Finite-Sample Linear Filter Optimization in Wireless Communications and Financial Systems

Mengyi Zhang, *Student Member, IEEE*, Francisco Rubio, *Member, IEEE*, Daniel P. Palomar, *Fellow, IEEE*, and Xavier Mestre, *Senior Member, IEEE*

***Abstract*—We study the problem of linear filter optimization with finite sample size, which has wide applications such as beamformer design in wireless communications and portfolio optimization in finance. Traditional methods in both fields are not robust against the imprecise channel vector and the noise covariance matrix (or the mean return and the covariance of assets in finance) due to finite sample size. We consider estimation errors both in the channel vector and the noise covariance matrix (or the mean return and the covariance) simultaneously. We resort to high-dimensional asymptotics to account for the fact that the observation dimension is of the same order of magnitude as the number of samples, and use the diagonal loading method (or the shrinkage estimator) to improve the robustness. The channel vector (or mean return) and the noise covariance matrix are estimated from the training data, and then corrected under several widely-used criteria. In an asymptotic setting where the number of samples is comparable to the observation dimension, we obtain linear filters that are as good as the optimal filters with a shrinkage structure and a perfect channel vector (or mean return) under different criteria. Monte Carlo simulations show the advantage of our linear filters in the finite sample size regime.**

***Index Terms*—Diagonal loading, shrinkage, finite sample size, imprecise channel vector, covariance matrix estimation, random matrix theory.**

I. Introduction

Robust adaptive beamforming is a classic and continuously developing topic in array signal processing applications such as communications, sonar, and radar [1], [2]. In financial systems, portfolio optimization has a rich history of theoretical research and practical applications [3] that dates back to Markowitz's work [4]. The problems of robust beamforming and portfolio selection are regarded as linear filter optimization. In these applications, the linear filters (i.e., the beamformer and the portfolio) are optimized to satisfy some criteria, e.g., maximum signal-to-noise ratio (SNR) in array signal processing and the corresponding Sharpe ratio in financial applications, which is the ratio of the mean and the standard deviation.

The filter design problem in wireless communications and financial systems can be analyzed together since they share similar signal models. In wireless communications, the received signal is the transmitted signal multiplied by the channel vector plus multi-dimensional noise. In financial systems, the asset return can be modeled as the mean asset return plus some volatility component. The optimal filter is constructed with the channel vector (or the mean return) and the noise covariance matrix (or asset volatilities), and the specific form varies according to the criteria. In the stationary case, i.e., when the channel vector and the noise covariance in wireless communication (or the mean and covariance of assets in finance) remains constant in the training period and evaluation period, we can construct the filter based on the in-sample observations and then apply it to the out-of-sample signal.

However, the optimal filter cannot be obtained in practice since the channel vector and the noise covariance matrix (or the mean and covariance of assets) are unknown and have to be estimated. The accuracy of their estimation directly affects the performance of the filter. In wireless communications, if an infinite number of samples are available and the channel vector is perfectly known, the traditional minimum variance (MV) beamformer is optimal since it maximizes the signal-to-noise ratio (SNR) at the output of the beamformer. But if the number of samples available at the receiver is not sufficiently high, the traditional MV beamformer based on sample matrix inversion (SMI) is known to have a detrimental effect on the performance since there is mismatch between the true correlation matrix of observations and the sample correlation matrix. Moreover, the traditional MV beamformer lacks robustness against even small mismatches in the desired channel vector. In wireless communications, for instance, the imprecise knowledge of either the sample correlation matrix or the desired channel vector leads to the "signal cancellation effect", sometimes making it even worse than the traditional phased array [5]. In financial systems, the Markowitz's framework provides an efficient frontier as a tradeoff between the mean return and risk of the portfolio. Traditionally, the sample means and covariances of the asset returns are used to implement these portfolios based on the true mean and covariance, but due to estimation error, these estimated portfolios typically perform poorly.

A. Literature Review

In array signal processing, one of the most popular solutions to compensate this error consists in expressing the covariance

Manuscript received April 09, 2013; revised July 28, 2013; accepted July 29, 2013. Date of publication August 08, 2013; date of current version September 09, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Maria Sabrina Greco. This work was supported by the Hong Kong RGC 617911 Research Grant.

M. Zhang is with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong (e-mail: zmyannie@ust.hk).

F. Rubio is a Quantitative Analyst at Genetic Finance, Ltd., Hong Kong (e-mail: franciscoerubio@gmail.com).

D. P. Palomar is with the Hong Kong University of Science and Technology (HKUST). He is also a consultant with Centre Tecnològic de Telecomnicacions de Catalunya—Hong Kong (CTTC-HK), Hong Kong (e-mail: palomar@ust.hk).

X. Mestre is with the Centre Tecnològic de Telecomnicacions de Catalunya (CTTC), Castelldefels, Barcelona 08860, Spain (e-mail: xavier.mestre@cttc.es).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2013.2277835

matrix as a linear combination of the sample covariance matrix and the identity matrix, which is referred to as the diagonal loading method [6], [7]. It has been shown that the diagonal loading method can improve the robustness against mismatches caused by the imprecise channel vector and the noise covariance matrix [8], [9]. In financial systems, the corresponding method is regarded as the shrinkage estimator [10].

There are some limitations in the literature work employing diagonal loading or shrinkage in both fields of wireless communications and finance. Most of them focus only on one aspect of the estimation, either the channel vector (mean returns) or the noise covariance matrix. For example, the robust MV beamformers proposed in [5] and [11] incorporated uncertainty constraints on the channel vector but did not consider the finite sample size effect on the covariance matrix. The presence of random steering vector was considered in [12] and a generalized loading method was applied. In other papers such as [9], the main focus was to deal with the finite sample size effect, and a perfect channel vector was assumed. Both types of mismatches were considered and handled in [13] in a deterministic way, i.e., the worst-case design. However, the performance could be affected by improper uncertainty set modeling and the choice of some parameters. In financial applications, the minimum variance portfolio was considered in [14] instead of the general mean-variance portfolio, where the mean return was not formulated. It was proposed in [15] to estimate the covariance matrix in the criterion of the quadratic loss with respect to the true covariance matrix, and it was proposed in [16] to estimate the inverse of the covariance matrix directly in the criteria of quadratic loss and Stein's loss, but these methods are suboptimal compared to the method of directly estimating the portfolio according to the criterion of Sharpe ratio, which is the ultimate objective. Moreover, constraints on portfolio norms are included in [17] in addition to the Markowitz's framework, however, the threshold cannot be determined properly before we have the out-of-sample data and its improper choice would lead to bad performance of the portfolio.

B. *Methodology and Contributions*

In this paper, we develop a unified method to estimate the optimal linear filter for both multiantenna array signals and financial assets that provide significant improvements with respect to other methods published in the literature. We handle both estimation error in the channel vector and the noise covariance (or the mean and covariance of assets in finance) simultaneously. In wireless communications, we assume the transmitted signal is known to the receiver; in practice, it could be a training sequence or information data estimated through a feedback loop. The channel vector and the noise covariance matrix are estimated with the transmitted signal and the observations, and then corrected with diagonal loading. In financial applications, we directly define the portfolio in terms of the sample mean and covariance, and then form the covariance matrix estimator which is a linear combination of the sample covariance and the identity matrix. We select the diagonal loading factor (or equivalently, the shrinkage factor) based on random matrix theory: to reflect the fact that the sample size is comparable to the dimension of the observations, we employ high-dimensional asymptotics where both of them go to infinity. We derive the convergence of the two types of coupled errors and then correct them in the asymptotic regime. With the observable estimators of the criteria quantities, we can estimate the optimal factors and finally obtain the linear filters.

Our main contributions can be summarized as follows: We handle robustness issues with respect to both the channel vector and the noise covariance matrix (or the mean and covariance of assets in finance) together. In an asymptotic setting where the number of samples is comparable to the observation dimension, we obtain linear filters that are as good as the optimal shrinkage filters under different criteria.

C. *Organization of the Paper*

The rest of the paper is organized as follows. Section II is devoted to the signal models in wireless communications and finance as well as the estimation of the linear filter. Important theoretical results on the asymptotic equivalents and consistent estimators of the criteria quantities for the linear filter are provided in Section III. In Section IV, numerical results in wireless communications and financial systems are presented, where both synthetic data and the real market data are tested. Section V concludes the paper and all technical details and derivations are relegated to the appendices.

D. *Notation*

In this paper, $s, \mathbf{x}, \mathbf{M}$ denote scalars, vectors, and matrices, respectively. The superscripts $(\cdot)^T$ and $(\cdot)^H$ denote, respectively, the transpose and the conjugate transpose. The trace of $\mathbf{M}$ is denoted by $\mathrm{tr}[\mathbf{M}]$ and the mathematical expectation operator is denoted by $\mathsf{E}(\cdot)$. $\mathbb{R}$ and $\mathbb{C}$ denote the real and complex fields of dimension specified by superscripts. $\|\mathbf{x}\|$ is the Euclidean norm of a vector $\mathbf{x}$; for a matrix $\mathbf{M}$, $\|\mathbf{M}\|_2 = \sqrt{\lambda_{max}(\mathbf{M}^H\mathbf{M})}$ denotes the spectral norm; $\|\mathbf{M}\|_F = \sqrt{\mathrm{tr}[\mathbf{M}^H\mathbf{M}]}$ denotes the Frobenius norm; and $\|\mathbf{M}\|_{\mathrm{tr}} = \mathrm{tr}[(\mathbf{M}^H\mathbf{M})^{1/2}]$ denotes the trace norm. Given two quantities a and b, $a \asymp b$ denotes they are asymptotic equivalents, i.e., $|a-b| \to 0$ almost surely. Finally, $\xi, \bar{\xi}$, and $\hat{\xi}$ denote a certain random scalar quantity, its asymptotic equivalent, and its consistent estimator, respectively.

II. Problem Formulation

A. *Signal Model*

We consider the following discrete-time linear model. Let $\mathbf{y}(n) \in \mathbb{C}^{M\times 1}$, $n = 1,\dots,N$, denote a collection of received signal. At snapshot n, it can be expressed as

$$\mathbf{y}(n) = \mathbf{h}s(n) + \mathbf{n}(n). \tag{1}$$

Here $s(n)$ is the input signal, $\mathbf{h} \in \mathbb{C}^{M\times 1}$ is the channel, and $\mathbf{n}(n) \in \mathbb{C}^{M\times 1}$ is the random noise which is assumed to be temporally independent and identical distributed (i.i.d.) with mean zero and a covariance matrix $\mathbf{R}_n$. Without loss of generality, we can assume $\mathsf{E}(|s(n)|^2) = 1$. From a statistical perspective, $\mathbf{y}(n)$ is a random process with mean $\mathbf{h}s(n)$ and covariance $\mathbf{R}_n$.

The above signal model can be applied to both fields of wireless communication and finance. In wireless communication, the model in (1) is complex-valued: $s(n)$ represents the transmitted signal, $\mathbf{h}$ is the channel vector, and $\mathbf{y}(n)$ is the received signal. (We do not assume that the channel vector $\mathbf{h}$ is perfectly known, which is different from [5] and [9]. Instead, we estimate it from the observations and the training signal.) On the other hand, in financial applications, the model in (1) is real-valued: $s(n)$ does not play any role and can be regarded as $1, \mathbf{y}(n)$ is the vector of asset returns, $\mathbf{h}$ is the mean return, and $\mathbf{n}(n)$ is the risky component. Usually the mean return $\mathbf{h}$ is unknown and has to be estimated from the asset returns $\mathbf{y}(n)$. We consider a general complex-valued case since the model and the theory can be applied to both real and complex cases. If the samples are real-valued, the same results hold.

We then process the received signal $\mathbf{y}(n)$ through a linear filter $\mathbf{w}$ and the output is

$$x(n) = \mathbf{w}^H \mathbf{y}(n). \tag{2}$$

In wireless communications, $\mathbf{w}$ is the beamformer and the processed signal $x(n)$ is the estimate of the transmitted signal $s(n)$. In financial applications, components of $\mathbf{w}$ represent the money invested in respective assets, and $x(n)$ is the portfolio return.

We will focus on the problem of designing $\mathbf{w}$ based on different criteria through first- and second-order statistics. In the following subsection, we introduce several widely-used criteria in wireless communications and finance.

B. Design Criteria

1) Criteria in Wireless Communications: In wireless communications, the minimum variance (MV) beamformer obtains a linear filter which is the optimal solution of the following problem [2]:

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^H \mathbf{R}_y \mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^H \mathbf{h} = 1 \end{aligned} \tag{3}$$

where $\mathbf{R}_y = \mathsf{E}[\mathbf{y}(n)\mathbf{y}(n)^H]$ is the correlation matrix of the observations. In (3), the correlation matrix of the observations can also be replaced with the noise covariance matrix $\mathbf{R}_n$ [18] due to the identity $\mathbf{R}_y = \mathbf{R}_n + \mathbf{h}\mathbf{h}^H$, and the formulation is equivalent to the original one. In the following, we mainly focus on the noise covariance matrix $\mathbf{R}_n$, and the objective of (3) becomes

$$\mathsf{VAR} = \mathbf{w}^H \mathbf{R}_n \mathbf{w}, \tag{4}$$

which is the variance of the signal at the output of the beamformer. Moreover, the minimum mean-square error (MMSE) receiver is also widely considered in wireless communications [19], which has the following objective:

$$\begin{aligned} \mathsf{MSE} &= \mathsf{E}(|s(n) - \mathbf{w}^H \mathbf{y}(n)|^2) \\ &= |1 - \mathbf{w}^H \mathbf{h}|^2 + \mathbf{w}^H \mathbf{R}_n \mathbf{w}. \end{aligned} \tag{5}$$

Both the minimum variance beamformer and the linear MMSE receiver optimize the signal-to-noise ratio (SNR), i.e.,

$$\mathsf{SNR} = \frac{\mathbf{w}^H \mathbf{h} \mathsf{E}(|s(n)|^2) \mathbf{h}^H \mathbf{w}}{\mathbf{w}^H \mathbf{R}_n \mathbf{w}} = \frac{|\mathbf{w}^H \mathbf{h}|^2}{\mathbf{w}^H \mathbf{R}_n \mathbf{w}}. \tag{6}$$

Loosely speaking, criteria (3)–(6) are equivalent, since the optimal solutions are scaled versions of each other. The optimal solutions can be written in the following unified form:

$$\mathbf{w}^\star = \kappa \mathbf{R}_n^{-1} \mathbf{h}, \quad \kappa = \begin{cases} \frac{1}{\mathbf{h}^H \mathbf{R}_n^{-1} \mathbf{h}} & \text{for MV} \\ \frac{1}{1 + \mathbf{h}^H \mathbf{R}_n^{-1} \mathbf{h}} & \text{for MMSE} \\ \text{any constant} & \text{for max SNR.} \end{cases} \tag{7}$$

The linear filters in (7) are referred to as clairvoyant, since they are constructed using $\mathbf{R}_n$ and $\mathbf{h}$, which are unknown in practice.

2) Criteria in Finance: In financial systems, the returns of M financial assets in terms of the mean return and covariance is modeled as the following vector stochastic process [20]

$$\mathbf{y}(n) = \mathbf{h} + \mathbf{n}(n), \tag{8}$$

where $\mathbf{h}$ is the mean return and $s(n) = 1$ corresponding to the general model in (1). The design criteria of optimal portfolio is discussed in what follows[1].

If a very conservative investor wants to invest a portfolio with the least amount of risk, and does not care about the return, then the design results in the global minimum variance portfolio (GMVP) problem [3] (note that it is the same as in (3)):

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^T \mathbf{R}_n \mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^T \mathbf{1} = 1. \end{aligned} \tag{9}$$

The GMVP problem in (9) disregards the return. A general mean-variance portfolio optimization problem that incorporates the return constraint in (9) is [10]

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^T \mathbf{R}_n \mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^T \mathbf{h} = h_p \\ & \mathbf{w}^T \mathbf{1} = 1 \end{aligned} \tag{10}$$

where h_p is the expected return threshold and different values of h_p generate portfolios on the efficient frontier [2]. Moreover, the Sharpe ratio (SR) is also a popular criterion to measure the ratio between the mean and the standard deviation [21] (disregarding the risk free asset [3]):

$$\mathsf{SR} = \frac{\mathbf{w}^T \mathbf{h}}{\sqrt{\mathbf{w}^T \mathbf{R}_n \mathbf{w}}}, \tag{11}$$

which is the square root of the SNR in (6) for real-valued $\mathbf{w}, \mathbf{h}$, and $\mathbf{R}_n$.

The unified form of optimal portfolio for criteria (9)–(11) is as follows:

$$\mathbf{w}^\star = \kappa_1 \mathbf{R}_n^{-1} \mathbf{1} + \kappa_2 \mathbf{R}_n^{-1} \mathbf{h}, \tag{12}$$

[1]In the formulations (9)–(11), we allow for short selling, i.e., the portfolio weight $\mathbf{w}_i$ can be negative.

[2]The efficient frontier is the curve denoting the highest return for each given risk.

[3]When we disregard the risk-free asset, the Sharpe ratio coincides with the original definition of the information ratio that is simply the mean over the standard deviation of a series of measurements.

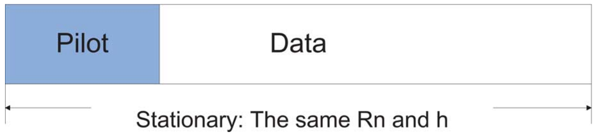

Fig. 1. Pilot signal and information data in wireless communications.

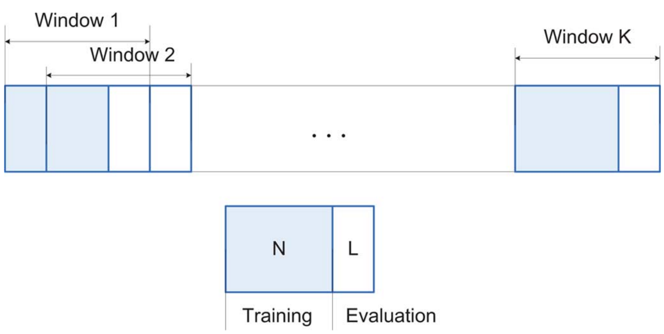

Fig. 2. The rolling-window procedure: In the kth window, the first N observations are used as the training samples and the following L observations are used for evaluation. The window shifts until the end of the data set.

with

$$\begin{cases} \kappa_1 = 1/A, & \kappa_2 = 0 & \text{for GMVP} \\ \kappa_1 = \frac{C - h_p B}{AC - B^2}, & \kappa_2 = \frac{h_p A - B}{AC - B^2} & \text{for mean} - \text{var} \\ \kappa_1 = 0, & \kappa_2 = \text{any constant} & \text{for max SR} \end{cases} \tag{13}$$

where $A = \mathbf{1}^T \mathbf{R}_n^{-1} \mathbf{1}$, $B = \mathbf{1}^T \mathbf{R}_n^{-1} \mathbf{h}$ and $C = \mathbf{h}^T \mathbf{R}_n^{-1} \mathbf{h}$. Similarly, the portfolios in (12) are referred to as clairvoyant, since they are constructed using $\mathbf{R}_n$ and $\mathbf{h}$, which are unknown in practice.

C. The Procedure for Designing $\mathbf{w}$

In the beamforming problem, the data are divided into two parts: the training period and the evaluation period. We use a deterministic and known signal (pilot) in the training period to construct such a beamformer, and then apply this beamformer to the observations that contain the random signal (data). This is illustrated in Fig. 1 using the wireless communication example. It is assumed that channel and noise are stationary, i.e., the pilot and data share that same $\mathbf{R}_n$ and $\mathbf{h}$. Since the optimal linear filter $\mathbf{w}^\star$ depends only on $\mathbf{R}_n$ and $\mathbf{h}$, we first use the pilot signal in the training period to construct an estimator of $\mathbf{w}^\star$, then apply it to the observations that contain the information data.

In the portfolio optimization problem, the principle is different since there is no transmitted signal. But there are still training period and evaluation period. We use the history of the returns as the training period to construct the portfolio, and then apply the portfolio in the evaluation period.

In practice, the above process can be implemented using a rolling-window approach, so that the most recent observations are used for mean return and the covariance estimation. The rolling-window procedure is illustrated in Fig. 2. In the kth window, we use the first N observations as the training samples to construct the linear filter, and then apply the linear filter to the following L observations and evaluate its performance. The window moves forward until the evaluation period does not overlap with the previous one, so that we obtain the $k+1$th window. The window shifts until the end of the data set (or the frame). This procedure is also applicable to the wireless communications scenario. In this case, if the SNR is reasonably good to begin with, a decision-feedback scheme can be used to estimate the updated training data in the next window based on the observations in the current window.

D. Optimal Solution Estimation

We begin with the estimation of the channel vector and noise covariance matrix (or the mean return and the covariance). In some scenarios of wireless communications, the channel vector is also referred to as the steering vector, which is deterministic and can be known (possibly up to some unknown parameters, such as the signal's Direction of Arrival) based on the geometry of array of antennas. Due to the fact that the mismatch in the presumed channel vector leads to severe performance degradation of the beamformer, we consider the case that the channel vector is estimated with the pilot signal.

First of all, we define the following matrix notations for compactness: let $\mathbf{Y} = [\mathbf{y}(1), \dots, \mathbf{y}(N)], \mathbf{s} = [s(1), \dots, s(N)]^T$ and $\mathbf{N} = [\mathbf{n}(1), \dots, \mathbf{n}(N)]$, then (1) can be written in a matrix form:

$$\mathbf{Y} = \mathbf{h}\mathbf{s}^T + \mathbf{N}. \tag{14}$$

Without loss of generality, we can assume [4] $\frac{1}{N}\|\mathbf{s}\|_2^2 = 1$. If it is not the case, we can absorb that constant in $\mathbf{h}$. We next introduce the structure of the estimated optimal linear filter. We focus on all the three criteria (i.e., (4)–(6)) in wireless communications and only the Sharpe ratio criteria (i.e., (11)) in finance systems, since the minimum variance portfolio is already considered in [14] and is a special case of our results. Moreover, regarding the mean-variance portfolio, we do not give the explicit solution here because it is too involved, but it can be easily obtained with our theoretical results.

Traditionally, we estimate the channel vector or the mean return $\mathbf{h}$ using [5]

$$\hat{\mathbf{h}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{y}(n) s(n)^* = \frac{1}{N}\mathbf{Y}\mathbf{s}^*. \tag{15}$$

Recall that in the portfolio optimization problem, $s(n) = 1$. Then the estimator in (15) coincides with the sample mean of $\mathbf{y}(n)$.

The traditional way to estimate the covariance matrix is the sample covariance matrix (SCM),

$$\hat{\mathbf{R}}_{n,\mathsf{SCM}} = \frac{1}{N}(\mathbf{Y} - \hat{\mathbf{h}}\mathbf{s}^T)(\mathbf{Y} - \hat{\mathbf{h}}\mathbf{s}^T)^H, \tag{16}$$

and the corresponding linear filter $\mathbf{w}_{\mathsf{trd}}$ based on $\hat{\mathbf{R}}_{n,\mathsf{SCM}}$ and $\hat{\mathbf{h}}$ is in the same form as in (7), but replacing the unknown $\mathbf{R}_n$ and $\mathbf{h}$ with $\hat{\mathbf{R}}_{n,\mathsf{SCM}}$ and $\hat{\mathbf{h}}$, respectively. To mitigate the estimation

[4] This assumption is automatically satisfied in portfolio optimization applications since $\mathbf{s} = \mathbf{1}$.

[5] If $\mathbf{h}$ is assumed to follow a Gaussian distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{R}_h$, the MMSE estimator is $\hat{\mathbf{h}} = \mathbf{R}_h(\mathbf{R}_n + N\mathbf{R}_h)^{-1}\mathbf{Y}\mathbf{s}^*$. However, in this paper, we focus on the case of a deterministic steering vector $\mathbf{h}$, and the estimator in (15) is an unbiased estimator.

error of the covariance matrix caused by the finite-sample-size effect, we use the shrinkage estimator:

$$\hat{\mathbf{R}}_n = \alpha\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I}, \tag{17}$$

and the linear filter we propose is:

$$\mathbf{w} = \hat{\kappa}\hat{\mathbf{R}}_n^{-1}\hat{\mathbf{h}}, \tag{18}$$

where $\hat{\kappa}$ is in the same form as in (7) and (12) replacing $\mathbf{R}_n$ and $\mathbf{h}$ with $\hat{\mathbf{R}}_n$ and $\hat{\mathbf{h}}$, respectively.

To calibrate α and ρ in (18), we use directly the VAR, MSE, and SNR expressions in (4)–(6) and SR in (11) as the objectives:

$$\mathsf{SNR} = \frac{|\mathbf{h}^H\hat{\mathbf{R}}_n^{-1}\hat{\mathbf{h}}|^2}{\hat{\mathbf{h}}^H\hat{\mathbf{R}}_n^{-1}\mathbf{R}_n\hat{\mathbf{R}}_n^{-1}\hat{\mathbf{h}}} \tag{19}$$

$$\mathsf{MSE} = \frac{|1+(\hat{\mathbf{h}}-\mathbf{h})^H\hat{\mathbf{R}}_n^{-1}\hat{\mathbf{h}}|^2 + \hat{\mathbf{h}}^H\hat{\mathbf{R}}_n^{-1}\mathbf{R}_n\hat{\mathbf{R}}_n^{-1}\hat{\mathbf{h}}}{(1+\hat{\mathbf{h}}^H\hat{\mathbf{R}}_n^{-1}\hat{\mathbf{h}})^2} \tag{20}$$

$$\mathsf{VAR} = \frac{\hat{\mathbf{h}}^H\hat{\mathbf{R}}_n^{-1}\mathbf{R}_n\hat{\mathbf{R}}_n^{-1}\hat{\mathbf{h}}}{(\hat{\mathbf{h}}^H\hat{\mathbf{R}}_n^{-1}\hat{\mathbf{h}})^2} \tag{21}$$

$$\mathsf{SR} = \frac{\mathbf{h}^T\hat{\mathbf{R}}_n^{-1}\hat{\mathbf{h}}}{\sqrt{\hat{\mathbf{h}}^T\hat{\mathbf{R}}_n^{-1}\mathbf{R}_n\hat{\mathbf{R}}_n^{-1}\hat{\mathbf{h}}}}. \tag{22}$$

For benchmark purposes, we will also consider the performance of a beamformer that uses perfect knowledge of h, namely:

$$\mathbf{w}_{\mathsf{dl-h}} = \hat{\kappa}\hat{\mathbf{R}}_n^{-1}\mathbf{h} \tag{23}$$

where $\hat{\mathbf{R}}_n$ is given in (17) and $\hat{\kappa}$ is in the same form as in (7) and (12) while replacing $\mathbf{R}_n$ with $\hat{\mathbf{R}}_n$.

Ideally, if we knew the true $\mathbf{R}_n$ and $\mathbf{h}$, we could directly select α and ρ to optimize (19)–(22) for each realization of the observations and obtain the corresponding beamformer. However, the quantities in (19)–(22) cannot be evaluated since $\mathbf{R}_n$ and $\mathbf{h}$ are unknown. We tackle this problem using random matrix theory: we consider a practical scenario where the sample size N is comparable to the array dimension M. Mathematically speaking, this is formulated as the asymptotic regime where M and N both go to infinity with certain ratio. We first derive deterministic asymptotic equivalents of (19)–(22), and then provide estimators which also approach the corresponding deterministic quantities in the double limit. Since the consistent estimators of (19)–(22) depend only on the observations $\mathbf{y}(n)$, the transmitted signal $s(n)$, and the parameters (α, ρ), we can maximize them to obtain the optimal (α, ρ) and hence the linear filter.

In the following we will provide the main results on the deterministic equivalents and consistent estimators of the criteria in (19)–(22).

III. Main Results

In this section we will discuss the calibration of α and ρ based on random matrix theory. We first provide the following lemma that gives a simplified form of the sample covariance matrix in (16).

Lemma 1: The sample covariance matrix in (16) can be written as

$$\hat{\mathbf{R}}_{n,\mathsf{SCM}} = \frac{1}{N}\mathbf{N}\mathbf{W}\mathbf{N}^H, \tag{24}$$

where $\mathbf{W} = \mathbf{I} - \frac{1}{N}\mathbf{s}^*\mathbf{s}^T$.

Proof: We substitute $\hat{\mathbf{h}}$ in (15) into the expression in (16), and obtain

$$\hat{\mathbf{R}}_{n,\mathsf{SCM}} = \frac{1}{N}\mathbf{Y}\mathbf{W}\mathbf{W}^H\mathbf{Y}^H \tag{25}$$

where $\mathbf{W} = \mathbf{I} - \frac{1}{N}\mathbf{s}^*\mathbf{s}^T$. Note that $\mathbf{Y} = \mathbf{h}\mathbf{s}^T + \mathbf{N}$, then

$$\mathbf{Y}\mathbf{W} = \mathbf{h}\mathbf{s}^T\mathbf{W} + \mathbf{N}\mathbf{W}. \tag{26}$$

Note that $\mathbf{W}\mathbf{s}^* = 0$ due to $\mathbf{s}^T\mathbf{s}^* = N$, hence the first term on the right hand side of (26) becomes zero. Additionally, $\mathbf{W}\mathbf{W}^H = \mathbf{W}$ yields the result in (24). ■

Similarly, the shrinkage estimator in (17) which is a linear combination of the sample covariance matrix and the identity matrix can be written as $\hat{\mathbf{R}}_n = \alpha\frac{1}{N}\mathbf{N}\mathbf{W}\mathbf{N}^H + \rho\mathbf{I}$.

With the simplified notations, we provide the technical hypotheses and some further definitions in the following subsection.

A. Assumptions and Further Definitions

The following set of assumptions will be maintained throughout the paper.

(A1) Let the spectral norm of $\mathbf{R}_n$ and the Euclidean norm of $\mathbf{h}$ be bounded uniformly in M.

(A2) Let $\mathbf{Z}$ be an $M \times N$ matrix whose elements $\mathbf{Z}_{ij}$ are i.i.d. standardized Gaussian random variables. Then the noise matrix can be written as $\mathbf{N} = \mathbf{R}_n^{1/2}\mathbf{Z}$.

We consider the limiting regime defined by both M and N growing large without bound at the same rate, i.e., $M, N \to \infty$ such that $0 < \liminf c \leq \limsup c < \infty$, with $c = M/N$.

In the rest of the paper, we normalize α to 1 to simplify the notations. The reason is that in the expressions (19)–(22), SNR, VAR, and SR are invariant under the scaling of $\hat{\mathbf{R}}_n$. Therefore, we can normalize α to 1 under these criteria and the performance of the proposed beamformer is the same. Regarding MSE, we can write

$$\hat{\mathbf{R}}_n = \alpha\left(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \frac{\rho}{\alpha}\mathbf{I}\right). \tag{27}$$

In the main theorems we will provide, we can absorb α into ρ, and then $\xi \asymp \bar{\xi}$ implies $\alpha^{-1}\xi \asymp \alpha^{-1}\bar{\xi}$.

Before proceeding to the main theorems in this paper, we introduce some further definitions: we define $\{\delta, \tilde{\delta}\}$ as the unique positive solutions to the following system of equations [16], [14]:

$$\begin{cases} \tilde{\delta} = \frac{1}{N}\mathsf{tr}[\mathbf{W}(\mathbf{I}+\delta\mathbf{W})^{-1}] \\ \delta = \frac{1}{N}\mathsf{tr}[\mathbf{R}_n(\tilde{\delta}\mathbf{R}_n + \rho\mathbf{I})^{-1}] \end{cases} \tag{28}$$

and also

$$\begin{cases} \tilde{\gamma} = \frac{1}{N}\mathsf{tr}[(\mathbf{W}(\mathbf{I}+\delta\mathbf{W})^{-1})^2] \\ \gamma = \frac{1}{N}\mathsf{tr}[(\mathbf{R}_n(\tilde{\delta}\mathbf{R}_n + \rho\mathbf{I})^{-1})^2] \end{cases} \tag{29}$$

which are essential quantities for the asymptotic equivalents and estimators of the constitutive quantities in (19)–(22).

We next decompose the quantities in (19)–(22) into some simpler elementary parts that allow us to simplify the analysis. Let $\mathbf{v} = \frac{1}{N}\mathbf{N}\mathbf{s}^*$, and recall that $\hat{\mathbf{h}} = \frac{1}{N}\mathbf{Y}\mathbf{s}^* = \mathbf{h} + \mathbf{v}$, so that these components can be written as

$$\xi_1 = \mathbf{h}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\mathbf{h} \tag{30}$$

$$\xi_2 = \mathbf{h}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\mathbf{v} \tag{31}$$

$$\xi_3 = \mathbf{v}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\mathbf{v} \tag{32}$$

$$\xi_4 = \mathbf{h}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\mathbf{R}_n(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\mathbf{h} \tag{33}$$

$$\xi_5 = \mathbf{h}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\mathbf{R}_n(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\mathbf{v} \tag{34}$$

$$\xi_6 = \mathbf{v}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\mathbf{R}_n(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\mathbf{v}. \tag{35}$$

With (30)–(35), the VAR, MSE, SNR and SR in (19)–(22) can be written as follows:

$$\mathsf{SNR} = \frac{|\xi_1 + \xi_2|^2}{\xi_4 + \xi_5 + \xi_5^* + \xi_6} \tag{36}$$

$$\mathsf{MSE} = \frac{|1 + \xi_2^* + \xi_3|^2 + \xi_4 + \xi_5 + \xi_5^* + \xi_6}{(1 + \xi_1 + \xi_2 + \xi_2^* + \xi_3)^2} \tag{37}$$

$$\mathsf{VAR} = \frac{\xi_4 + \xi_5 + \xi_5^* + \xi_6}{(\xi_1 + \xi_2 + \xi_2^* + \xi_3)^2} \tag{38}$$

$$\mathsf{SR} = \frac{\xi_1 + \xi_2}{\sqrt{\xi_4 + 2\xi_5 + \xi_6}}. \tag{39}$$

B. Asymptotic Equivalents of the Quantities in (36)–(39)

The following theorem establishes the asymptotic behavior of quantities in (30)–(35), which are essential for the convergence of VAR, MSE, SNR and SR in (36)–(39).

Theorem 1: Define the following deterministic quantities,

$$\bar{\xi}_1 = \mathbf{h}^H(\tilde{\delta}\mathbf{R}_n + \rho\mathbf{I})^{-1}\mathbf{h} \tag{40}$$

$$\bar{\xi}_2 = 0 \tag{41}$$

$$\bar{\xi}_3 = \delta \tag{42}$$

$$\bar{\xi}_4 = \frac{1}{1 - \gamma\tilde{\gamma}}\mathbf{h}^H\mathbf{R}_n^{1/2}(\tilde{\delta}\mathbf{R}_n + \rho\mathbf{I})^{-2}\mathbf{R}_n^{1/2}\mathbf{h} \tag{43}$$

$$\bar{\xi}_5 = 0 \tag{44}$$

$$\bar{\xi}_6 = \frac{\gamma}{1 - \gamma\tilde{\gamma}}. \tag{45}$$

Under Assumptions (A1)–(A2), we have $\xi_i \asymp \bar{\xi}_i,\ i = 1, \ldots, 6$, i.e., the random quantities in (30)–(35) behave as the deterministic quantities (40)–(45) in the double limit.

Proof: See the Appendix. ■

The convergence results of ξ_1 and ξ_4 coincide with the results in [9] and [14]. Theorem 1 enables us to analyze the asymptotic criteria in (36)–(39) because it is enough to analyze their asymptotic equivalents, which are easier to characterize because of their deterministic nature. We will next show that, Theorem 1 helps to derive the estimators of the quantities in (36)–(39). Since the estimated criteria will depend on the sample observations, the training sequence and the shrinkage parameters, we will be able to estimate the shrinkage parameters to optimize the asymptotic criteria. The next subsection will provide the estimators of (36)–(39).

C. Consistent Estimators of the Quantities in (36)–(39)

In order to calibrate the shrinkage parameters, we have to obtain observable estimators of the criteria quantities in (36)–(39). We begin with the following lemma which provides a consistent estimator of δ.

Lemma 2: ([14]) Under Assumptions (A1)–(A2), a consistent estimator of δ, denoted by $\hat{\delta}$, is given by the solution to

$$\frac{1}{N}\mathsf{tr}[\hat{\mathbf{R}}_{n,\mathsf{SCM}}(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}] = \hat{\delta}\frac{1}{N}\mathsf{tr}[\mathbf{W}(\mathbf{I} + \hat{\delta}\mathbf{W})^{-1}]. \tag{46}$$

Note that in [14] and [16], the above result holds for a general positive semidefinite $\mathbf{W}$ with bounded spectral norm. However in our case, it can be simplified further with the particular structure of $\mathbf{W}$. Recall that $\mathbf{W} = \mathbf{I} - \frac{1}{N}\mathbf{s}^*\mathbf{s}^T$, so that $\mathbf{W}$ is a projection matrix with an eigenvalue 1 with multiplicity $N-1$ and an eigenvalue 0 with multiplicity 1. Therefore, letting $D = \frac{1}{N}\mathsf{tr}[\hat{\mathbf{R}}_{n,\mathsf{SCM}}(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}]$, we can show from Lemma 2 that

$$\hat{\delta} = \frac{D/(1 - 1/N)}{1 - D/(1 - 1/N)}. \tag{47}$$

Actually, $\hat{\delta}$ can be simplified as

$$\hat{\delta} = \frac{D}{1 - D}, \tag{48}$$

because the contribution of $1/N$ will vanish in the limit.

Now we discuss the consistent estimators of criteria in (36)–(39). We use the original expressions in (19)–(22) instead of the expressions in (36)–(39), since we do not need to estimate $\xi_1 - \xi_6$ separately.

We first estimate SNR in (19), the other estimators can be obtained with the constituents of the estimator of SNR. We propose to estimate the numerator and denominator of (19) separately. We first deal with the numerator, which is easier because only $\mathbf{h}$ cannot be observed. Let us first consider the asymptotic behavior of the quantity $\hat{\mathbf{h}}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\hat{\mathbf{h}}$. Note that

$$\mathbf{h}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\hat{\mathbf{h}} = \hat{\mathbf{h}}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\hat{\mathbf{h}} - \xi_2 - \xi_3. \tag{49}$$

Moreover, $\hat{\delta}$ is a consistent estimator of ξ_3, since they have the same asymptotic equivalent according to Theorem 1 (42) and Lemma 2 (46). Additionally, $\xi_2 \asymp 0$ as stated in Theorem 1 (41), therefore we conclude that

$$\mathbf{h}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\hat{\mathbf{h}} \asymp \hat{\mathbf{h}}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\hat{\mathbf{h}} - \hat{\delta} \tag{50}$$

and the observable quantity on the right hand side of (50) is a consistent estimator of the quantity in the numerator of (19).

Let us now consider an estimator of the denominator of (19). We first consider the following conventional (plug-in) estimator which replaces the unknown $\mathbf{R}_n$ with $\hat{\mathbf{R}}_{n,\mathsf{SCM}}$, i.e.,

$$\mathsf{DNM}_{\mathsf{cnv}} = \hat{\mathbf{h}}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\hat{\mathbf{R}}_{n,\mathsf{SCM}} \times (\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\hat{\mathbf{h}}. \quad (51)$$

Similarly, we can decompose $\mathsf{DNM}_{\mathsf{cnv}}$ into the following components:

$$\mathsf{DNM}_{\mathsf{cnv}} = \xi_{4,\mathsf{cnv}} + \xi_{5,\mathsf{cnv}} + \xi^*_{5,\mathsf{cnv}} + \xi_{6,\mathsf{cnv}} \quad (52)$$

where

$$\xi_{4,\mathsf{cnv}} = \mathbf{h}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\hat{\mathbf{R}}_{n,\mathsf{SCM}}(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\mathbf{h} \quad (53)$$

$$\xi_{5,\mathsf{cnv}} = \mathbf{h}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\hat{\mathbf{R}}_{n,\mathsf{SCM}}(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\mathbf{v} \quad (54)$$

$$\xi_{6,\mathsf{cnv}} = \mathbf{v}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\hat{\mathbf{R}}_{n,\mathsf{SCM}}(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\mathbf{v}. \quad (55)$$

Regarding the components in (53)–(55), we have the following convergence results.

Theorem 2: Under Assumptions (A1)–(A2), the following convergence results hold true:

$$b\xi_{4,\mathsf{cnv}} \asymp \bar{\xi}_4,\ \xi_{5,\mathsf{cnv}} \asymp 0,\ \ and\ \ b\xi_{6,\mathsf{cnv}} \asymp \bar{\xi}_6, \quad (56)$$

where $b = \frac{1}{\frac{1}{N}\mathsf{tr}[\mathbf{W}(\mathbf{I}+\hat{\delta}\mathbf{W})^{-2}]}$.

Proof: See the Appendix. ■

Theorem 2 shows that $b\mathsf{DNM}_{\mathsf{cnv}}$ is a consistent estimator of the denominator of (19) since they have the same asymptotic equivalent. Similarly, we can derive the estimators of (20)–(22):

$$\mathsf{SNR}_{\mathsf{est}} = \frac{|\hat{\mathbf{h}}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\hat{\mathbf{h}} - \hat{\delta}|^2}{b\mathsf{DNM}_{\mathsf{cnv}}} \quad (57)$$

$$\mathsf{MSE}_{\mathsf{est}} = \frac{|1+\hat{\delta}|^2 + b\mathsf{DNM}_{\mathsf{cnv}}}{(1+\hat{\mathbf{h}}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\hat{\mathbf{h}})^2} \quad (58)$$

$$\mathsf{VAR}_{\mathsf{est}} = \frac{b\mathsf{DNM}_{\mathsf{cnv}}}{(\hat{\mathbf{h}}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\hat{\mathbf{h}})^2} \quad (59)$$

$$\mathsf{SR}_{\mathsf{est}} = \frac{\hat{\mathbf{h}}^T(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\hat{\mathbf{h}} - \hat{\delta}}{\sqrt{b\mathsf{DNM}_{\mathsf{cnv}}}} \quad (60)$$

where $\mathsf{DNM}_{\mathsf{cnv}}$ is given in (51) and b is defined in Theorem 2.

Having derived consistent estimators of the quantities in (57)–(60), one can use exhaustive search to find the optimal ρ.

IV. Simulation Results

In this section we will numerically study the convergence of criteria quantities and the nonasymptotic performance of the proposed beamformers.

A. Convergence

First we investigate the performance of the estimators derived from the main Theorem 2. We compare the original SNR, MSE, and VAR expressions in (19)–(21) and the proposed estimators in (57)–(59) in the nonasymptotic regime. Moreover, we use the plug-in SNR, MSE, and VAR as benchmarks to illustrate the estimation performance of traditional approaches. The plug-in expressions are given by replacing $\mathbf{R}_n$ and $\mathbf{h}$ in (19)–(21) with $\hat{\mathbf{R}}_{n,\mathsf{SCM}}$ and $\hat{\mathbf{h}}$, respectively. Since the SR expression in (22) is the square root of the SNR expression in (19), it is enough to analyze the estimation of SNR expression in (19).

In this set of experiments, we use synthetic data generated according to (A2). We let the entries of $\mathbf{h}$ be complex Gaussian distributed with mean zero and variance one, and assume that $\mathbf{s}$ has norm $\sqrt{N}$. The covariance matrix of noise vector is generated as $\mathbf{R}_n(i,j) = 0.7^{|i-j|}$, which is the covariance matrix of a Gaussian AR process. We consider the diagonal loading beamformer in (18), whose SNR, MSE, and VAR are given in (19)–(21). We fix $\alpha = 1$, and let the range of ρ be $[0,\ 3 \cdot \frac{1}{M}\mathsf{tr}(\hat{\mathbf{R}}_{n,\mathsf{SCM}})]$, here the scalar $\frac{1}{M}\mathsf{tr}(\hat{\mathbf{R}}_{n,\mathsf{SCM}})$ is used to make the sample covariance matrix $\hat{\mathbf{R}}_{n,\mathsf{SCM}}$ and the shrinkage target $\mathbf{I}$ of similar scale. We fix the observation dimension $M = 20$, and show the performance of the estimators with number of samples $N = 10$ and $N = 30$, respectively. We plot the true SNR, MSE, and VAR in (19)–(21), the estimated SNR, MSE, and VAR in (57)–(59), and the plug-in SNR, MSE, and VAR as functions of ρ. Each experiment is repeated 100 times and the averaged quantities are plotted.

It can be seen in Fig. 3(a)–(c) that the traditional plug-in SNR, MSE, and VAR estimates not only deviate significantly from the true values, but always overestimate the performance.

The optimal ρ for the plug-in criteria is 0, which leads to the traditional sample covariance matrix. The proposed estimators of SNR, MSE, and VAR as functions of ρ are pretty close to the true ones when $M = 20, N = 30$. When the number of samples is smaller, i.e., $N = 10$, the proposed estimators deviate from the the true criteria quantities, especially for MSE and VAR. However, the optimal ρ of the estimated criteria does not deviate too much from the optimal ρ of the true criteria. Moreover, it can be seen that the performance of the diagonal loading beamformer is very sensitive to the choice of ρ, especially when ρ is smaller than the optimal one, the SNR and VAR performance degrade seriously. Hence, it is important to chose a proper diagonal loading factor ρ instead of using the sample covariance matrix which corresponds to $\rho = 0$.

B. Simulation in Beamforming

We focus on the beamforming problem in wireless communications with finite sample size settings. In this set of experiments, we let the channel vector $\mathbf{h}$ have complex Gaussian distributed entries with zero mean and variance one, which is unknown to the receiver and estimated from the receiving signal and the training signal. $\mathbf{s}$ is the same as in the previous subsection, but the entries of $\mathbf{R}_n$ are divided by the transmitting SNR, i.e., $\mathbf{R}_n(i,j) = \sigma_n^2 \cdot 0.7^{|i-j|}$, where the transmitting SNR is $1/\sigma_n^2$. The proposed RMT estimators have diagonal factors that optimize (57)–(59), and are compared with the following estimators:

(Clairvoyant): It is the upper bound for all types of beamformers, refer to (7).

(Diagonal loading with known $\mathbf{h}$ and $\mathbf{R}_n$): It is the upper bound for all beamformers with a diagonal loading structure, which is given in (23). The true covariance matrix $\mathbf{R}_n$ is known and a perfect channel vector $\mathbf{h}$ is assumed, but the covariance

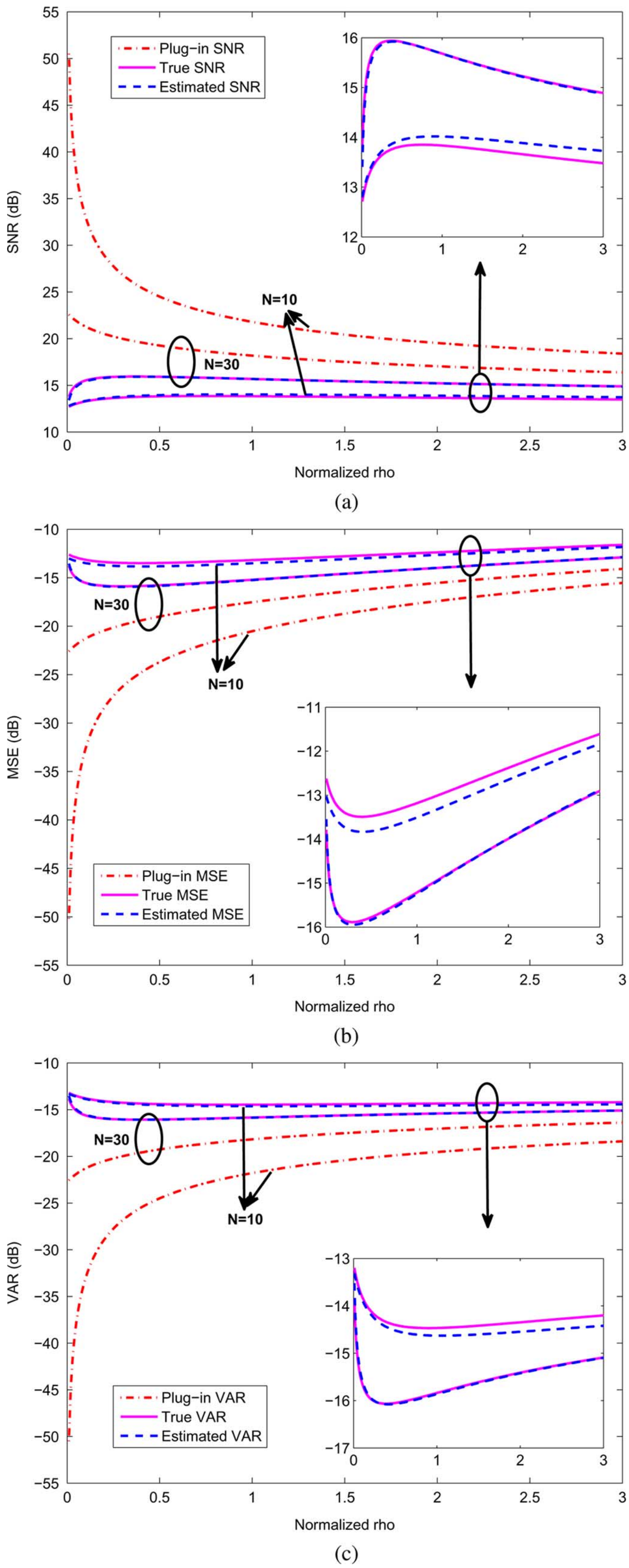

Fig. 3. True SNR, MSE, VAR and their estimators versus normalized ρ, $M = 20$, $N = 10$ and 30.

matrix estimator is restricted to be a linear combination of the sample covariance matrix and the identity matrix.

(LSMI beamformer): Loaded Sample Matrix Inversion beamformer. The loading factor of the noise covariance matrix is chosen to be $10\sigma_n^2$ which has been empirically shown to be a suitable value, and the shrinkage target is $\mathbf{I}$. The channel vector is estimated from the receiving signal and the training signal. LSMI beamformer is also a benchmark in [5]:

$$\mathbf{w}_{\mathsf{LSMI}} = \left(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + 10\sigma_n^2\mathbf{I}\right)^{-1}\hat{\mathbf{h}}.$$

(Traditional SMI): It is the traditional sample covariance matrix inversion method, and the estimated channel vector is used.

In the proposed RMT method, we replace $\mathbf{I}$ with $\frac{1}{M}\mathsf{tr}(\hat{\mathbf{R}}_{n,\mathsf{SCM}})\mathbf{I}$ in (57)–(60) to simplify the search process. Since the corrected noise covariance matrix is a linear combination of $\hat{\mathbf{R}}_{n,\mathsf{SCM}}$ and $\mathbf{I}$ with weights α and ρ, respectively, it is better to make $\hat{\mathbf{R}}_{n,\mathsf{SCM}}$ and $\mathbf{I}$ in similar scale (which could differ substantially due to different transmitting SNR).

We evaluate SNR, MSE, and VAR. Each experiment is conducted 200 times and the average criteria are compared.

We first plot the output SNR, MSE, and VAR versus the number of training samples in Fig. 4(a)–(c), respectively. It can be seen in Fig. 4(a) that the output SNR of the clairvoyant method is nearly constant (and the oscillation comes from different realizations of the channel), since this method does not depend on the samples. The performance of our proposed beamformer is very close to that of diagonal loading with known $\mathbf{h}$ and $\mathbf{R}_n$, even when the sample size is small. Furthermore, the proposed method dominates the LSMI beamformer and the traditional SMI method. When N becomes larger, the performance of the SMI method improves but the performance of LSMI does not change much, since it exploits a fixed and large diagonal loading factor. Similar conclusions can be drawn for Fig. 4(b) and (c), but we emphasize that under MSE criteria, the advantage of the proposed beamformer is even more obvious with respect to the LSMI and the SMI methods when N is smaller than M.

We finally plot the output SNR, MSE, and VAR versus the transmitting SNR in Fig. 5(a)–(c), respectively. It can be seen in Fig. 5(a) that the output SNR of the clairvoyant increases linearly with respect to the transmitting SNR. The performance of our proposed beamformer is very close to that of diagonal loading with known $\mathbf{h}$ and $\mathbf{R}_n$ under all transmitting SNRs. When transmitting SNR is low, the difference between our proposed beamformer and diagonal loading with known $\mathbf{h}$ and $\mathbf{R}_n$ is larger because of the error in estimating the channel vector. Moreover, in this scenario, our proposed method outperforms LSMI by approximately 2 dB and SMI by nearly 3 dB. Similar conclusions can be drawn for Fig. 4(b) and (c).

C. *Simulation in Portfolio Optimization*

In this section, we use real market data to evaluate the proposed portfolio. We consider the stocks conforming the Hang Seng Index. The data we use are Hang Seng Index of 45 stocks of Yahoo Finance daily close prices in the period from Jan. 1, 2008 to July 31, 2011 (i.e., 720 days), and the log returns are calculated. We disregard the first 200 days since the returns are very unstable. We use a rolling window procedure to evaluate the out-of-sample performance of the portfolios as follows: At

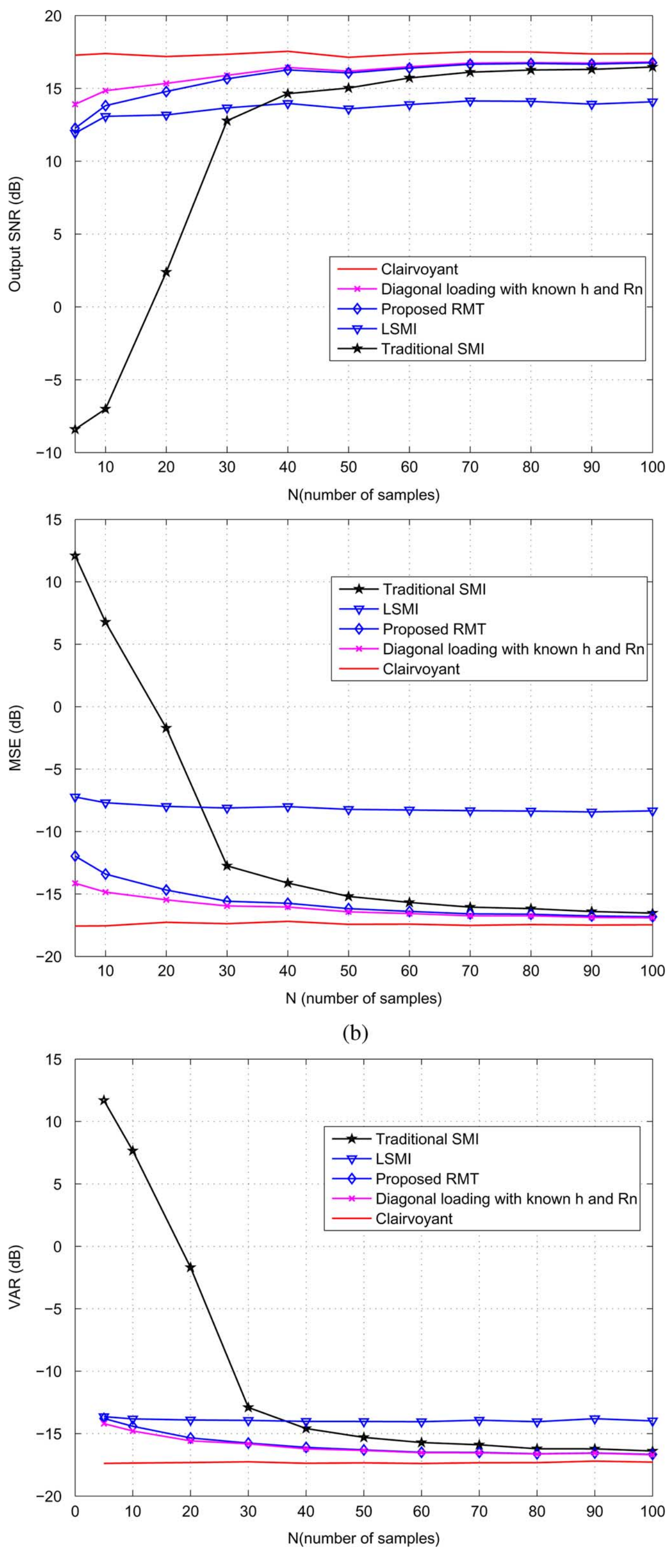

Fig. 4. SNR, MSE and VAR in the beamforming problem of wireless communications: transmitting SNR is 0 dB, $M = 20$, N varies from 5 to 100.

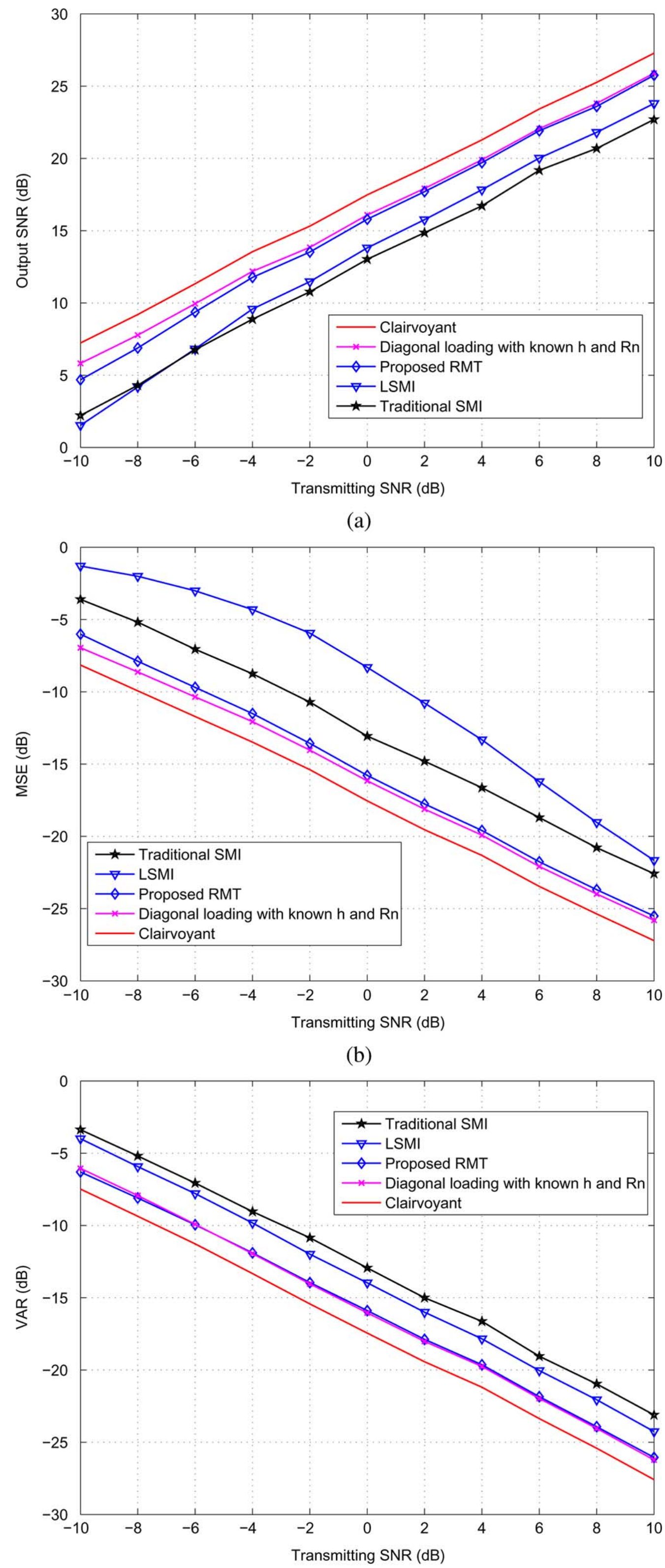

Fig. 5. SNR, MSE and VAR in the beamforming problem of wireless communications: transmitting SNR varies from -10 to 10 dB, $M = 20,\ N = 30$.

a particular day t (t can be considered as a window index), we use the previous N days (i.e., $t-N$ to $t-1$) as the training period to construct the estimated portfolio $\hat{\mathbf{w}}$. In the following 10 days, which is the test period, we use this portfolio and obtain its return. Then this window shifts until the end of the data. With this series of returns, we can compute the mean and the standard deviation, and the ratio is exactly the Sharpe ratio.

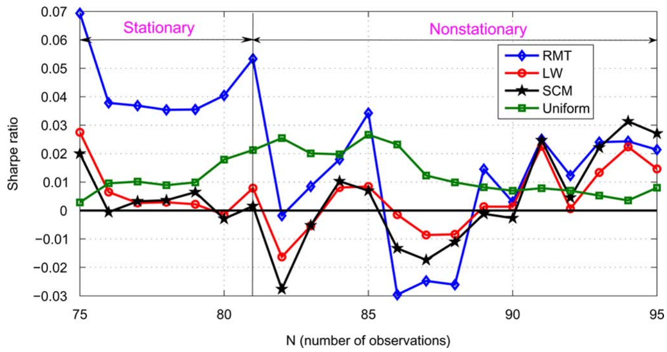

Fig. 6. Sharpe Ratio of RMT, LW, SCM, and a uniform portfolio: $M = 45$, N varies from 75 to 95.

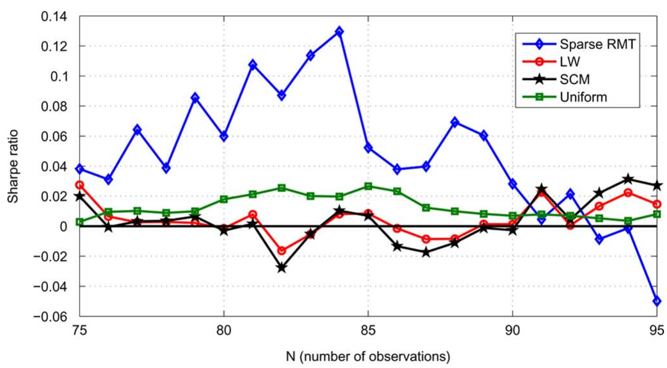

Fig. 7. Sharpe Ratio of sparse RMT, LW, SCM, and a uniform portfolio: $M = 45, N$ varies from 75 to 95.

The proposed RMT estimator has shrinkage factors that optimize (60), and is compared with the following estimators: (LW): The estimator of the sample covariance matrix of asset returns proposed in [15] is used to construct the portfolio.

(SCM): The covariances of the asset returns are estimated with the sample covariance matrix of the data in the training period. Then the portfolio is constructed based on the sample mean and the sample covariance matrix.

(Uniform): The naive portfolio that assigns the same weight to each asset. Note that it is shown that the uniform portfolio has good performance and it is not easy to beat [22].

We plot the Sharpe ratio versus the number of training samples (or the window length) in Figs. 6 and 7.

It can be seen Fig. 6 that the proposed RMT portfolio is sensitive to the choice of the window length: When N varies from 75 to 81, the proposed RMT portfolio outperforms the other methods, but when N is larger than 81, the performance of the RMT portfolio is worse and unstable. This is mainly because when the training window length is too long, our assumptions are violated: the mean return and the covariance cannot be stationary in a long period. Therefore, we can divide data in Fig. 6 into the stationary part and the nonstationary part, according to stationarity of the mean return and the covariance in the training period. From this point of view, the proposed RMT portfolio performs well if the training period is stationary.

Moreover, we can improve the proposed RMT portfolio so that even in a longer training period it still outperforms the other portfolios. It is shown in [23] that sparsity can stabilize the portfolio, since the uncertainty to be estimated is decreased. Based on that, we modify the proposed RMT method so that, after obtaining the optimal $\mathbf{w}$, we then set 0 to those weights whose absolute values are less than 5 percent of the summed absolute values of all the weights. This is a common post-processing step in sparse methods [6]. The improved RMT portfolio is referred to as (Sparse RMT). It can be seen in Fig. 7 that Sparse RMT portfolio outperforms all the other methods a lot when N varies from 75 to 90.

V. Conclusions and Discussion

In this paper, we have studied the problem of linear filter optimization with finite sample size in wireless communications and finance. We have considered the degradation effects caused by the estimation error in the channel vector and the noise covariance matrix (or the mean return and the asset covariance matrix). We have used the diagonal loading method (or the shrinkage estimator) to mitigate the degradation and estimated the optimal loading (shrinkage) factor based on random matrix theory. Under several widely-used criteria in wireless communications and finance, we have obtained linear filters that are as good as the optimal filters with a shrinkage structure and a perfect channel vector (or mean return) under different criteria. We have performed Monte Carlo simulations with both synthetic data and real market data, and have shown the advantage of our linear filters compared with some well-known linear filters in wireless communications and finance.

There are other applications that can be included in the proposed linear filter optimization framework, such as speech recognition and radar signal processing. Speech recognition performance degrades significantly in distant-talking environments [24], [25], and microphone array processing techniques can improve the quality of the output signal and increase the SNR. Moreover, in radar systems, both transmit and receive beamforming are used to improve the system performance [26], [27]. The proposed method on channel estimation and noise covariance matrix estimation can be applied in these applications, which are regarded as our potential directions of future work.

Appendix

We first provide some useful notations and stochastic convergence results, which are essential for the proof of the main theorems. Here are some notations: We apply eigenvalue decomposition to $\mathbf{W}$ so that we obtain $\mathbf{W} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$. Then we absorb $\mathbf{U}$ into $\mathbf{N}$ and have $\tilde{\mathbf{N}} = \mathbf{N}\mathbf{U}$. Due to the rotation invariant property of Gaussian distributed vectors, we also have columns of $\tilde{\mathbf{N}}$ are i.i.d. distributed with covariance matrix $\mathbf{R}_n$. Therefore $\hat{\mathbf{R}}_{n,\mathsf{SCM}}$ can be written in the following form:

$$\hat{\mathbf{R}}_{n,\mathsf{SCM}} = \frac{1}{N}\mathbf{Y}\mathbf{W}\mathbf{Y}^H = \frac{1}{N}\mathbf{N}\mathbf{W}\mathbf{N}^H = \frac{1}{N}\tilde{\mathbf{N}}\mathbf{\Lambda}\tilde{\mathbf{N}}^H. \tag{61}$$

In the sequel, $\mathbf{\Theta}$ will be a nonrandom $M \times M$ matrix whose trace norm is bounded uniformly in M. The following lemma will be instrumental in the proof of our results; see [28], [14] for a proof. Note that in [28] and [14] the weighting matrix $\mathbf{W}$ is

[6]Furthermore, sparsity in beamforming has been considered as a way to reduce the RF circuits.

required to be diagonal. Hence, we can use $\mathbf{\Lambda}$ and $\tilde{\mathbf{N}}$ instead of $\mathbf{W}$ and $\mathbf{N}$, and the same results hold true.

Lemma 3: Assume that Assumptions (A1)–(A2) hold. Then, for each $\rho > 0$ the following results hold true almost surely:

$$\begin{aligned}&\mathsf{tr}[\mathbf{\Theta}(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}]\\ &\asymp \mathsf{tr}[\mathbf{\Theta}(\tilde{\delta}\mathbf{R}_n + \rho\mathbf{I})^{-1}] \end{aligned}\tag{62}$$

$$\begin{aligned}&\mathsf{tr}[\mathbf{\Theta}(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\mathbf{R}_n(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}]\\ &\asymp \frac{1}{1-\gamma\tilde{\gamma}}\mathsf{tr}[\mathbf{\Theta}\mathbf{R}_n(\tilde{\delta}\mathbf{R}_n + \rho\mathbf{I})^{-2}] \end{aligned}\tag{63}$$

$$\begin{aligned}&\mathsf{tr}[\mathbf{\Theta}\hat{\mathbf{R}}_{n,\mathsf{SCM}}(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-2}]\\ &\asymp \frac{\mathsf{tr}[\mathbf{\Theta}\mathbf{R}_n(\tilde{\delta}\mathbf{R}_n + \rho\mathbf{I})^{-2}]}{(1-\gamma\tilde{\gamma})\frac{1}{N}\mathsf{tr}[\alpha\mathbf{W}(\mathbf{I}+\hat{\delta}\alpha\mathbf{W})^{-2}]} \end{aligned}\tag{64}$$

where $\hat{\mathbf{R}}_{n,\mathsf{SCM}}$ is expressed as in (24), $\{\delta_M, \tilde{\delta}_M\}$ and $\{\gamma_M, \tilde{\gamma}_M\}$ are defined as in (28) and (29).

Moreover, we give the simplified expressions of $\hat{\mathbf{R}}_{n,\mathsf{SCM}}$ and $\mathbf{v}$. Recall that $\mathbf{W} = \mathbf{I} - \frac{1}{N}\mathbf{s}^*\mathbf{s}^T$, we conclude that $\mathbf{\Lambda} = \begin{bmatrix}\mathbf{I}_{N-1} & \mathbf{0}_{N-1}\\ \mathbf{0}_{N-1}^T & 0\end{bmatrix}$. Letting $\tilde{\mathbf{v}} = \frac{1}{\sqrt{N}}\mathbf{U}^H\mathbf{s}^*$, and noting that $\mathbf{W}\mathbf{s}^* = \mathbf{0}$ implies $\mathbf{\Lambda}\tilde{\mathbf{v}} = \mathbf{0}$, due to the particular form of $\mathbf{\Lambda}$, we have $\tilde{\mathbf{v}}$ in the form of $\tilde{\mathbf{v}} = [\mathbf{0}_{N-1},\ v_2]^T$. Additionally, $|v_2|^2 = \|\tilde{\mathbf{v}}\|_2^2 = 1$.

Now, we decompose $\tilde{\mathbf{N}}$ into two independent parts, i.e., $\tilde{\mathbf{N}} = [\tilde{\mathbf{N}}_1,\ \tilde{\mathbf{n}}_2]$, then $\hat{\mathbf{R}}_{n,\mathsf{SCM}}$ and $\mathbf{v}$ can be written as:

$$\hat{\mathbf{R}}_{n,\mathsf{SCM}} = \frac{1}{N}\tilde{\mathbf{N}}_1\tilde{\mathbf{N}}_1^H, \quad \mathbf{v} = \frac{v_2}{\sqrt{N}}\tilde{\mathbf{n}}_2. \tag{65}$$

It can be seen in (65) that $\hat{\mathbf{R}}_{n,\mathsf{SCM}}$ and $\mathbf{v}$ are independent. This conclusion would be used in the proofs in the following.

A. Proof of Theorem 1

First of all, we provide the derivations of $\bar{\xi}_1$ and $\bar{\xi}_4$. Noting that $\|\mathbf{h}\|_2$ is bounded in $M, \bar{\xi}_1$ can be easily obtained from (62) in Lemma 3 by letting $\mathbf{\Theta} = \mathbf{h}\mathbf{h}^H$. Moreover,

$$\mathbf{R}_n(\tilde{\delta}\mathbf{R}_n + \rho\mathbf{I})^{-2} = \mathbf{R}_n^{1/2}(\tilde{\delta}\mathbf{R}_n + \rho\mathbf{I})^{-2}\mathbf{R}_n^{1/2}, \tag{66}$$

along with (63) in Lemma 3 where $\mathbf{\Theta} = \mathbf{h}\mathbf{h}^H$, we obtain $\bar{\xi}_4$.

Then we provide the derivation of $\bar{\xi}_3$ and the derivation of $\bar{\xi}_6$ follows the same way. Note that

$$\xi_3 = \frac{1}{N}\tilde{\mathbf{n}}_2^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\tilde{\mathbf{n}}_2. \tag{67}$$

Moreover, recall that $\tilde{\mathbf{n}}_2$ is a random Gaussian vector with zero mean and Covariance matrix $\mathbf{R}_n$, then we can conclude that for any deterministic matrix $\mathbf{\Theta}$ whose trace norm is bounded uniformly in $M, \mathsf{tr}[\mathbf{\Theta}(\tilde{\mathbf{n}}_2\tilde{\mathbf{n}}_2^H)] \asymp \mathsf{tr}[\mathbf{\Theta}\mathbf{R}_n]$. Additionally, $\hat{\mathbf{R}}_{n2,\mathsf{SCM}}$ and $\mathbf{v}$ are independent. From (62) in Lemma 3, we can conclude that

$$\xi_3 \asymp \frac{1}{N}\mathsf{tr}[\mathbf{R}_n(\tilde{\delta}\mathbf{R}_n + \rho\mathbf{I})^{-1}] \tag{68}$$

and the right hand side of (68) is exactly δ.

Moreover, the derivations of ξ_2 and ξ_5 are similar and we use ξ_2 to illustrate. Noting that

$$\xi_2 = \mathbf{h}^H(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\frac{v_2}{\sqrt{N}}\tilde{\mathbf{n}}_2, \tag{69}$$

along with the zero mean of $\tilde{\mathbf{n}}_2\mathbf{h}^H$, we have ξ_2 converges to 0 since $\mathsf{tr}[\mathbf{\Theta}(\tilde{\mathbf{n}}_2\mathbf{h}^H)] \asymp 0$.

B. Proof of Theorem 2

We first provide the convergence results of $\hat{\xi}_{4,\mathsf{cnv}}$ in Theorem 2. Noting that

$$\begin{aligned}&(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\hat{\mathbf{R}}_{n,\mathsf{SCM}}(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-1}\\ &\qquad = \hat{\mathbf{R}}_{n,\mathsf{SCM}}(\hat{\mathbf{R}}_{n,\mathsf{SCM}} + \rho\mathbf{I})^{-2}, \end{aligned}\tag{70}$$

we can easily obtain $b\hat{\xi}_{4,\mathsf{cnv}} \asymp \bar{\xi}_4$ with (64) in Lemma 3 where $\mathbf{\Theta} = \mathbf{h}\mathbf{h}^H$.

Regarding the convergence results of $\hat{\xi}_{5,\mathsf{cnv}}$ and $\hat{\xi}_{6,\mathsf{cnv}}$, recall that $\mathsf{tr}[\mathbf{\Theta}(\tilde{\mathbf{n}}_2\mathbf{h}^H)] \asymp 0$ and $\mathsf{tr}[\mathbf{\Theta}(\tilde{\mathbf{n}}_2\tilde{\mathbf{n}}_2^H)] \asymp \mathsf{tr}[\mathbf{\Theta}\mathbf{R}_n]$, along with (64) in Lemma 3, we finally get $\xi_{5,\mathsf{cnv}} \asymp 0$ and $b\xi_{6,\mathsf{cnv}} \asymp \bar{\xi}_6$.

References

[1] H. Van Trees and J. Wiley, *Optimum Array Processing*. New York, NY, USA: Wiley, 2002.
[2] A. Gershman, N. Sidiropoulos, S. Shahbazpanahi, M. Bengtsson, and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Process. Mag.* , vol. 27, no. 5, pp. 62–75, May 2010.
[3] G. Cornuejols and R. Tutuncu, *Optimization Methods in Finance*. Cambridge, U.K.: Cambridge Univ. Press, 2006.
[4] H. Markowitz, "Portfolio selection," *J. Finance*, vol. 7, no. 1, pp. 77–91, 1952.
[5] S. Vorobyov, A. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 313–324, Feb. 2003.
[6] Y. Abramovich, "Controlled method for adaptive optimization of filters using the criterion of maximum snr," *Radio Eng. Electron. Phys.*, vol. 26, pp. 87–95, 1981.
[7] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 35, pp. 1365–1376, 1987.
[8] J. Li, P. Stoica, and Z. Wang, "On robust capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, Jul. 2003.
[9] X. Mestre and M. Lagunas, "Finite sample size effect on minimum variance beamformers: Optimum diagonal loading factor for large arrays," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 69–82, Jan. 2006.
[10] O. Ledoit and M. Wolf, "Improved estimation of the covariance matrix of stock returns with an application to portfolio selection," *J. Empir. Finance*, vol. 10, no. 5, pp. 603–621, 2003.
[11] L. Lei, J. Lie, A. Gershman, and C. See, "Robust adaptive beamforming in partly calibrated sparse sensor arrays," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1661–1667, Mar. 2010.
[12] O. Besson and F. Vincent, "Performance analysis of beamformers using generalized loading of the covariance matrix in the presence of random steering vector errors," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 452–459, Feb. 2005.
[13] S. Shahbazpanahi, A. Gershman, Z. Luo, and K. Wong, "Robust adaptive beamforming for general-rank signal models," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2257–2269, Sep. 2003.
[14] F. Rubio, X. Mestre, and D. P. Palomar, "Performance analysis and optimal selection of large minimum variance portfolios under estimation risk," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 8, pp. 337–350, Aug. 2012.
[15] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivar. Anal.*, vol. 88, pp. 365–411, Feb. 2004.
[16] M. Zhang, F. Rubio, and D. P. Palomar, "Improved calibration of high-dimensional precision matrices," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1509–1519, Jun. 2013.
[17] V. DeMiguel, L. Garlappi, F. Nogales, and R. Uppal, "A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms," *Manage. Sci.*, vol. 55, no. 5, pp. 798–812, 2009.
[18] D. Johnson and D. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Englewood Cliffs, NJ, USA: PTR/Prentice-Hall, 1993.
[19] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[20] F. Fabozzi, P. Kolm, and D. Pachamanova, *Robust Portfolio Optimization and Management*. New York, NY, USA: Wiley, 2007.
[21] D. Ruppert, Statistics and Data Analysis for Financial Engineering 2010, Springer Texts in Statistics.

[22] V. DeMiguel, L. Garlappi, and R. Uppal, “Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?,” *Rev. Financial Studies*, vol. 22, no. 5, pp. 1915–1953, 2009.
[23] J. Brodie, I. Daubechies, C. De Mol, D. Giannone, and I. Loris, “Sparse and stable Markowitz portfolios,” *Proc. Nat. Acad. Sci.*, vol. 106, no. 30, pp. 12267–12272, 2009.
[24] M. Wölfel and J. McDonough*, Distant Speech Recognition*. New York: Wiley, 2009.
[25] X. A. Miro, S. Bozonnet, N. Evans, C. Fredouille, G. Friedland, and O. Vinyals, “Speaker diarization: A review of recent research,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 2, pp. 356–370, Feb. 2012.
[26] I. Bekkerman and J. Tabrikian, “Target detection and localization using MIMO radars and sonars,” *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 3873–3883, Oct. 2006.
[27] H. Godrich, A. M. Haimovich, and R. S. Blum, “Target localization accuracy gain in MIMO radar-based systems,” *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2783–2803, Jun. 2010.
[28] F. Rubio and X. Mestre, “Spectral convergence for a general class of random matrices,” *Statist. Probabil. Lett.*, vol. 81, pp. 592–602, May 2011.

Mengyi Zhang (S’09) received the B.Sc. degree in electronic information from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2009.

Currently, she has finished her Ph.D. defense and will receive the Ph.D. degree soon in 2013 with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST). Her research interests are mainly on statistical signal processing and optimization in wireless communications and financial systems.

Francisco Rubio (M’07) received the B.S. degree in electrical engineering from the Technical University of Madrid, the M.S. degree in information science from the Mannheim University of Applied Sciences, and the Ph.D. degree from the Technical University of Catalonia, in 2001, 2002, and 2008, respectively.

Since March 2012, he is a Quantitative Analyst with Genetic Finance Ltd., Hong Kong. From January 2011 to March 2012, he was a Research Associate with the Hong Kong University of Science and Technology, and from June 2008 to January 2011, he was a Research Associate and the head of the Engineering Department of the Centre Tecnològic de Telecomunicacions de Catalunya, Barcelona. From July to December 2007, he held a visiting appointment as a Predoctoral Research Fellow with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL. Prior to his Ph.D., he held research and professional appointments at Retevisión Móvil (now France Telecom’s Orange) and Siemens in Madrid, and T-Nova Deutsche Telekom (Darmstadt), MABEL Biomedical Engineering Labs (Mannheim), and BMW AG (Munich) in Germany.

Dr. Rubio has actively participated in several European research projects and industrial contracts, and is a corecipient of an academic performance award from T-Systems.

Daniel P. Palomar (S’99-M’03-SM’08-F’12) received the electrical engineering and Ph.D. degrees (both with honors) from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1998 and 2003, respectively.

He joined the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology (HKUST), in 2006, where he is an Associate Professor. Since 2013, he has been a Fellow of the Institute for Advanced Study (IAS) at HKUST. He had previously held several research appointments, namely, at King’s College London (KCL), London, U.K.; Technical University of Catalonia (UPC), Barcelona; Stanford University, Stanford, CA; Telecommunications Technological Center of Catalonia (CTTC), Barcelona; Royal Institute of Technology (KTH), Stockholm, Sweden; University of Rome “La Sapienza,” Rome, Italy; and Princeton University, Princeton, NJ. His current research interests include applications of convex optimization theory, game theory, and variational inequality theory to financial systems and communication systems.

Dr. Palomar is an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION THEORY, and has been an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, a Guest Editor of the IEEE SIGNAL PROCESSING MAGAZINE 2010 Special Issue on “Convex Optimization for Signal Processing,” the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2008 Special Issue on “Game Theory in Communication Systems,” and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS 2007 Special Issue on “Optimization of MIMO Transceivers for Realistic Communication Networks.” He serves on the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications (SPCOM). He is a recipient of a 2004/2006 Fulbright Research Fellowship; the 2004 Young Author Best Paper Award by the IEEE Signal Processing Society; the 2002/2003 best Ph.D. prize in Information Technologies and Communications by the Technical University of Catalonia (UPC); the 2002/2003 Rosina Ribalta first prize for the Best Doctoral Thesis in Information Technologies and Communications by the Epson Foundation; and the 2004 prize for the best Doctoral Thesis in Advanced Mobile Communications by the Vodafone Foundation and COIT.

Xavier Mestre (S’96–M’03–SM’09) received the M.S. and Ph.D. degrees in electrical engineering from the Technical University of Catalonia (UPC) in 1997 and 2002, respectively, and the Licenciate Degree in mathematics in 2011. During the pursuit of his Ph.D., he was recipient of a 1998–2001 Ph.D. scholarship (granted by the Catalan Government) and was awarded the 2002 Rosina Ribalta second prize for the Best Doctoral Thesis Project within areas of Information Technologies and Communications by the Epson Iberica foundation.

From January 1998 to December 2002, he was with UPC’s Communications Signal Processing Group, where he has worked as a Research Assistant and has participated actively in the European-funded projects. In January 2003, he joined the Telecommunications Technological Center of Catalonia (CTTC), where he currently holds a position as a Senior Research Associate in the area of Radio Communications. During this time, he has actively participated in multiple European projects and several contracts with the local industry. Currently, he is head of the Advanced Signal and Information Processing Department.

Dr. Mestre has been an Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING, 2007–2011 and since 2012, has been a member of the IEEE Sensor Array and Multichannel Technical Committee.