisfies

$$\text{MSE}(\mathbf{s}) - \overline{\text{MSE}}(\mathbf{s}) \xrightarrow[n \to \infty]{a.s.} 0.$$

Lemma 2. *[17, Proposition 4.2] Let δ be defined as in (12). Define $\overline{\text{LCE}}(\mathbf{s})$ as*

$$\overline{\text{LCE}}(\mathbf{s}) = -\frac{1}{m}\log\det\left(\mathbf{I}_n + \delta\mathbf{R}\text{diag}(\mathbf{s})\right) + \log(c\delta) + 1$$

---

[1] The normalization factor $\frac{1}{m}$ is considered herein to comply with the asymptotic growth regime of random matrix theory

Then, under assumptions 1 and 2,

$$\mathrm{LCE}(\mathbf{s}) - \overline{\mathrm{LCE}}(\mathbf{s}) \xrightarrow[n\to\infty]{a.s.} 0$$

To obtain an asymptotic equivalent for the WEV, the following technical assumption is additonally needed:

Assumption 3. *Let* $\lambda_1(\mathbf{s}), \cdots, \lambda_n(\mathbf{s})$ *be the eigenvalues of* $\mathbf{R}^{\frac{1}{2}}\mathrm{diag}(\mathbf{s})\mathbf{R}^{\frac{1}{2}}$. *We assume that the probability measure* $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\delta}_{\lambda_i}(\mathbf{s})$ *converge weakly to a probability measure* μ_s. *Moreover, we assume all the eigenvalues of* $\mathbf{R}^{\frac{1}{2}}\mathrm{diag}(\mathbf{s})\mathbf{R}^{\frac{1}{2}}$ *to be almost surely contain in the support of* μ_s $Supp(\mu_s)$ *or equivalently:*

$$\max_{i=1,\cdots,n} \mathrm{dist}(\lambda_i(\mathbf{R}^{\frac{1}{2}}\mathrm{diag}(\mathbf{s})\mathbf{R}^{\frac{1}{2}}), \mu_s)) \xrightarrow[n\to\infty]{a.s.} 0$$

With this assumption at hand, the WEV can be approximated as

Lemma 3. *[18] Under assumptions 1, 2 and 3,*

$$\mathrm{WEV}(\mathbf{s}) - \overline{\mathrm{WEV}}(\mathbf{s}) \xrightarrow[n\to\infty]{a.s.} 0$$

where $\overline{\mathrm{WEV}}(\mathbf{s})$ *is given by:*

$$\overline{\mathrm{WEV}}(\mathbf{s}) = -\frac{1}{\eta} + \frac{1}{m}\mathrm{tr}\left[\mathbf{R}\,\mathrm{diag}\,(\mathbf{s})\,(\mathbf{I}+\eta\mathbf{R}\,\mathrm{diag}\,(\mathbf{s}))^{-1}\right]$$

and η *is the solution to the following equation in* $(0,\infty)$

$$\eta^2 = \left(\frac{1}{m}\mathrm{tr}\left[\left(\mathbf{R}^{\frac{1}{2}}\mathrm{diag}(\mathbf{s})\mathbf{R}^{\frac{1}{2}}\right)^2\left(\mathbf{I}_n+\eta\mathbf{R}^{\frac{1}{2}}\mathrm{diag}(\mathbf{s})\mathbf{R}^{\frac{1}{2}}\right)^{-2}\right]\right)^{-1}.$$

B. Blind selection techniques

The asymptotic analysis carried out in the previous section is now leveraged to build efficient blind methods for measurement selection. Our blind approaches are based on solving the following selection problem:

$$\mathbf{s}^* = \underset{\mathbf{s}}{\mathrm{argmin}}\,\overline{f}(\mathbf{s})$$
$$\mathrm{s.t.} \qquad \mathbf{1}^T\mathbf{s} = k \qquad (13)$$
$$s_i \in \{0,1\}, i = 1,\cdots,n.$$

where $\overline{f}$ refers to one of the asymptotic approximations for MSE, LCE or the WEV, that have been computed in the previous section. We present in the sequel two different methods. The first one, termed blind convex relaxation technique relies on the use of the convex relaxation approach used in [1], while the second one, is merely based on the use of a greedy algorithm that solves the Problem in (13)

1) Blind convex relaxation technique: This method builds upon the use of the convex relaxation concept. It replaces the boolean constraints in (13) by the convex constraints $0 \leq s_i \leq 1$. In doing so, we obtain the following optimization problem:

$$\mathbf{s}^* = \underset{\mathbf{s}}{\mathrm{argmin}}\,\overline{f}(\mathbf{s})$$
$$\mathrm{s.t.} \qquad \mathbf{1}^T\mathbf{s} = k \qquad (14)$$
$$0 \leq s_i \leq 1, i = 1,\cdots,n.$$

It is worth mentioning that, although the non-convex relaxation techniques are now replaced by the convex ones $0 \leq s_i \leq 1$, it is not clear whether the obtained problem (14) remains convex. This is because we are not sure whether the objective function, representing the almost sure deterministic equivalent of one of the error measures is still (14) convex. The following theorem answers this question and establishes the convexity of the deterministic approximation of the MSE, LCE and WEV.

Theorem 1. *Define* $\overline{f}$ *as:*

$$\overline{f} : \mathbb{R}_+^n \to \mathbb{R}_+$$
$$\mathbf{s} \mapsto \overline{f}(\mathbf{s})$$

where $\overline{f}$ *is either* $\overline{\mathrm{MSE}}(\mathbf{s})$, $\overline{\mathrm{LCE}}(\mathbf{s})$ *or* $\overline{\mathrm{WEV}}(\mathbf{s})$. *Then,* $\overline{f}$ *is convex in* $\mathbb{R}_+^n$.

Proof: See Appendix ∎

Based on this Theorem, one can thus resort to standard convex tools to solve Problem (14). This step often requires the knowldege of the gradient of $\overline{f}$ with respect to $\mathbf{s}$ which can be found in the top of the next page where $\mathbf{R_s} = \mathbf{R}^{\frac{1}{2}}\,\mathrm{diag}\,(\mathbf{s})\,\mathbf{R}^{\frac{1}{2}}$ (A proof can be found in Appendix A) The solution of (14) is a vector with real positive values not necessarily zeros and ones. To obtain the indexes of the selected measurements, we should order its entries and then set the greatest ones to 1 and set the remaining to zero.

2) Greedy blind algorithms: Greedy algorithms have been widely applied to the framework of wireless communications, particularly in scheduling where the aim is to select the set of users that maximizes a certain utility function [19]. The use of greedy algorithms for measurements selection is, however, less common. In order to stress the wide scope of applicability of the proposed algorithm, we consider here the problem of selecting the index of measurements that minimizes a pre-defined error measure $f(\mathcal{H},\mathcal{S})$, where $\mathcal{H}$ is some information about the measurement matrix $\mathbf{H}$ [2] and $\mathcal{S}$ is a set of k indexes from $\{1,\cdots n\}$. The principle of the proposed greedy algorithm is as follows. First, we start by choosing an initial candidate set $\mathcal{S}$ obtained by randomly selecting a pattern (set of measurements indexes) of size k. Then, select from the set of the remaining indexes ($\overline{\mathcal{S}} = \{1,\cdots,n\}\setminus\mathcal{S}$), the first value that, when replaced with one of the indexes in $\mathcal{S}$ leads to a reduction in $f(\mathcal{H},\mathcal{S})$. When this occurs, $\mathcal{S}$ is updated by replacing the index that presents the largest reduction in $f(\mathcal{H},\mathcal{S})$. This procedure is repeated for a predetermined number of iterations K. The corresponding algorithm is detailed in Algorithm 1. It can be applied for any metric $f(\mathcal{H},\mathcal{S})$. This implies that the greedy algorithm might be considered as a channel aware algorithm when $\mathcal{H}$ is given by $\mathbf{H}$ and entirely blind when $\mathcal{H}$ contains only statistical information about the channel.

Proposition 1. *The greedy algorithm described by the steps of Algorithm 1 is guaranteed to converge.*

---

[2] $\mathcal{H}$ could be for example the statistics given by $\mathbf{R}$ or the full channel matrix $\mathbf{H}$.

$$\frac{\partial \overline{\text{MSE}}\left(\mathbf{s}\right)}{\partial s_i} = \frac{\partial \delta}{\partial s_i} = \delta_i' = -\frac{\delta \left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I}+\delta\mathbf{R}_s\right)^{-2}\mathbf{R}^{\frac{1}{2}}\right]_{i,i}}{\text{tr}\left[\mathbf{R_s}\left(\mathbf{I}+\delta\mathbf{R_s}\right)^{-2}\right]}.$$

$$\frac{\partial \overline{\text{LCE}}\left(\mathbf{s}\right)}{\partial s_i} = \left(1-c\right)\frac{\delta_i'}{\delta} + \frac{\delta_i'}{m\delta}\text{tr}\left(\mathbf{I}+\delta\mathbf{R}_s\right)^{-1} - \frac{\delta}{m}\left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I}+\delta\mathbf{R}_s\right)^{-1}\mathbf{R}^{\frac{1}{2}}\right]_{i,i}.$$

$$\frac{\partial \eta}{\partial s_i} = \eta_i' = -\eta\left(\text{tr}\left[\mathbf{R_s}\left(\mathbf{I}+\eta\mathbf{R_s}\right)^{-2}\right] - \text{tr}\left[\mathbf{R_s}\left(\mathbf{I}+\eta\mathbf{R_s}\right)^{-3}\right]\right)^{-1}\left(\left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I}+\eta\mathbf{R}_s\right)^{-2}\mathbf{R}^{\frac{1}{2}}\right]_{i,i} - \left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I}+\eta\mathbf{R}_s\right)^{-3}\mathbf{R}^{\frac{1}{2}}\right]_{i,i}\right).$$

$$\frac{\partial \overline{\text{WEV}}\left(\mathbf{s}\right)}{\partial s_i} = -\frac{\left(c-1\right)\eta_i'}{\eta^2} + \frac{\eta_i'}{m\eta^2}\left[2\,\text{tr}\left(\mathbf{I}+\eta\mathbf{R}_s\right)^{-1} - \text{tr}\left(\mathbf{I}+\eta\mathbf{R}_s\right)^{-2}\right] + \frac{1}{m}\left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I}+\eta\mathbf{R}_s\right)^{-2}\mathbf{R}^{\frac{1}{2}}\right]_{i,i}.$$

Algorithm	Complexity
Convex Optimization(Channel-aware)	$N \times \mathcal{O}\left(n^3\right)$
Convex Optimization(Blind)	$\mathcal{O}\left(n^3\right)$
Greedy(Channel-aware)	$K \times N \times \mathcal{O}\left(n^2\right)$
Greedy(Blind)	$K \times \mathcal{O}\left(n^2\right)$

Table I: Computational complexity of the different proposed algorithms.

Proof: By construction of Algorithm 1, the computation metric $f\left(\mathcal{H},\mathcal{S}\right)$ decreases on each iteration. Since the performance is bounded by the optimal performance achieved through the exhaustive search, the algorithm produces a decreasing bounded sequence, which implying its convergence. ∎

C. Complexity Analysis

In this part, we discuss the complexity of the different selection algorithms. Consider first the case of full blind methods. The convex approach requires $\mathcal{O}(n^3)$ operations which is the cost of using interior-point methods. As for the greedy approach, the computational complexity is governed by two factors. i) the complexity needed at every iteration and ii) the total number of iterations until convergence, which we denote by K. At every iteration, we need to perform $k\left(n-k\right)$ computations, thus in total, we need $K \times k\left(n-k\right)$ computations. As k and n are assumed to be commensurable, the computation complexity is thus $K \times \mathcal{O}\left(n^2\right)$. Now, if the greedy approach and convex relaxation based techniques are applied when the channel is perfectly known, complexity has to be multiplied by N which represents the number of times over which the channel changes. To sum up, we present the complexity achieved by the proposed selection algorithms in Table I in both full blind and CSI aware scenarios. Figure 1 represents the MSE performance achieved by the greedy algorithm for both cases (channel-aware and blind) as a function of the number of iterations. For both cases, the greedy algorithm requires a number of iterations, $K = 2$ to converge. This value of K will be implemented in all the next simulations for the greedy algorithm.

IV. POTENTIAL APPLICATIONS

In this section, we show two potential applications in which the proposed blind measurement selection algorithms
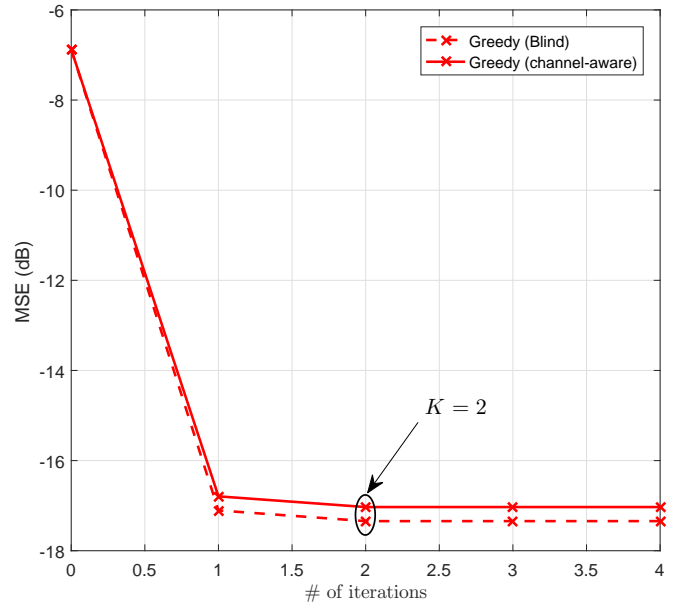

Figure 1: Average MSE performance (over 100 realizations) of the greedy approach for both cases (channel-aware and blind) as function of the number of iterations. $\mathbf{R}$ follows the model in (17) ($m = 30$, $n = 100$, $k = 50$ and $d = 2$).

can be applied to help reducing the computational cost. The first application concerns antenna selection in massive MIMO systems while the second focuses on the problem of sensor selection in WSN. In the following, we provide a detailed description of the system model in each application and analyze the performance in terms of the error measures proposed in section II.

A. Antenna Selection for Single-cell Uplink Massive MIMO Systems

Consider the uplink of a single cell MU-MIMO system in which m single-antenna users are served by a single base station (BS) equipped with n antennas with $m < n$, as sketched in Figure 2. Assuming that the users' signals are perfectly synchronized in time and frequency, the received vector at the BS is given by

$$\mathbf{y} = \sqrt{\rho}\mathbf{H}\mathbf{x} + \mathbf{e}, \tag{15}$$

where $\mathbf{y} \in \mathbb{C}^{n \times 1}$ is the received vector at the BS, ρ is the average transmit power per user and $\mathbf{x} \in \mathbb{C}^{m \times 1}$ is the data vector. Matrix $\mathbf{H} = \{h_{i,j}\} \in \mathbb{C}^{n \times m}$ denotes the narrow-band uplink channel matrix where $h_{i,j}$ is the channel coefficient between the j-th user and the i-th BS's antenna. Moreover, we assume that the random channel $\mathbf{H}$ exhibits the one-sided Kronecker model given by

$$\mathbf{H} = \mathbf{R}^{\frac{1}{2}} \mathbf{W}, \qquad (16)$$

where $\mathbf{W} \in \mathbb{C}^{n \times m}$ is a matrix with $i.i.d$ circularly symmetric zero mean unit-variance complex Gaussian entries, $\mathbf{R}$ models the spatial receive correlation matrix, whose elements represent the correlation between the antennas of the BS and $\mathbf{e}$ denotes noise vector at the BS with $i.i.d$ circularly symmetric zero mean unit-variance complex Gaussian entries, i.e., $\mathbf{e} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_n)$. At the receiver side, the BS estimates the transmitted vector $\mathbf{x}$ using $\mathbf{y}$. Several detection procedures can be used, among which are the optimal maximum likelihood (ML) detector and the least squares. The latter achieves a good balance between complexity and performance. In communication parlance, it is referred to as zero-forcing (ZF) detection and is given by

$$\widehat{\mathbf{x}} = \frac{1}{\sqrt{\rho}} \mathbf{H}^{\dagger} \mathbf{y},$$

where $\mathbf{H}^{\dagger} = \left(\mathbf{H}^H \mathbf{H}\right)^{-1} \mathbf{H}^H$ is the pseudo-inverse of $\mathbf{H}$. Even with the use of a ZF detector in place of the optimal ML decoder, the complexity of the decoding might be prohibitively high as a result of the high number of antennas n. Antenna selection appears thus as a valuable technique that can allow decoding with a lower complexity. In this respect, we evaluate the performance of the aforementioned antenna selection procedures for this practical scenario.
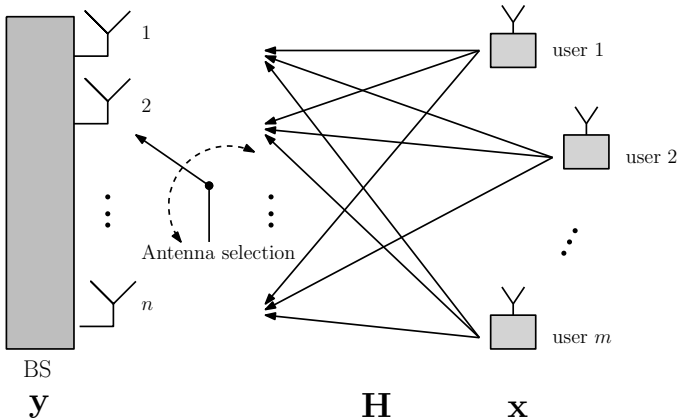

Figure 2: System model of an uplink MU-MIMO system composed of a BS equipped with n antennas and serving m single-antenna users.

Numerical Example: In all experiments, we assume that the number of users m is 30 and the total budget of antennas is
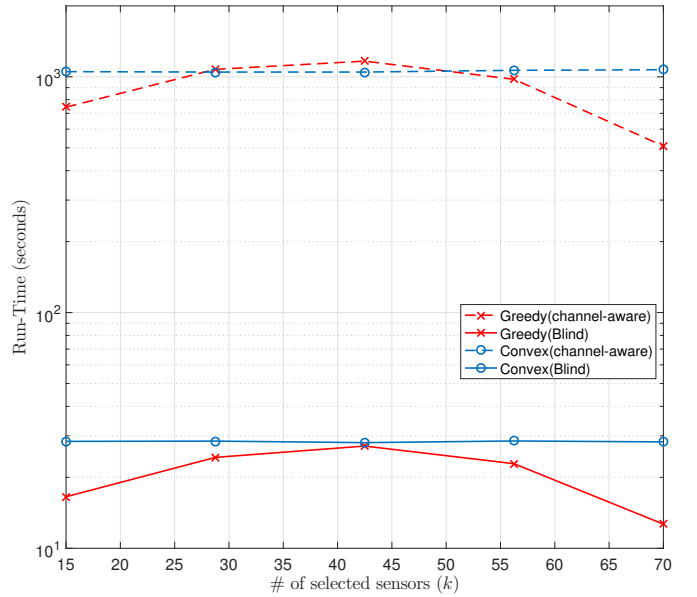

Figure 3: Run-time in seconds vs the number of selected antennas k for the considered selection algorithms.

$n = 100$. We also set the SNR to $\rho = 20$ dB. The following spatial correlation model [20] is considered:

$$\mathbf{R}_{i,j} = \exp\left(-0.05.d^2 (i-j)^2\right), \ 1 \le i,j \le n. \qquad (17)$$

that models a broadside Gaussian power azimuth spectrum with a root-mean-square spread of $2°$ where d corresponds to the antenna separation in wavelength units. The greedy and the convex relaxation based algorithms are considered in channel-aware ($\mathcal{H} = \mathbf{H}$) and fully blind scenarios ($\mathcal{H} = \mathbf{R}$). Figure 4 reports the achieved averaged MSE (over 100 realizations) for all proposed selection algorithms along with the random selection algorithm that randomly picks a set of k antennas out of n. As a major observation, we note that when the correlation between antennas is low ($d = 4$), the proposed blind algorithms are not that advantageous as compared to the random selection algorithm. This is kind of expected since the rows of $\mathbf{H}$ become almost statistically independent and identically distributed. They are thus statistically equivalent and selecting any k rows would asymptotically achieve the same MSE, as can be evidenced from the deterministic equivalent of the MSE shown in Lemma 1. However, with the impact of correlation becoming more important ($d \downarrow$), the gain of blind approaches over the random selection approach increases. They constitute thus a valuable option, given the fact that they only entail a loss of a up to 1 dB as compared to channel-aware algorithms. This can be clearly seen in Figure 5, where we plot the average MSE against the antennas' separation d. It is worth mentioning that for $d = 1$ and $d = 2$, the proposed blind greedy approach outperforms the channel-aware convex approach. This may sounds counter intuitive, but this is in fact due to the quantization effect at the output of the relaxed optimization problem. Also, it is worth mentioning that the blind greedy algorithms perform antenna selection at the pace of the variation of the large scale statistics. This must
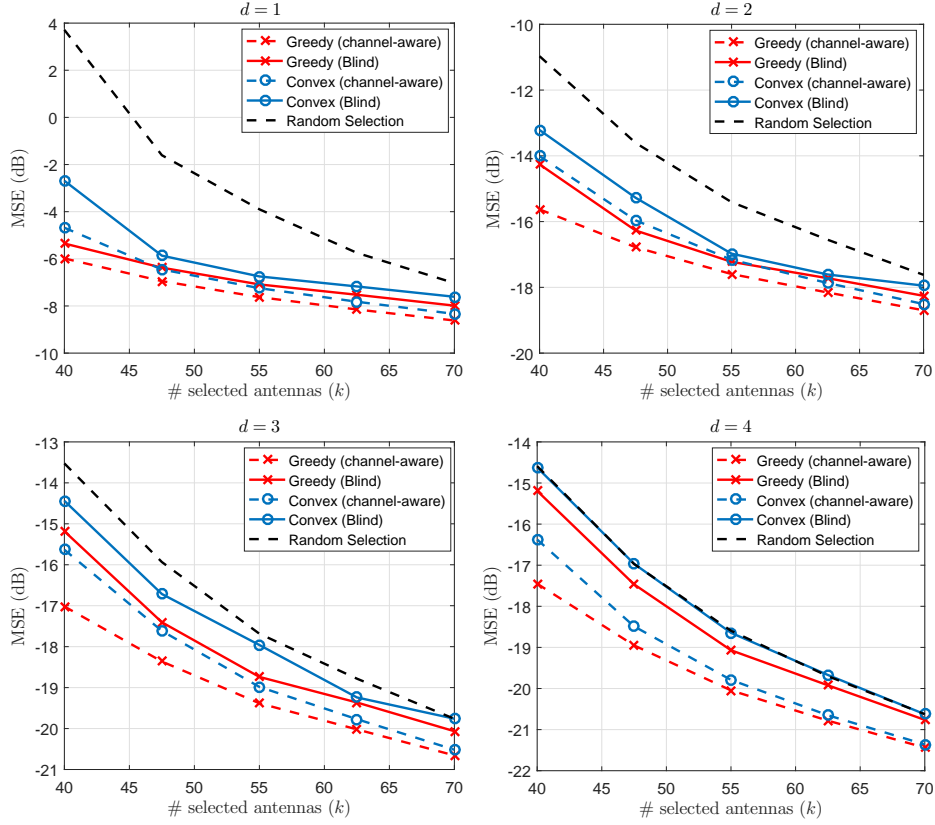
Figure 4: Average MSE achieved by the proposed selection techniques versus k for different values of the antennas' separation d.

be compared with the channel aware algorithms which are required to perform antenna selection for every channel realization. A high reduction in the computational complexity is thus achieved as evidenced by Figure 3, showing the run-time in seconds consumed by the different selection algorithms. All in all, it appears that the proposed blind selection techniques present in reality a better trade-off between complexity and performance.

B. Sensor Selection in WSN

We consider a wireless sensor network (WSN) with total number of $n = 100$ sensor nodes sensing a phenomena of dimension $m = 30$ where sensors are randomly deployed over a circular area of radius 30m. Thus, we have the same linear system as in (1). Unlike the previous application, we assume that the rows of $\mathbf{H}$ are statistically independent and consider the correlation in the measurement noise. The error covariance matrix in this case becomes

$$\mathbf{\Sigma} = \left(\mathbf{W}^H \mathbf{\Phi}^{-1} \mathbf{W}\right)^{-1}, \tag{18}$$

where $\mathbf{\Phi}$ is the noise covariance matrix given by [21]

$$\mathbf{\Phi}_{i,j} = \sigma^2 \exp\left(-\rho \left\|S_i - S_j\right\|_2\right), \ 1 \le i, j \le n, \tag{19}$$

where $\left\|S_i - S_j\right\|_2$ denotes the Euclidean distance between nodes' locations in the 2D plane S_i and S_j, $\sigma^2 = 1$ and ρ is the correlation parameter that controls the strength of the spatial correlation. Obviously, a larger ρ results in a weaker
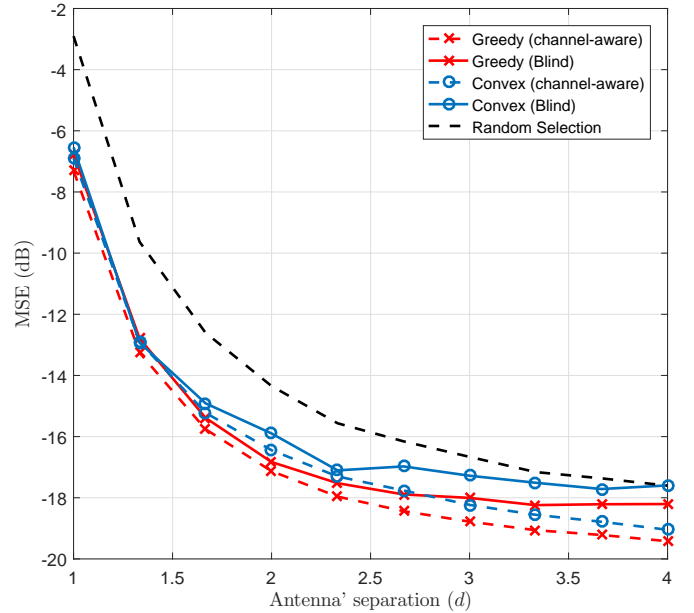

Figure 5: Average MSE achieved by the proposed selection techniques versus the antennas' separation d with $k = 50$.

correlation and vice-versa. As shown in (18), $\mathbf{\Sigma}$ is function of $\mathbf{\Phi}^{-1}$ which presents a slight difference as compared to the
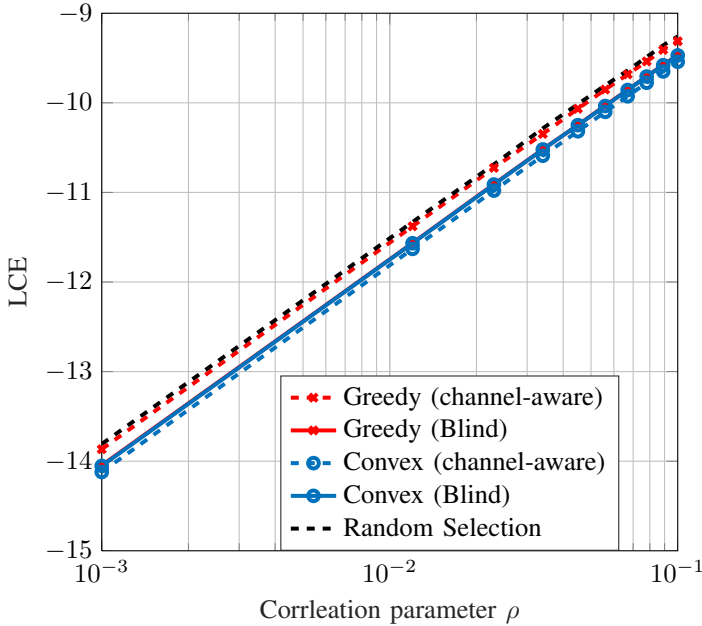
Figure 7: Average LCE achieved by the proposed selection techniques versus the correlation parameter ρ in the case of WSN with $k = 50$.
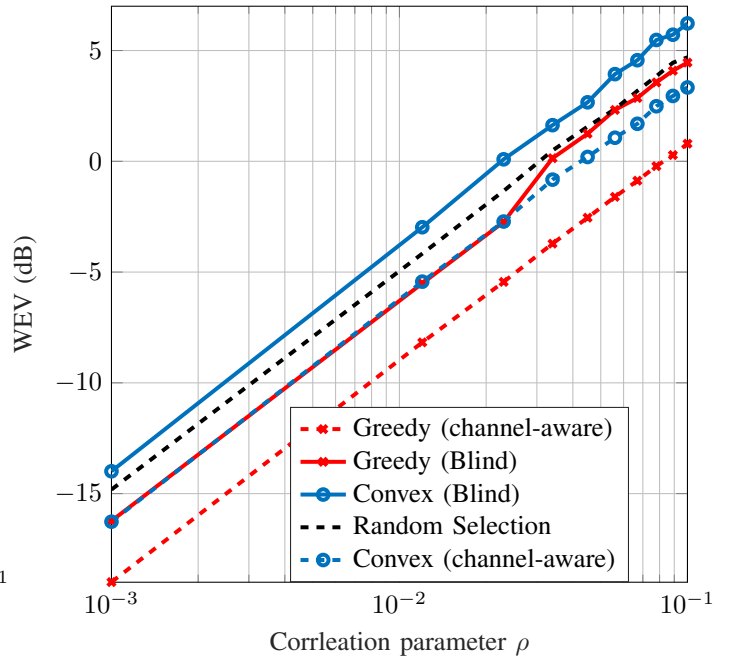

Figure 8: Average WEV in dB achieved by the proposed selection techniques versus the correlation parameter ρ in the case of WSN with $k = 50$.

previous application, where we use $\mathbf{\Phi}^{-1}$ instead of $\mathbf{R}$ [3].

We examine the performance in terms of the MSE as well as the LCE and the WEV. As shown in Figure 6, similar observations to the application of massive MIMO can be conducted. We notice that increasing the correlation ($\rho \downarrow$) results in a better performance of the proposed blind selection algorithms as compared to the random selection algorithm and vice versa. We also observe that the proposed blind convex approach fails to achieve a good performance for both the MSE and the WEV and yields a performance that is worse than that of random selection. This is due to the same quantization effect as explained in the case of massive MIMO. In Figures 7 and 8, we show the performance of the LCE and the WEV against the correlation parameter ρ respectively. Figures 7 and 8 clearly show that the performance for both the LCE and the WEV improves with increasing the correlation ($\rho \downarrow$) for the proposed blind approach. This is suitable in such application, since a central node can perform sensor selection without knowledge of the channel matrix which may require a huge overhead.

V. CONCLUSION

In this paper, we introduced blind techniques for measurement selection. In particular, we showed that using tools from random matrix theory, it is possible to asymptotically approximate error measures that are commonly used in this context. As such, perfect knowledge of the measurement matrix is not needed and only statistics are required to perform measurement selection. We proposed two techniques: the first is based on a greedy approach and the second is based on

a convex relaxation heuristic. The proposed blind selection techniques have been tested in two applications related to wireless communications: the first is antenna selection in uplink multiusers massive MIMO systems and the second is sensor selection in wireless sensor networks. Numerical results showed that the blind techniques have a comparable performance to techniques that require full knowledge of the measurement matrix, especially at high correlation.

REFERENCES

[1] S. Joshi and S. Boyd, "Sensor Selection via Convex Optimization," *IEEE Trans. on Signal Processing*, vol. 57, no. 2, pp. 451–462, Feb 2009.
[2] G. Hovland and B. McCarragher, "Dynamic Sensor Selection for Robotic Systems," in *IEEE Int. Conf. Robotics Automation*, vol. 1, 1997, pp. 272–277.
[3] D. Kammer, "Sensor Placement for On-Orbit Modal Identification and Correlation of Large Space Structures," *J. Guid., Control, Dynam*, vol. 14, pp. 251–259, 1991.
[4] L. Yao, W. Sethares, and D. Kammer, "Sensor Placement for On-Orbit Modal Identification via a Genetic Algorithm," *Amer. Inst. Aeronaut. Astronaut. J.*, vol. 31, p. 19221928, 1993.
[5] N. Nguyen and A. Miller, "A Review of Some Exchange Algorithms for Constructing Discrete D-Optimal Designs," *Comput. Statist. Data Anal.*, vol. 14, pp. 489–498, 1992.
[6] R. John and N. Draper, "D-Optimality for Regression Designs: A review," *Technometrics*, vol. 17, no. 1, pp. 15–23, 1997.
[7] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for Next Generation Wireless Systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.
[8] 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, November 2010.
[9] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling Up MIMO: Opportunities and Challenges with Very Large Arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan 2013.

---

[3]All the derived results concerning the asymptotic equivalents and their convergence can be applied in a straightforward manner to the case $\mathbf{R} = \mathbf{\Phi}^{-1}$.
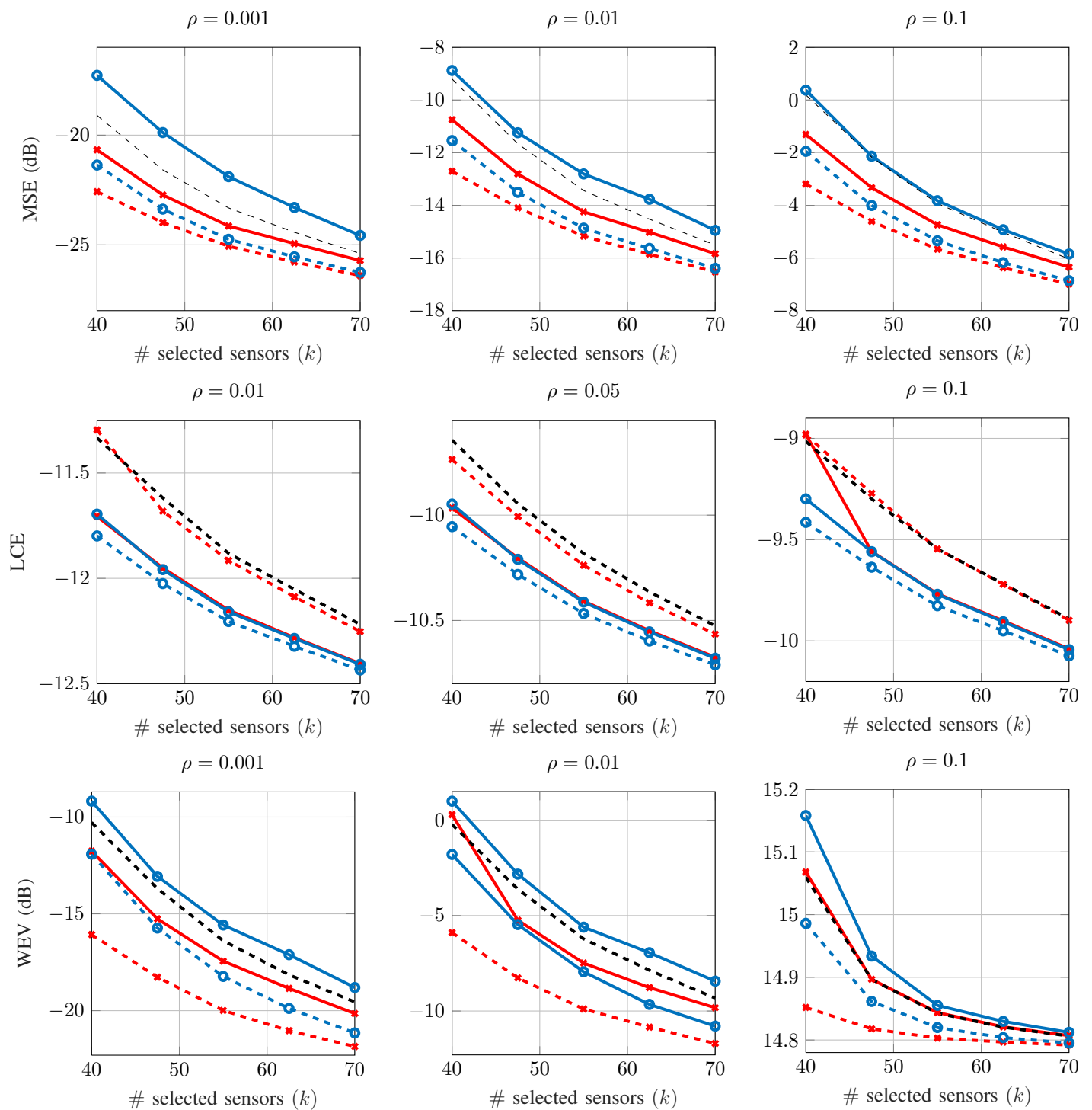
Figure 6: Average error MSE, LCE and WEV) achieved by the proposed selection techniques for different values of the correlation parameter ρ in the WSN scenario.

[10] X. Gao, O. Edfors, J. Liu, and F. Tufvesson, "Antenna Selection in Measured Massive MIMO Channels using Convex Optimization," in *Globecom Workshops (GC Wkshps), 2013 IEEE*, Dec 2013, pp. 129–134.

[11] A. F. Molisch and M. Z. Win, "MIMO Systems with Antenna Selection," *IEEE Microwave Magazine*, vol. 5, no. 1, pp. 46–56, Mar 2004.

[12] S. Sanayei and A. Nosratinia, "Antenna Selection in MIMO Systems," *IEEE Communications Magazine*, vol. 42, no. 10, pp. 68–73, Oct 2004.

[13] L. Yao, W. A. Sethares, and D. C. Kammer, "Sensor Placement for On-orbit Modal Identification via a Genetic Algorithm," *Amer. Inst.*

Aeronaut. Astronaut. J., vol. 31, no. 10, pp. 1922–1928, 1993.

[14] A. Sayed, *Fundamentals of Adaptive Filtering*. John Wiley & Sons, 2003.

[15] R. J. Serfling, *Approximations Theorems of Mathematical Statistics*. John Wiley & Sons, 2002.

[16] 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.

[17] *High SNR Approximations of the Capacity of MIMO Correlated Rician channels: A large System Approach*, 2007.

[18] Z. D. Bai and J. W. Silverstein, "Exact Separation of Eigenvalues of Large Dimensional Sample Covariance Matrices," *The Annals of Probability*, vol. 27, no. 3, pp. 1536–1555, 1999.
[19] S. Sigdel, W. A. Krzymie, and M. Al-Shalash, "Greedy and Progressive User Scheduling for CoMP Wireless Networks," in *Communications (ICC), 2012 IEEE International Conference on*, June 2012, pp. 4218–4223.
[20] G. Alfano, A. M. Tulino, A. Lozano, and S. Verdu, "Capacity of MIMO Channels with One-sided Correlation," *ISSSTA*, August 2004.
[21] S. Liu, S. P. Chepuri, M. Fardad, E. Maazade, G. Leus, and P. K. Varshney, "Sensor Selection for Estimation with Correlated Measurement Noise," *IEEE Transactions on Signal Processing*, vol. 64, no. 13, pp. 3509–3522, July 2016.
[22] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim, "On the Capacity Achieving Covariance Matrix for Rician MIMO Channels: An Asymptotic Approach," *IEEE Transactions on Information Theory*, vol. 56, no. 3, pp. 1048–1069, March 2010.

APPENDIX A
GRADIENT DERIVATION OF THE DIFFERENT DETERMINISTIC EQUIVALENTS

Useful Lemmas

Inversion Lemma: For $\mathbf{A}$ an invertible square matrix and column vector $\mathbf{u}$ such that $1 + \mathbf{u}^T \mathbf{A}^{-1}\mathbf{u} \neq 0$, we have

$$\left(\mathbf{A} + \mathbf{u}\mathbf{u}^T\right)^{-1} = \mathbf{A}^{-1} - \frac{1}{1 + \mathbf{u}^T\mathbf{A}^{-1}\mathbf{u}}\mathbf{A}^{-1}\mathbf{u}\mathbf{u}^T\mathbf{A}^{-1}. \tag{20}$$

$$\det\left(\mathbf{A} + \mathbf{u}\mathbf{u}^T\right) = \det\left(\mathbf{A}\right)\left(1 + \mathbf{u}^T\mathbf{A}^{-1}\mathbf{u}\right). \tag{21}$$

Implicit Function Theorem: Let $f : U \times V \to \mathbb{R}$ be a continuously differentiable function and let $g : U \to V$ be the implicit function defined as follows

$$\left\{\left(\mathbf{x}, g\left(\mathbf{x}\right)\right)|\mathbf{x} \in U\right\} = \left\{\left(\mathbf{x}, \mathbf{y}\right) \in U \times V | f\left(\mathbf{x}, \mathbf{y}\right) = \mathbf{c}\right\}. \tag{22}$$

Then, g is continuously differentiable and

$$\frac{\partial g}{\partial x_i}\left(\mathbf{x}\right) = -\left(\frac{\partial f}{\partial y}\left(\mathbf{x}, g\left(\mathbf{x}\right)\right)\right)^{-1}\frac{\partial f}{\partial x_i}\left(\mathbf{x}, g\left(\mathbf{x}\right)\right). \tag{23}$$

Some Useful Notations
- $\mathbf{R_s} = \mathbf{R}^{\frac{1}{2}}\operatorname{diag}\left(\mathbf{s}\right)\mathbf{R}^{\frac{1}{2}}$.
- $\mathbf{r}_i$ is the ith column of $\mathbf{R}^{\frac{1}{2}}$.
- $\mathbf{\Psi}_i = \mathbf{R_s} - s_i\mathbf{r}_i\mathbf{r}_i^T$.

Gradient of the MSE

For the MSE, we have the following fixed-point equation in terms of δ

$$\delta\operatorname{tr}\left[\mathbf{R_s}\left(\mathbf{I} + \delta\mathbf{R_s}\right)^{-1}\right] - m = 0.$$

Define

$$f\left(\mathbf{s}, y\right) = y\operatorname{tr}\left[\mathbf{R_s}\left(\mathbf{I} + y\mathbf{R_s}\right)^{-1}\right] - m$$
$$= n - m - \operatorname{tr}\left[\left(\mathbf{I} + y\mathbf{R_s}\right)^{-1}\right].$$

Then,

$$\frac{\partial f}{\partial y}\left(\mathbf{s}, y\right) = \operatorname{tr}\left[\mathbf{R_s}\left(\mathbf{I} + y\mathbf{R_s}\right)^{-2}\right].$$

On the other hand

$$\operatorname{tr}\left[\left(\mathbf{I} + y\mathbf{R_s}\right)^{-1}\right] = \operatorname{tr}\left(\mathbf{I} + y\mathbf{\Psi}_i + ys_i\mathbf{r}_i\mathbf{r}_i^T\right)^{-1}$$
$$= \operatorname{tr}\left(\mathbf{I} + y\mathbf{\Psi}_i\right)^{-1} - \frac{ys_i\mathbf{r}_i^T\left(\mathbf{I} + y\mathbf{\Psi}_i\right)^{-2}\mathbf{r}_i}{1 + ys_i\mathbf{r}_i^T\left(\mathbf{I} + y\mathbf{\Psi}_i\right)^{-1}\mathbf{r}_i}$$

Thus,

$$\frac{\partial f}{\partial s_i}\left(\mathbf{s}, y\right) = \frac{y\mathbf{r}_i^T\left(\mathbf{I} + y\mathbf{\Psi}_i\right)^{-2}\mathbf{r}_i}{\left(1 + ys_i\mathbf{r}_i^T\left(\mathbf{I} + y\mathbf{\Psi}_i\right)^{-1}\mathbf{r}_i\right)^2}$$
$$= y\mathbf{r}_i^T\left(\mathbf{I} + y\mathbf{R_s}\right)^{-2}\mathbf{r}_i$$
$$= y\left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I} + y\mathbf{R_s}\right)^{-2}\mathbf{R}^{\frac{1}{2}}\right]_{i,i}$$

Finally,

$$\delta_i' = \frac{\partial\delta}{\partial s_i} = -\frac{\delta\left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I} + \delta\mathbf{R_s}\right)^{-2}\mathbf{R}^{\frac{1}{2}}\right]_{i,i}}{\operatorname{tr}\left[\mathbf{R_s}\left(\mathbf{I} + \delta\mathbf{R_s}\right)^{-2}\right]}.$$

and

$$\frac{\partial\overline{\mathrm{MSE}}\left(\mathbf{s}\right)}{\partial s_i} = -\frac{\delta\left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I} + \delta\mathbf{R_s}\right)^{-2}\mathbf{R}^{\frac{1}{2}}\right]_{i,i}}{\operatorname{tr}\left[\mathbf{R_s}\left(\mathbf{I} + \delta\mathbf{R_s}\right)^{-2}\right]}.$$

Gradient of the LCE

The LCE converges a.s. to the following quantity

$$\overline{\mathrm{LCE}}\left(\mathbf{s}\right) = -\frac{1}{m}\log\det\left(\mathbf{I} + \delta\mathbf{R_s}\right) + \log\left(c\delta\right) + 1$$
$$= -\frac{1}{m}\log\left|\mathbf{I} + \delta\mathbf{\Psi}_i\right|$$
$$- \frac{1}{m}\log\left(1 + \delta s_i\mathbf{r}_i^T\left(\mathbf{I} + \delta\mathbf{\Psi}_i\right)^{-1}\mathbf{r}_i\right)$$
$$+ \log\left(c\delta\right) + 1.$$

We have

$$\frac{\partial}{\partial s_i}\frac{1}{m}\log\left|\mathbf{I} + \delta\mathbf{\Psi}_i\right| = \frac{\delta_i'}{m}\operatorname{tr}\left[\mathbf{\Psi}_i\left(\mathbf{I} + \delta\mathbf{\Psi}_i\right)^{-1}\right].$$

and

$$\frac{\partial}{\partial s_i}\frac{1}{m}\log\left(1 + \delta s_i\mathbf{r}_i^T\left(\mathbf{I} + \delta\mathbf{\Psi}_i\right)^{-1}\mathbf{r}_i\right)$$
$$= \frac{\frac{\delta}{m}\mathbf{r}_i^T\left(\mathbf{I} + \delta\mathbf{\Psi}_i\right)^{-1}\mathbf{r}_i + \frac{\delta_i's_i}{m}\mathbf{r}_i^T\left(\mathbf{I} + \delta\mathbf{\Psi}_i\right)^{-2}\mathbf{r}_i}{1 + \delta s_i\mathbf{r}_i^T\left(\mathbf{I} + \delta\mathbf{\Psi}_i\right)^{-1}\mathbf{r}_i}$$
$$= \frac{\delta}{m}\left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I} + \delta\mathbf{R_s}\right)^{-1}\mathbf{R}^{\frac{1}{2}}\right]_{i,i}$$
$$+ \frac{\delta_i's_i}{m}\frac{\left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I} + \delta\mathbf{R_s}\right)^{-2}\mathbf{R}^{\frac{1}{2}}\right]_{i,i}}{1 - \delta s_i\left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I} + \delta\mathbf{R_s}\right)^{-1}\mathbf{R}^{\frac{1}{2}}\right]_{i,i}}$$

Finally,

$$
\begin{aligned}
\frac{\partial \overline{\mathrm{LCE}}\left(\mathbf{s}\right)}{\partial s_i} &= -\frac{\delta_i'}{m}\,\mathrm{tr}\left[\boldsymbol{\Psi}_i\left(\mathbf{I}+\delta\boldsymbol{\Psi}_i\right)^{-1}\right] \\
&\quad -\frac{\frac{\delta}{m}\mathbf{r}_i^T\left(\mathbf{I}+\delta\boldsymbol{\Psi}_i\right)^{-1}\mathbf{r}_i+\frac{\delta_i's_i}{m}\mathbf{r}_i^T\left(\mathbf{I}+\delta\boldsymbol{\Psi}_i\right)^{-2}\mathbf{r}_i}{1+\delta s_i\mathbf{r}_i^T\left(\mathbf{I}+\delta\boldsymbol{\Psi}_i\right)^{-1}\mathbf{r}_i} \\
&\quad +\frac{\delta_i'}{\delta}. \\
&= -\frac{\delta_i'}{m}\,\mathrm{tr}\left[\boldsymbol{\Psi}_i\left(\mathbf{I}+\delta\boldsymbol{\Psi}_i\right)^{-1}\right] \\
&\quad -\frac{\delta}{m}\left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I}+\delta\mathbf{R}_s\right)^{-1}\mathbf{R}^{\frac{1}{2}}\right]_{i,i} \\
&\quad -\frac{\delta_i's_i}{m}\frac{\left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I}+\delta\mathbf{R}_s\right)^{-2}\mathbf{R}^{\frac{1}{2}}\right]_{i,i}}{1-\delta s_i\left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I}+\delta\mathbf{R}_s\right)^{-1}\mathbf{R}^{\frac{1}{2}}\right]_{i,i}} \\
&\quad +\frac{\delta_i'}{\delta} \\
&= (1-c)\frac{\delta_i'}{\delta}+\frac{\delta_i'}{m\delta}\,\mathrm{tr}\left(\mathbf{I}+\delta\mathbf{R}_s\right)^{-1} \\
&\quad -\frac{\delta}{m}\left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I}+\delta\mathbf{R}_s\right)^{-1}\mathbf{R}^{\frac{1}{2}}\right]_{i,i}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \overline{\mathrm{LCE}}\left(\mathbf{s}\right)}{\partial s_i} &= (1-c)\frac{\delta_i'}{\delta}+\frac{\delta_i'}{m\delta}\,\mathrm{tr}\left(\mathbf{I}+\delta\mathbf{R}_s\right)^{-1} \\
&\quad -\frac{\delta}{m}\left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I}+\delta\mathbf{R}_s\right)^{-1}\mathbf{R}^{\frac{1}{2}}\right]_{i,i}
\end{aligned}
$$

Gradient of the WEV

The WEV converges a.s. to the following quantity

$$
\overline{\mathrm{WEV}}\left(\mathbf{s}\right)=-\frac{1}{\eta}+\frac{1}{m}\,\mathrm{tr}\left[\mathbf{R}_{\mathbf{s}}\left(\mathbf{I}+\eta\mathbf{R}_{\mathbf{s}}\right)^{-1}\right],
$$

where

$$
\eta^2\,\mathrm{tr}\left[\mathbf{R}_{\mathbf{s}}^2\left(\mathbf{I}+\eta\mathbf{R}_{\mathbf{s}}\right)^{-2}\right]-m=0.
$$

Define

$$
\begin{aligned}
f\left(\mathbf{s},y\right)&=y^2\,\mathrm{tr}\left[\mathbf{R}_{\mathbf{s}}^2\left(\mathbf{I}+y\mathbf{R}_{\mathbf{s}}\right)^{-2}\right]-m \\
&= n-m+\mathrm{tr}\left[\left(\mathbf{I}+y\mathbf{R}_{\mathbf{s}}\right)^{-2}\right]-2\,\mathrm{tr}\left[\left(\mathbf{I}+y\mathbf{R}_{\mathbf{s}}\right)^{-1}\right]
\end{aligned}
$$

Then,

$$
\frac{\partial f}{\partial y}\left(\mathbf{s},y\right)=2y\,\mathrm{tr}\left[\mathbf{R}_{\mathbf{s}}^2\left(\mathbf{I}+y\mathbf{R}_{\mathbf{s}}\right)^{-2}\right]-2y^2\,\mathrm{tr}\left[\mathbf{R}_{\mathbf{s}}^3\left(\mathbf{I}+y\mathbf{R}_{\mathbf{s}}\right)^{-3}\right]
$$

Note that

$$
\begin{aligned}
\left(\mathbf{I}+y\mathbf{R}_{\mathbf{s}}\right)^{-1}&=\left(\mathbf{I}+y\boldsymbol{\Psi}_i+ys_i\mathbf{r}_i\mathbf{r}_i^T\right)^{-1} \\
&=\left(\mathbf{I}+y\boldsymbol{\Psi}_i\right)^{-1} \\
&\quad -\frac{ys_i\left(\mathbf{I}+y\boldsymbol{\Psi}_i\right)^{-1}\mathbf{r}_i\mathbf{r}_i^T\left(\mathbf{I}+y\boldsymbol{\Psi}_i\right)^{-1}}{1+ys_i\mathbf{r}_i^T\left(\mathbf{I}+y\boldsymbol{\Psi}_i\right)^{-1}\mathbf{r}_i}
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\mathrm{tr}\left(\mathbf{I}+y\mathbf{R}_{\mathbf{s}}\right)^{-2}&=\mathrm{tr}\left(\mathbf{I}+y\boldsymbol{\Psi}_i\right)^{-2} \\
&\quad +\frac{y^2s_i^2\mathbf{r}_i^T\left(\mathbf{I}+y\boldsymbol{\Psi}_i\right)^{-2}\mathbf{r}_i\mathbf{r}_i^T\left(\mathbf{I}+y\boldsymbol{\Psi}_i\right)^{-2}\mathbf{r}_i}{\left(1+ys_i\mathbf{r}_i^T\left(\mathbf{I}+y\boldsymbol{\Psi}_i\right)^{-1}\mathbf{r}_i\right)^2} \\
&\quad -2\frac{ys_i\mathbf{r}_i^T\left(\mathbf{I}+y\boldsymbol{\Psi}_i\right)^{-3}\mathbf{r}_i}{1+ys_i\mathbf{r}_i^T\left(\mathbf{I}+y\boldsymbol{\Psi}_i\right)^{-1}\mathbf{r}_i}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\frac{\partial f}{\partial s_i}\left(\mathbf{s},y\right)&=\frac{2y\mathbf{r}_i^T\left[\left(\mathbf{I}+y\boldsymbol{\Psi}_i\right)^{-2}-\left(\mathbf{I}+y\boldsymbol{\Psi}_i\right)^{-3}\right]\mathbf{r}_i}{\left(1+ys_i\mathbf{r}_i^T\left(\mathbf{I}+y\boldsymbol{\Psi}_i\right)^{-1}\mathbf{r}_i\right)^2} \\
&\quad +\frac{2y^2s_i\left[\mathbf{r}_i^T\left(\mathbf{I}+y\boldsymbol{\Psi}_i\right)^{-2}\mathbf{r}_i\right]^2}{\left(1+ys_i\mathbf{r}_i^T\left(\mathbf{I}+y\boldsymbol{\Psi}_i\right)^{-1}\mathbf{r}_i\right)^3}.
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\eta_i' &= \frac{\partial \eta}{\partial s_i} \\
&= -\left(2\eta\,\mathrm{tr}\left[\mathbf{R}_{\mathbf{s}}^2\left(\mathbf{I}+\eta\mathbf{R}_{\mathbf{s}}\right)^{-2}\right]-2\eta^2\,\mathrm{tr}\left[\mathbf{R}_{\mathbf{s}}^3\left(\mathbf{I}+\eta\mathbf{R}_{\mathbf{s}}\right)^{-3}\right]\right)^{-1} \\
&\quad \times\left[\frac{2\eta\mathbf{r}_i^T\left[\left(\mathbf{I}+\eta\boldsymbol{\Psi}_i\right)^{-2}-\left(\mathbf{I}+\eta\boldsymbol{\Psi}_i\right)^{-3}\right]\mathbf{r}_i}{\left(1+\eta s_i\mathbf{r}_i^T\left(\mathbf{I}+\eta\boldsymbol{\Psi}_i\right)^{-1}\mathbf{r}_i\right)^2}\right. \\
&\quad \left.+\frac{2\eta^2s_i\left[\mathbf{r}_i^T\left(\mathbf{I}+\eta\boldsymbol{\Psi}_i\right)^{-2}\mathbf{r}_i\right]^2}{\left(1+\eta s_i\mathbf{r}_i^T\left(\mathbf{I}+\eta\boldsymbol{\Psi}_i\right)^{-1}\mathbf{r}_i\right)^3}\right] \\
&= -\left(\mathrm{tr}\left[\mathbf{R}_{\mathbf{s}}\left(\mathbf{I}+\eta\mathbf{R}_{\mathbf{s}}\right)^{-2}\right]-\mathrm{tr}\left[\mathbf{R}_{\mathbf{s}}\left(\mathbf{I}+\eta\mathbf{R}_{\mathbf{s}}\right)^{-3}\right]\right)^{-1} \\
&\quad \times\eta\left(\left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I}+\eta\mathbf{R}_s\right)^{-2}\mathbf{R}^{\frac{1}{2}}\right]_{i,i}-\left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I}+\eta\mathbf{R}_s\right)^{-3}\mathbf{R}^{\frac{1}{2}}\right]_{i,i}\right)
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\frac{\partial \overline{\mathrm{WEV}}\left(\mathbf{s}\right)}{\partial s_i}&=-\frac{(c-1)\eta_i'}{\eta^2}+\frac{\eta_i'}{m\eta^2} \\
&\quad \times\left[2\,\mathrm{tr}\left(\mathbf{I}+\eta\mathbf{R}_s\right)^{-1}-\mathrm{tr}\left(\mathbf{I}+\eta\mathbf{R}_s\right)^{-2}\right] \\
&\quad +\frac{1}{m}\left[\mathbf{R}^{\frac{1}{2}}\left(\mathbf{I}+\eta\mathbf{R}_s\right)^{-2}\mathbf{R}^{\frac{1}{2}}\right]_{i,i}
\end{aligned}
$$

APPENDIX B
PROOF OF THEOREM 1

We prove the convexity of the $\overline{\mathrm{MSE}}$ using a similar argument as the one proposed in [22]. Let p be a non negative integer, and $\widetilde{\mathbf{W}}$ be a $np \times mp$ matrix with i.i.d $\mathcal{CN}\left(0,1\right)-$ distributed entries. For $\mathbf{s}\in\mathbb{R}_+^n$, let $\mathbf{R}\left(\mathbf{s}\right)=\mathbf{R}^{\frac{1}{2}}\,\mathrm{diag}\left(\mathbf{s}\right)\mathbf{R}^{\frac{1}{2}}$ and

$$
\widetilde{\mathbf{R}}\left(\mathbf{s}\right)=\mathbf{I}_p\otimes\mathbf{R}\left(\mathbf{s}\right),
$$

where $\otimes$ is the Kronecker product of matrices. Then, $\widetilde{\mathbf{R}}$ is a $np \times np$ positive definite block diagonal matrix. Since p affects both dimensions m and n at the same pace, we have the following convergence

$$
\lim_{p\to\infty}\mathrm{tr}\left[\left(\widetilde{\mathbf{W}}^H\widetilde{\mathbf{R}}\left(\mathbf{s}\right)\widetilde{\mathbf{W}}\right)^{-1}\right]=\widetilde{\mathrm{MSE}}\left(\mathbf{s}\right), \tag{24}
$$

where $\widetilde{\text{MSE}}\left(\widetilde{\mathbf{R}}\left(\mathbf{s}\right)\right) = \widetilde{\delta}\left(\mathbf{s}\right)$ is solution to the following fixed-point equation

$$\widetilde{\delta}\left(\mathbf{s}\right) = \frac{mp}{\text{tr}\left[\widetilde{\mathbf{R}}\left(\mathbf{s}\right)\left(\mathbf{I} + \widetilde{\delta}\left(\mathbf{s}\right)\widetilde{\mathbf{R}}\left(\mathbf{s}\right)\right)^{-1}\right]}. \qquad (25)$$

Using the block-diagonal structure of $\widetilde{\mathbf{R}}\left(\mathbf{s}\right)$, we have

$$\left(\mathbf{I} + \widetilde{\delta}\left(\mathbf{s}\right)\widetilde{\mathbf{R}}\left(\mathbf{s}\right)\right)^{-1} = \mathbf{I}_p \otimes \left(\mathbf{I} + \widetilde{\delta}\left(\mathbf{s}\right)\mathbf{R}\left(\mathbf{s}\right)\right)^{-1}.$$

Then,

$$\begin{aligned}
\widetilde{\mathbf{R}}\left(\mathbf{s}\right)\left(\mathbf{I} + \widetilde{\delta}\left(\mathbf{s}\right)\widetilde{\mathbf{R}}\left(\mathbf{s}\right)\right)^{-1} &= \left[\mathbf{I}_p \otimes \mathbf{R}\left(\mathbf{s}\right)\right] \\
&\quad \times \left[\mathbf{I}_p \otimes \left(\mathbf{I} + \widetilde{\delta}\left(\mathbf{s}\right)\mathbf{R}\left(\mathbf{s}\right)\right)^{-1}\right] \\
&= \mathbf{I}_p \otimes \left[\mathbf{R}\left(\mathbf{s}\right)\left(\mathbf{I} + \widetilde{\delta}\left(\mathbf{s}\right)\mathbf{R}\left(\mathbf{s}\right)\right)^{-1}\right]
\end{aligned}$$

Thus,

$$\text{tr}\left[\widetilde{\mathbf{R}}\left(\mathbf{s}\right)\left(\mathbf{I} + \widetilde{\delta}\left(\mathbf{s}\right)\widetilde{\mathbf{R}}\left(\mathbf{s}\right)\right)^{-1}\right] = p\,\text{tr}\left[\mathbf{R}\left(\mathbf{s}\right)\left(\mathbf{I} + \widetilde{\delta}\left(\mathbf{s}\right)\mathbf{R}\left(\mathbf{s}\right)\right)^{-1}\right]$$

and,

$$\begin{aligned}
\widetilde{\delta}\left(\mathbf{s}\right) &= \frac{m}{\text{tr}\left[\mathbf{R}\left(\mathbf{s}\right)\left(\mathbf{I} + \widetilde{\delta}\left(\mathbf{s}\right)\mathbf{R}\left(\mathbf{s}\right)\right)^{-1}\right]} \\
&= \overline{\text{MSE}}\left(\mathbf{s}\right).
\end{aligned}$$

In other words,

$$\lim_{p\to\infty} \text{tr}\left[\left(\widetilde{\mathbf{W}}^H\widetilde{\mathbf{R}}\left(\mathbf{s}\right)\widetilde{\mathbf{W}}\right)^{-1}\right] = \overline{\text{MSE}}\left(\mathbf{s}\right). \qquad (26)$$

As a matter of fact, $\overline{\text{MSE}}\left(\mathbf{s}\right)$ is the pointwise limit of convex functions, which implies that it is also convex. The proof of convexity of the LCE and the WEV follow using the same argument. This concludes the proof of Theorem 1.