

Random Vandermonde Matrices-Part II: Applications

Øyvind Ryan, *Member, IEEE*

Centre of Mathematics for Applications,
University of Oslo, P.O. Box 1053 Blindern, 0316 Oslo, Norway

Phone: +47 93 24 83 21

Fax: +47 22 85 43 49

Email: oyvindry@ifi.uio.no

Mérouane Debbah, *Member, IEEE*

SUPELEC,

Alcatel-Lucent Chair on Flexible Radio, Plateau de Moulon,

3 rue Joliot-Curie,

91192 GIF SUR YVETTE CEDEX, France

Phone: +33 1 69 85 20 07

Fax: +33 1 69 85 12 59

Email: merouane.debbah@supelec.fr

Abstract—In this paper, we review some potential applications of random Vandermonde matrices in the field of signal processing and wireless communications. Using asymptotic results based on the theory of random Vandermonde matrices, we show through several application examples, namely deconvolution, wireless capacity analysis and sampling theory, the research potential of this theory. Quite surprisingly, in nearly all the cases, the asymptotic results turn out to be valid for dimensions which are of interest for the community. The simulations confirm that random matrix theory/free probability theory are once more a unique tool to better understand the behavior of the eigenvalues of matrices.

Index Terms—Vandermonde matrices, Random Matrices, deconvolution, limiting eigenvalue distribution, MIMO.

I. INTRODUCTION

Vandermonde matrices have had for a long time a central position in signal processing due to their connections with other important matrices in the field such as the FFT [1] or Hadamard [2] transforms to name a few. The matrices have various applications in different fields [3], [4], [5], [6]. The applied research has been somewhat tempered by the fact that very few theoretical results were available. For example, until the recent results in [7], only results on the determinants and the moments of the determinant of Vandermonde matrices were known [8]. For a given deterministic Vandermonde matrix \mathbf{V} of dimension $N \times L$ defined by:

$$\mathbf{V} = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & \cdots & 1 \\ e^{-j\omega_1} & \cdots & e^{-j\omega_L} \\ \vdots & \ddots & \vdots \\ e^{-j(N-1)\omega_1} & \cdots & e^{-j(N-1)\omega_L} \end{pmatrix} \quad (1)$$

This project is partially sponsored by the project BIONET (INRIA).

This work was supported by Alcatel-Lucent within the Alcatel-Lucent Chair on flexible radio at SUPELEC

only the determinant of (1) when $N = L$ has a nice expression

$$\frac{1}{N^{N/2}} \prod_{1 \leq k < l \leq N} (e^{-j\omega_l} - e^{-j\omega_k}).$$

The result however is of little use in signal processing and wireless communications for example. One way of circumventing this problem was proposed by Ryan et al. [7] where $\omega_1, \dots, \omega_L$ are modeled as independent and identically distributed (phases) taking values on $[0, 2\pi)$. In this case, the random phases enable to predict neat expressions of the asymptotic (in the sense where $N \rightarrow \infty$, $L \rightarrow \infty$ and $\frac{L}{N} \rightarrow c$) moments of the Gram matrix associated to the Vandermonde matrix as well as more advanced models where products of the Vandermonde matrix with deterministic matrices are concerned. Remarkably, the results show that the moments depend only on the ratio c and the distribution of the entries of the phases and have explicit expressions. The self-averaging properties of these matrices provide therefore a neat tool to determine the parameters of interest in a problem where Vandermonde matrices are put forward. These results are reminiscent of similar results concerning i.i.d random matrices [9] which have shed light in the design of many important wireless communication problems such as CDMA [10], MIMO [11] or OFDM [12]. Building on the results of [7], this paper provides some useful applications showing the implications of these results in various applied fields. In section II, we show how Vandermonde matrices can be used to perform deconvolution and extend therefore the results of [13] restricted to the Gaussian case. The main examples are geared towards wireless systems, and include estimation of the number of paths, detection of the transmissions powers of the users, detection of the number of sources, and wavelength estimation. In section III, the asymptotic results are used to perform wireless capacity analysis. Capacity of line of sight

of multi-user MIMO systems and multifold scattering are the main examples which have not been dealt with in the literature before or up to some very coarse approximations [14]. Finally, section IV provides an important feature of the application of Vandermonde matrices to the very active field of sparse signal reconstruction. Interestingly, one can provide a general framework where only the sampling distribution matters asymptotically. The first example concerns the computation of the MMSE (Minimum Mean Square Error) whereas the second example focuses on the estimation of the sampling distribution. All the sections are illustrated by extensive simulations which discuss the validity of the asymptotic claims in the finite regime.

In the following, upper (lower boldface) symbols will be used for matrices (column vectors) whereas lower symbols will represent scalar values, $(\cdot)^T$ will denote transpose operator, $(\cdot)^*$ conjugation and $(\cdot)^H = ((\cdot)^T)^*$ hermitian transpose. \mathbf{I}_n will represent the $n \times n$ identity matrix. We let tr_n be the normalized trace for matrices of order $n \times n$, and Tr the non-normalized trace. \mathbf{V} will be used only to denote Vandermonde matrices with a given phase distribution. The dimensions of the Vandermonde matrices will always be $N \times L$ unless otherwise stated, and the phase distribution of the Vandermonde matrices will always be denoted by ω .

II. DECONVOLUTION

A. Detection of the number of sources

Let us consider a receiver with N antennas and L mobiles (each with a single antenna) in the cell. The received signal at the base station is given by

$$\mathbf{r}_i = \mathbf{V}\mathbf{P}^{\frac{1}{2}}\mathbf{s}_i + \mathbf{n}_i. \quad (2)$$

Here, \mathbf{r}_i , \mathbf{s}_i , \mathbf{n}_i are respectively the $N \times 1$ received vector, the $L \times 1$ transmit vector by the L users and the $N \times 1$ additive noise. In the case of a line of sight between the users and the base station (and considering a Uniform Linear Array), matrix \mathbf{V} has the following form:

$$\mathbf{V} = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & \dots & 1 \\ e^{-j2\pi\frac{d}{\lambda}\sin(\theta_1)} & \dots & e^{-j2\pi\frac{d}{\lambda}\sin(\theta_L)} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi(N-1)\frac{d}{\lambda}\sin(\theta_1)} & \dots & e^{-j2\pi(N-1)\frac{d}{\lambda}\sin(\theta_L)} \end{pmatrix} \quad (3)$$

Here, θ_i is the angle of the user in the cell and is supposed to be uniformly distributed over $[-\alpha, \alpha]$. $\mathbf{P}^{\frac{1}{2}}$ is an $L \times 1$ power matrix due to the different distances from which the users emit. In other words, we assume that the phase distribution has the form $2\pi\frac{d}{\lambda}\sin(\theta)$ with θ uniformly distributed on $[-\alpha, \alpha]$. It is easily seen, by taking inverse functions, that the density is, when $\frac{2d\sin\alpha}{\lambda} < 1$,

$$p_\omega(x) = \frac{1}{2\alpha\sqrt{\frac{4\pi^2d^2}{\lambda^2} - x^2}}$$

on $[-\frac{2\pi d\sin\alpha}{\lambda}, \frac{2\pi d\sin\alpha}{\lambda}]$, and 0 elsewhere. This density is shown in figure 1. The effect of a high concentration for this density near the origin is that the Vandermonde matrix has a

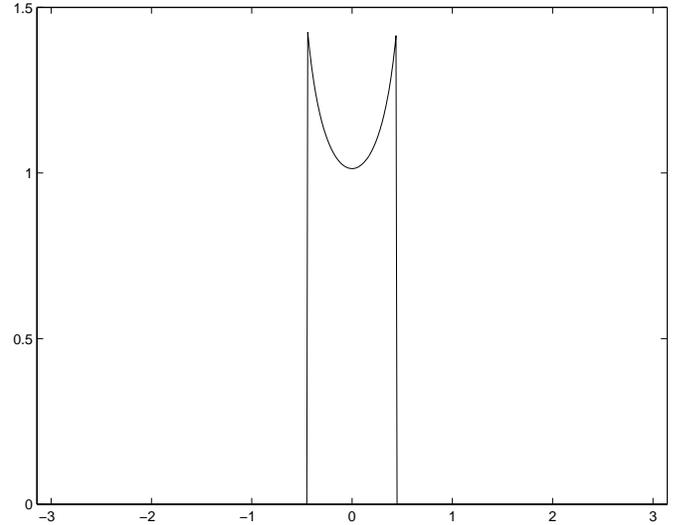


Fig. 1. The density $p_\omega(x)$ used in this paper. $\alpha = \frac{\pi}{4}$ and $\lambda = 10d$.

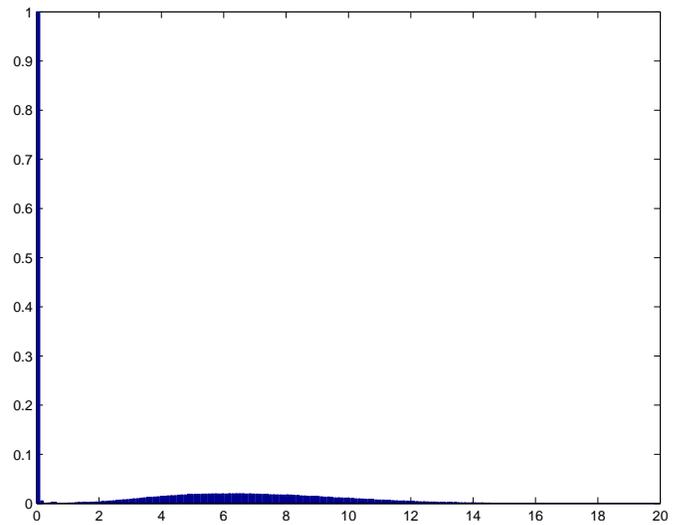


Fig. 2. Histogram of the mean eigenvalue distribution of 640 samples of $\mathbf{V}^H\mathbf{V}$, with \mathbf{V} a 1600×1600 Vandermonde matrix with phase distribution p_ω .

high concentration of the eigenvalues near the origin, and also a higher proportion of larger eigenvalues, when compared to the uniform phase distribution. This can be seen from figure 2, where the mean eigenvalue distribution of 640 samples of a 1600×1600 Vandermonde matrix with phase distribution p_ω with $\alpha = \frac{\pi}{4}$, $d = 1$, and $\lambda = 10d$ is shown. A corresponding eigenvalue histogram for uniform phase distribution can be found in [7]. Throughout the paper we will assume, as in figure 2, that $\alpha = \frac{\pi}{4}$, $d = 1$, and $\lambda = 10d$ when model (3) is used. With this assumption, $\frac{2d\sin\alpha}{\lambda} < 1$ is always fulfilled.

The goal is to detect the number of sources L and their respective power based on the sample covariance matrix supposing that we have K observations, of the same order as N . When the number of observation is quite higher than N (and the noise variance is known), classical subspace methods [15] provide tools to detect the number of sources. Indeed, let

\mathbf{R} be the true covariance matrix i.e

$$\mathbf{R} = \mathbf{V}\mathbf{P}\mathbf{V}^H + \sigma^2\mathbf{I}$$

The matrix \mathbf{R} has $N - L$ eigenvalues equal to σ^2 and L eigenvalues strictly superior to σ^2 . One can therefore determine the number of source by counting the number of eigenvalues different from σ^2 . However, in practice, one has only access to the sample covariance matrix given by

$$\hat{\mathbf{R}} = \frac{1}{K}\mathbf{Y}\mathbf{Y}^H,$$

with

$$\mathbf{Y} = [\mathbf{r}_1, \dots, \mathbf{r}_K] = \mathbf{V}\mathbf{P}^{\frac{1}{2}}[\mathbf{s}_1, \dots, \mathbf{s}_K] + [\mathbf{n}_1, \dots, \mathbf{n}_K]$$

If one simply has the sample covariance matrix $\hat{\mathbf{R}}$, (2) has three independent parts which must be dealt with in order to get an estimate of \mathbf{P} : the Gaussian matrices $\mathbf{S} = [\mathbf{s}_1, \dots, \mathbf{s}_K]$ and $\mathbf{N} = [\mathbf{n}_1, \dots, \mathbf{n}_K]$, and the Vandermonde part \mathbf{V} . It should thus be possible to combine Gaussian deconvolution [16] and Vandermonde deconvolution [7] by performing the following steps:

- 1) Estimate the moments of $\frac{1}{K}\mathbf{V}\mathbf{P}^{\frac{1}{2}}\mathbf{S}\mathbf{S}^H\mathbf{P}^{\frac{1}{2}}\mathbf{V}^H$ using multiplicative free convolution as described in [13]. This is the denoising part.
- 2) Estimate the moments of $\mathbf{P}\mathbf{V}^H\mathbf{V}$, again using multiplicative free deconvolution.
- 3) Estimate the moments of \mathbf{P} using Vandermonde deconvolution as described in [7].

Putting these steps together, we will prove the following:

Proposition 1: Define

$$I_n = (2\pi)^{n-1} \int_0^{2\pi} p_\omega(x)^n dx, \quad (4)$$

and let $m_P^i = tr_L(\mathbf{P}^i)$ be the moments of \mathbf{P} , and $m_{\hat{\mathbf{R}}}^i = tr_N(\hat{\mathbf{R}}^i)$ the moments of the sample covariance matrix. Then the equations

$$\begin{aligned} m_{\hat{\mathbf{R}}}^1 &= c_2 m_P^1 + \sigma^2 \\ m_{\hat{\mathbf{R}}}^2 &= c_2 m_P^2 + (c_2^2 I_2 + c_2 c_3)(m_P^1)^2 \\ &\quad + 2\sigma^2(c_2 + c_3)m_P^1 + \sigma^4(1 + c_1) \\ m_{\hat{\mathbf{R}}}^3 &= c_2 m_P^3 + (3c_2^2 I_2 + 3c_2 c_3)m_P^1 m_P^2 \\ &\quad + (c_2^3 I_3 + 3c_2^2 c_3 I_2 + c_2 c_3^2)(m_P^1)^3 \\ &\quad + 3\sigma^2(1 + c_1)c_2 m_P^2 \\ &\quad + 3\sigma^2((1 + c_1)c_2^2 I_2 + c_3(c_3 + 2c_2))(m_P^1)^2 \\ &\quad + 3\sigma^4(c_1^2 + 3c_1 + 1)c_2 m_P^1 \\ &\quad + \sigma^6(c_1^2 + 3c_1 + 1) \end{aligned}$$

provide an asymptotically unbiased estimator for the moments m_P^i from the moments of $m_{\hat{\mathbf{R}}}^i$ (or vice versa) when $\lim_{N \rightarrow \infty} \frac{N}{K} = c_1$, $\lim_{N \rightarrow \infty} \frac{L}{N} = c_2$, and where $\lim_{N \rightarrow \infty} \frac{L}{K} = c_3$.

The proof of this can be found in appendix A. Note that the statement applies to any ω with continuous density [7], not only the densities we restrict to here. In the simulations, proposition 1 is put to the test when \mathbf{P} has three sets of powers, 0.5, 1, and 1.5 (with equal probability), with phase distribution

given by (3). Both the number of sources and the powers are estimated. For the phase distribution (3), the integrals I_2 and I_3 can be computed exactly (for general phase distributions they are computed numerically), and are [17]

$$\begin{aligned} I_2 &= \frac{\lambda}{4d\alpha^2} \ln \left(\frac{1 + \sin \alpha}{1 - \sin \alpha} \right) \\ I_3 &= \frac{\lambda^2 \tan \alpha}{4d^2 \alpha^3}. \end{aligned}$$

Under the assumptions $\alpha = \frac{\pi}{4}$ and $\lambda = 10d$ used throughout this paper, the integrals above take the values

$$\begin{aligned} I_2 &= \frac{40}{\pi^2} \ln \left(\frac{2 + \sqrt{2}}{2 - \sqrt{2}} \right) \\ I_3 &= \frac{1600}{\pi^3}. \end{aligned}$$

For estimation of the powers, knowing that we have only three sets of powers with equal probability, it suffices to estimate the three lowest moments in order to get an estimate of the powers (which are the three distinct eigenvalues of \mathbf{P}). Therefore, in the following simulations, proposition 1 is first used to get an estimate of the moments of \mathbf{P} . Then these are used to obtain an estimate of the three distinct eigenvalues of \mathbf{P} using the Newton-Girard formulas [18]. These should then lie close to the three powers of \mathbf{P} .

For the model (3), it turns out that power estimation does not work particularly well. The result is shown in the first plot of figure 5. In the plot, $K = L = N = 576$, and $\sigma = \sqrt{0.1}$. Even though the matrices are quite large, the estimated powers are quite far from the actual powers. Actually, the estimation process is so far off that it computes eigenvalues which are complex conjugate pairs instead of the true, real ones (0.5, 1, 1.5) (this is an explanation for that the two lowest eigenvalues in the plot seem to coincide, since it is only the absolute values of the eigenvalues which are plotted). Increasing the matrix sizes further results in estimates which are closer to the true powers, but one would need matrices of size larger than 2000×2000 to get much closer to the true powers. As will be seen, power estimation works much better for the phase distribution model in the next section. A tentative explanation for this is the difference between the corresponding eigenvalue histograms of those two Vandermonde matrices, which are shown in figure 2 for model (3), and in figure 4 of [7] for the model of the next section.

For estimation of the number of users L , we assume that the power distribution of \mathbf{P} is known, but not L itself. Since L is unknown, in the simulations we enter different candidate values of it into the following procedure:

- 1) Computing the moments $m_P^i = tr_L(\mathbf{P}^i)$ of \mathbf{P} .
- 2) The moments $tr_L(\mathbf{P}^i)$ are fed into the formulas of proposition 1, and we thus obtain candidate moments m_i of the sample covariance matrix $\hat{\mathbf{R}}$.
- 3) Compute the sum of the square errors between the candidate moments of the sample covariance matrix, and the moments of the observed sample covariance

matrices, i.e. compute $\sum_{i=1}^3 |m_i - \hat{m}_i|^2$, where

$$\hat{m}_i = \frac{1}{n} \sum_{j=1}^n \text{tr}_N \left(\hat{\mathbf{R}}_j^i \right),$$

where $\hat{\mathbf{R}}_j$ are the observed sample covariance matrices. A more natural thing to do would perhaps be to compute a weighted sum of the square errors instead, i.e. compute $\sum_{i=1}^3 w_i |m_i - \hat{m}_i|^2$ for some choice of weights w_i . This strategy was used in [13] for Gaussian deconvolution, where it was argued that the *Catalan numbers* [19] are a good choice of weights. The argument was based on the fact that the limit of $\text{tr}_N \left(\left(\frac{1}{N} \mathbf{X} \mathbf{X}^H \right)^k \right)$ as $N \rightarrow \infty$ is the k 'th Catalan number [20], where \mathbf{X} is an $N \times N$ standard Gaussian matrix. We will not perform weighting of the sum of square errors in this paper, since the sum of square errors is computed up to three moments only. For higher moments, the weighting would be much more crucial (the Catalan numbers grow very fast in size).

The estimate L for the number of users is chosen as the one which gives the minimum value for the sum of square errors after these steps.

In figure 3, we have set $\sigma = \sqrt{0.1}$, $N = 100$, and $L = 36$. \mathbf{P} has three sets of powers, 0.5, 1, and 1.5 (with equal probability). We tried the procedure described above for 1 all the way up to 100 observations. It is seen that only a small number of observations are needed in order to get an accurate estimate of L . When $K = 1$, it is seen that more observations are needed to get an accurate estimate of L , when compared to $K = 10$.

B. Estimation of the number of paths

In many channel modeling applications, one needs to determine the number of paths of the channel [21]. For this purpose, consider a multi-path channel of the form:

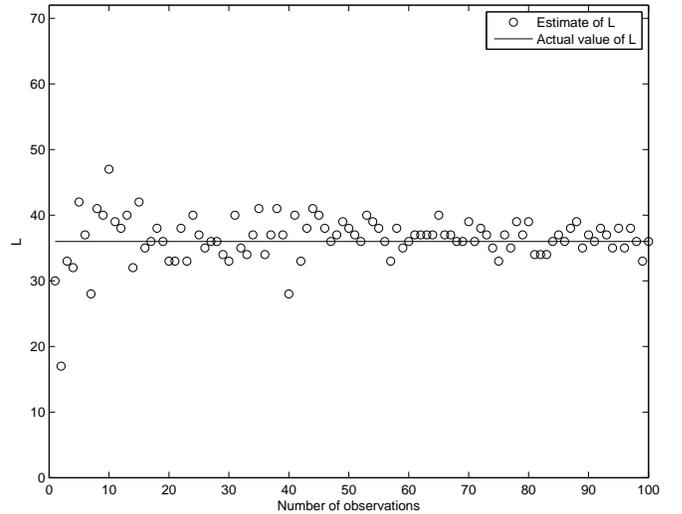
$$h(\tau) = \sum_{i=1}^L \alpha_i g(\tau - \tau_i)$$

Here, α_i are i.d Gaussian random variables with power P_i and τ_i are uniformly distributed delays over $[0, T]$. g is the low pass transmit filter. In the frequency domain, the channel is given by:

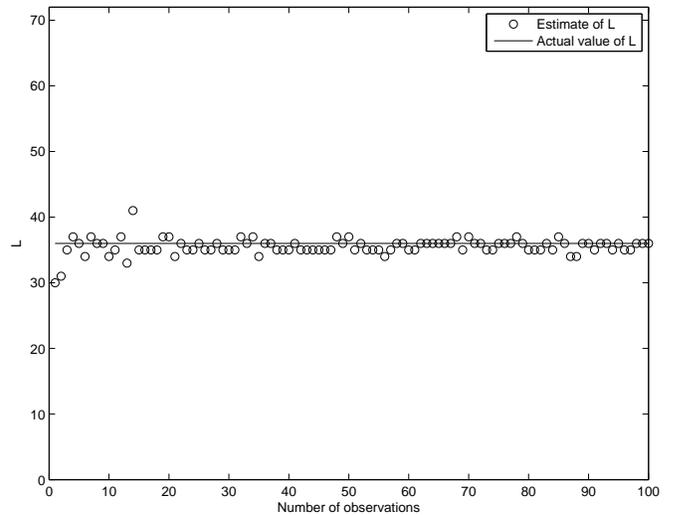
$$c(f) = \sum_{i=1}^L \alpha_i G(f) e^{-j2\pi f \tau_i}$$

For simplicity, we suppose the transmit filter to be ideal and therefore $G(f) = 1$. Sampling the continuous frequency signal at $f_i = i \frac{W}{N}$ where W is the bandwidth, the model becomes

$$\mathbf{V} = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & \cdots & 1 \\ e^{-j2\pi \frac{W\tau_1}{N}} & \cdots & e^{-j2\pi \frac{W\tau_L}{N}} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi(N-1) \frac{W\tau_1}{N}} & \cdots & e^{-j2\pi(N-1) \frac{W\tau_L}{N}} \end{pmatrix}, \quad (5)$$



(a) $K = 1$



(b) $K = 10$

Fig. 3. Estimate for the number of users. Actual value of L is 36. Also, $\sigma = \sqrt{0.1}$, $N = 100$.

We will here set $W = T = 1$, which means that the ω_i of (1) are uniformly distributed over $[0, 2\pi]$. Our model becomes

$$\mathbf{r} = \mathbf{V} \mathbf{P}^{\frac{1}{2}} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_L \end{pmatrix} + \begin{pmatrix} n_1 \\ \vdots \\ n_N \end{pmatrix}, \quad (6)$$

where L is the number of paths, N is the number of frequency samples, \mathbf{P} is the unknown $L \times L$ diagonal power matrix, and n_i is independent, additive, white, zero mean Gaussian noise of variance $\frac{\sigma}{\sqrt{N}}$. We take K observations of (6) and form the

observation matrix

$$\begin{aligned} \mathbf{Y} &= [\mathbf{r}_1 \cdots \mathbf{r}_K] \\ &= \mathbf{V}\mathbf{P}^{\frac{1}{2}} \begin{pmatrix} \alpha_1^{(1)} & \cdots & \alpha_1^{(K)} \\ \vdots & \ddots & \vdots \\ \alpha_L^{(1)} & \cdots & \alpha_L^{(K)} \end{pmatrix} \\ &\quad + \begin{pmatrix} n_1^{(1)} & \cdots & n_1^{(K)} \\ \vdots & \ddots & \vdots \\ n_N^{(1)} & \cdots & n_N^{(K)} \end{pmatrix}, \end{aligned} \quad (7)$$

which is the same model as (2), the only difference being that the phase distribution of the Vandermonde matrix now is uniform. In this case, we can do even better than proposition 1, in that one can write down estimators for the moments which are unbiased for any number of observations and frequency samples:

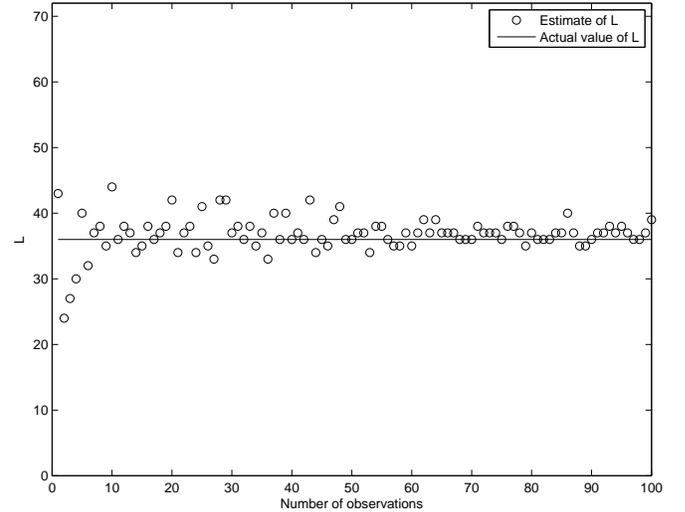
Proposition 2: Assume that \mathbf{V} has uniformly distributed phases, and let m_P^i be the moments of \mathbf{P} , and $m_R^i = \text{tr}_N(\hat{\mathbf{R}}^i)$ the moments of the sample covariance matrix. Define also $c_1 = \frac{N}{K}$, $c_2 = \frac{L}{N}$, and $c_3 = \frac{L}{K}$. Then

$$\begin{aligned} E[m_{\hat{R}}] &= c_2 m_P^1 + \sigma^2 \\ E[m_{\hat{R}}^2] &= c_2 \left(1 - \frac{1}{N}\right) m_P^2 + c_2(c_2 + c_3)(m_P^1)^2 \\ &\quad + 2\sigma^2(c_2 + c_3)m_P^1 + \sigma^4(1 + c_1) \\ E[m_{\hat{R}}^3] &= c_2 \left(1 + \frac{1}{K^2}\right) \left(1 - \frac{3}{N} + \frac{2}{N^2}\right) m_P^3 \\ &\quad + \left(1 - \frac{1}{N}\right) \left(3c_2^2 \left(1 + \frac{1}{K^2}\right) + 3c_2c_3\right) m_P^1 m_P^2 \\ &\quad + \left(c_2^3 \left(1 + \frac{1}{K^2}\right) + 3c_2^2c_3 + c_2c_3^2\right) (m_P^1)^3 \\ &\quad + 3\sigma^2 \left((1 + c_1)c_2 + \frac{c_1c_2^2}{KL}\right) \left(1 - \frac{1}{N}\right) m_P^2 \\ &\quad + 3\sigma^2 \left(\frac{c_1c_2^3}{KL} + c_2^2 + c_3^2 + 3c_2c_3\right) (m_P^1)^2 \\ &\quad + 3\sigma^4 \left(c_1^2 + 3c_1 + 1 + \frac{1}{K^2}\right) c_2 m_P^1 \\ &\quad + \sigma^6 \left(c_1^2 + 3c_1 + 1 + \frac{1}{K^2}\right) \end{aligned}$$

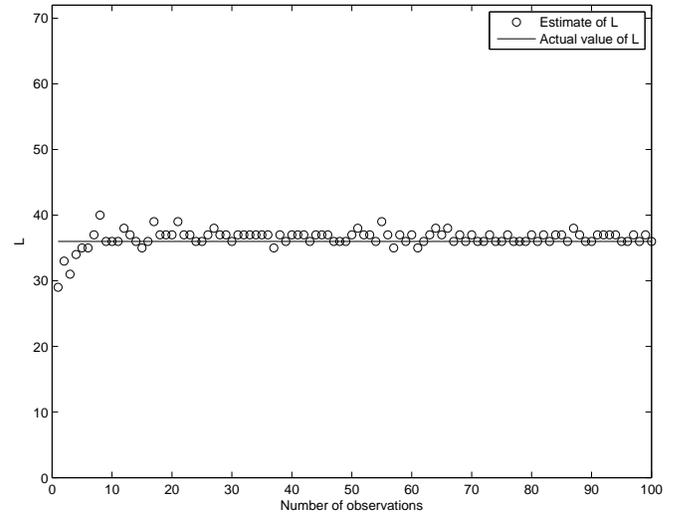
Just as proposition 1, this is proved in appendix A. In the following, this result is used in order to determine the number of paths as well as the power of each path. The different convergence rates of the approximations are clearly seen in the plots.

In figure 4, the number of paths is estimated based on the procedure sketched above. We have set $\sigma = \sqrt{0.1}$, $N = 100$, and $L = 36$. The procedure is tried for 1 all the way up to 100 observations. The plot is very similar to figure 3, in that only a small number of observations are needed in order to get an accurate estimate of L . When $K = 1$, it is seen that more observations are needed to get an accurate estimate of L , when compared to $K = 10$.

For the estimation of powers simulation, we have set $K = N = L = 144$, and $\sigma = \sqrt{0.1}$, following the procedure also described above, up to 1000 observations. The second plot in



(a) $K = 1$



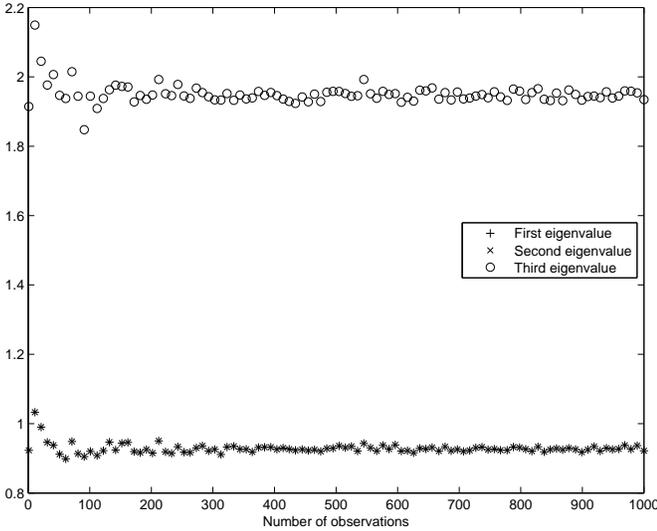
(b) $K = 10$

Fig. 4. Estimate for the number of paths. Actual value of L is 36. Also, $\sigma = \sqrt{0.1}$, $N = 100$.

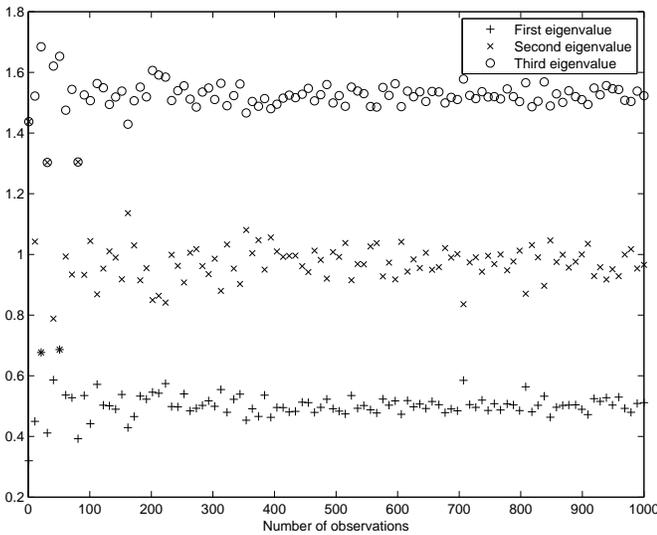
figure 5 shows the results which confirms the usefulness of the approach. We see that even for smaller matrix sizes than the model of the previous section, the estimates are much closer to the true powers.

C. Estimation of wavelength

In the field of MIMO cognitive sensing [22], [23], terminals must decide on the band on which to transmit and in particular sense which band is occupied. One way of doing is to find the wavelength λ in (3), based on some realizations of the sample covariance matrix. In our simulation, we have set $d = 1$ and $\lambda = 10$, $K = 10$, $L = 36$, $N = 100$, and $\sigma = \sqrt{0.1}$. We have tried the values 1, 2, ..., 100 as candidate wavelengths, and chosen the one which gives the smallest deviation (in the same sense as above, i.e. the sum of the squared errors of the first three moments are taken) from a different number of realizations of sample covariance matrices. The resulting plot is shown in figure 6, and shows that the



(a) Estimation of powers for various number of observations for the model (3) of section II-B. $K = N = L = 576$, and $\sigma = \sqrt{0.1}$.



(b) Estimation of powers for various number of observations for the model (5) of section II-A. $K = N = L = 144$, and $\sigma = \sqrt{0.1}$.

Fig. 5. Estimation of powers for the two models (3) and (5) of this section, for various number of observations.

Vandermonde deconvolution method can also be used for wavelength estimation. It is seen that the estimation gets better when the number of observations is increased.

III. WIRELESS CAPACITY ANALYSIS

A. General Results on Capacity and moments

For a general matrix \mathbf{W} , consider the mean capacity defined as

$$\begin{aligned} C_N &= \frac{1}{N} \mathbb{E} \left(\log_2 \det \left(\mathbf{I}_N + \frac{1}{\sigma^2} \mathbf{W} \mathbf{W}^H \right) \right) \\ &= \frac{1}{N} \sum_{k=1}^N \mathbb{E} \left(\log_2 \left(1 + \frac{1}{\sigma^2} \lambda_k \left(\mathbf{W} \mathbf{W}^H \right) \right) \right) \quad (8) \\ &= \int \log_2 \left(1 + \frac{1}{\sigma^2} t \right) \mu(dt), \end{aligned}$$

where μ is the mean empirical eigenvalue distribution of $\mathbf{W} \mathbf{W}^H$ i.e. $\mathbb{E} \left(\frac{1}{N} \sum_{i=1}^N \delta(\lambda - \lambda_i) \right)$ where λ_i are the eigenvalues of $\mathbf{W} \mathbf{W}^H$. In practice, we do not need the expectation in

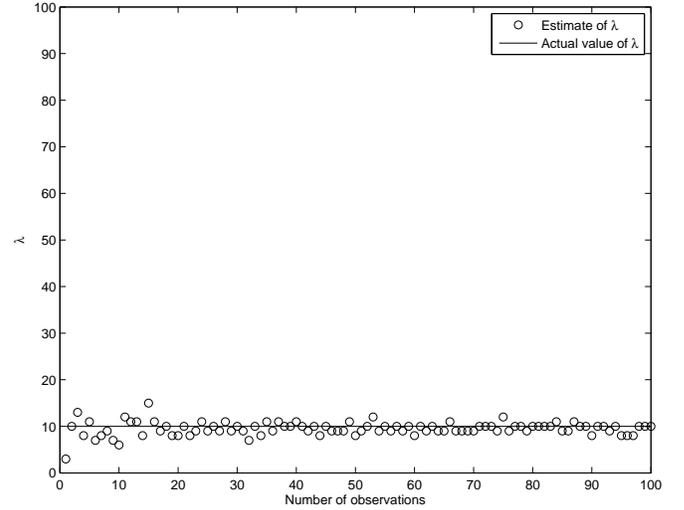


Fig. 6. Estimation of wavelength. Deconvolution was performed for varying number of observations, assuming different wavelengths. In the true model (3), $d = 1$, $\lambda = 10$, $K = 10$, $L = 36$, $N = 100$, and $\sigma = \sqrt{0.1}$.

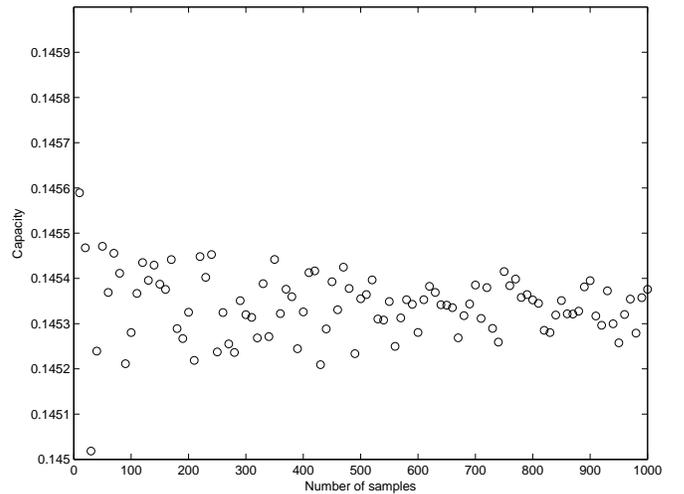


Fig. 7. Estimation of the true capacity $\frac{1}{N} \log_2 \det (\mathbf{I} + \rho \mathbf{V} \mathbf{V}^H)$ when the Vandermonde matrix \mathbf{V} has uniform phase distribution. \mathbf{V} has size 36×36 , and $\sigma = 3$ was used. The estimate is obtained from different number of samples approximating the true capacity.

(8), but rather asymptotic formulas (i.e. formulas which apply when $N \rightarrow \infty$ without the expectation operator). Asymptotic formulas for the capacity when \mathbf{W} is a Vandermonde matrix are not known, contrary to the case for Gaussian matrices: Since no exact formulas for the asymptotic capacity of Vandermonde matrices are known, we will instead obtain good estimates for it by taking an average of (8) (as defined) over many samples. In figure 7, capacity estimates for $\sigma = 3$ obtained this way up to 1000 samples are shown for 36×36 Vandermonde matrices with uniform phase distribution. It is seen that also for much small number of samples, all capacity estimates are between 0.145 and 0.146. 200 samples will be used in the simulations, since this gives a value close to the value the capacities seem to converge to.

Substituting the Taylor series

$$\log_2(1+t) = \frac{1}{\ln 2} \sum_{k=1}^{\infty} \frac{(-1)^{k+1} t^k}{k} \quad (9)$$

in (8), we obtain that

$$\begin{aligned} C_N &= \frac{1}{\ln 2} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k \sigma^{2k}} \int t^k \mu(dt) \\ &= \frac{1}{\ln 2} \sum_{k=1}^{\infty} \frac{(-1)^{k+1} m_k(\mu)}{k \sigma^{2k}} \\ &= \frac{1}{\ln 2} \sum_{k=1}^{\infty} \frac{(-1)^{k+1} m_k(\mu) \rho^k}{k}, \end{aligned} \quad (10)$$

where ρ is SNR, and where

$$m_k(\mu) = \int t^k d\mu(t) \text{ for } k \in \mathbb{Z}^+$$

are the moments of μ . When \mathbf{W} is a Vandermonde matrix, it is unknown whether this series converges. However, one can still assess if a finite partial sum of (10) is a good approximation to the actual capacity. In the following, this will be done for applications to capacity of line of sight of multi-user MIMO systems, and multifold scattering. Up to 7 terms in (10) are used in the assessments, since [7] has computed the seven first moments of the Vandermonde matrices. In the examples, low SNR values (i.e. high σ -values) will be used, since the series (10) will only converge then. This follows from the fact that the radius of convergence of (9) around 0 is 1, so that (10) can converge only when $\rho \leq \frac{1}{\gamma_2}$, where the support of the mean eigenvalue distribution of the Vandermonde matrix is denoted by $[\gamma_1, \gamma_2]$.

One can ask the question whether one can circumvent the problem with the convergence radius of (9) by using a different Taylor expansion, so that this expansion can be used to obtain the capacity at all SNR when one has the first order moments only. We will indeed show that this is possible, but that it in practice will require computation of many moments (the first seven moments as computed in [7] are not enough to obtain good approximations). Therefore, current methods do not suffice in capacity estimation for Vandermonde matrices at all SNR, so that an extension of the methods from [7], or new methods are needed.

To the end of coming up with a different Taylor expansion than (9), consider the following, where t_0 is any positive constant:

$$\begin{aligned} \log(1+\rho t) &= \log(1+t_0\rho + \rho(t-t_0)) \\ &= \sum_{k=1}^{\infty} \frac{(-1)^{k+1} \rho^k}{k(1+t_0\rho)^k} (t-t_0)^k. \end{aligned} \quad (11)$$

In order for this series to be accurate for few terms, we need to ensure that $\left| \frac{\rho(t-t_0)}{1+t_0\rho} \right| < 1$ when t is in the support $[\gamma_1, \gamma_2]$ of μ .

Assume first that $0 \leq t_0 \leq \frac{\gamma_2}{2}$. Then

$$\left| \frac{\rho(t-t_0)}{1+t_0\rho} \right| \leq \frac{\rho(\gamma_2-t_0)}{1+t_0\rho}$$

for all $t \in [\gamma_1, \gamma_2]$. It is easily checked that the latter is ≤ 1 when $\rho \leq \frac{1}{\gamma_2 - 2t_0}$. One easily checks also that $\frac{\rho(\gamma_2-t_0)}{1+t_0\rho}$ is a decreasing function of t_0 . We conclude from this that $t_0 = \frac{\gamma_2}{2}$ is the value in $[0, \frac{\gamma_2}{2}]$ which makes our power series converge fastest.

Assume now that $\frac{\gamma_2}{2} \leq t_0$. Then

$$\left| \frac{\rho(t-t_0)}{1+t_0\rho} \right| \leq \frac{\rho t_0}{1+t_0\rho} \leq 1$$

for all $t \in [\gamma_1, \gamma_2]$, so that our Taylor series converges for all ρ for such t_0 . Also, the $t_0 \geq \frac{\gamma_2}{2}$ which makes $\frac{\rho t_0}{1+t_0\rho}$ smallest is $t_0 = \frac{\gamma_2}{2}$. This "proves" that expansion around $t_0 = \frac{\gamma_2}{2}$ is the optimal choice for our Taylor expansion, and that it converges for all choices of ρ in this case.

Let us attempt to compute how many terms are needed in the Taylor expansion (11) in order to estimate the capacity at $\rho = 10$ with accuracy better than 0.5, for Vandermonde phase distributions given by (3) and (5). Due to the above considerations, we use $t_0 = \frac{\gamma_2}{2}$. If N terms in (11) are used to approximate the capacity, and for $t \leq \frac{\gamma_2}{2}$, consider the remainder term

$$\begin{aligned} &\left| \sum_{k=N+1}^{\infty} \frac{(-1)^{k+1} \rho^k}{k \left(1 + \frac{\gamma_2}{2} \rho\right)^k} \left(t - \frac{\gamma_2}{2}\right)^k \right| \\ &\leq \sum_{k=N+1}^{\infty} \frac{1}{k} \left(\frac{\rho \frac{\gamma_2}{2}}{1 + \frac{\gamma_2}{2} \rho} \right)^k \\ &\leq \frac{1}{N+1} \left(\frac{\rho \frac{\gamma_2}{2}}{1 + \frac{\gamma_2}{2} \rho} \right)^{N+1} \frac{1}{1 - \frac{\rho \frac{\gamma_2}{2}}{1 + \frac{\gamma_2}{2} \rho}} \\ &= \frac{\rho \gamma_2}{2(N+1)} \left(\frac{\rho \frac{\gamma_2}{2}}{1 + \frac{\gamma_2}{2} \rho} \right)^N, \end{aligned}$$

which is a good bound if much of a large proportion of the eigenvalues of the Vandermonde matrix is close to the origin (this can be inspected from the histograms of the eigenvalue of Vandermonde matrices in this paper and in [7]).

From figures 4,5, and 6 in [7], we see that $\gamma_2 \approx 5$ is a good guess for the upper bound of the support when the phase distribution has the form (5) (although we haven't proved that γ_2 even exists). For uniform phase distribution (5), the number of terms N thus needs to be chosen so that

$$\frac{25}{N+1} \left(\frac{25}{26} \right)^N \leq 0.5$$

for the given precision and ρ . It is easily checked that $N = 21$ is the lowest number of terms which makes this possible, so that 21 terms in the Taylor expansion (11), and thus the 21 first moments, are needed to obtain the required precision.

From figure 2 we see that $\gamma_2 \approx 15$ is a good guess for the upper bound of the support when the phase distribution has the form (3). For this case, the number of terms N thus needs to be chosen so that

$$\frac{75}{N+1} \left(\frac{75}{76} \right)^N \leq 0.5$$

for the given precision and ρ . It is easily checked that $N = 64$ is the lowest number of terms which makes this possible, so that 64 terms in the Taylor expansion (11), and thus the 64 first moments, are needed to obtain the required precision. This clearly indicates that the moments computed in [7] do not suffice.

B. Capacity of line of sight of multi-user MIMO systems

We consider a base station equipped with N antennas and L mobiles (each with a single antenna) in the cell. The received signal at the base station is given by

$$\mathbf{r}_i = \mathbf{C}\mathbf{P}^{\frac{1}{2}}\mathbf{s}_i + \mathbf{n}_i$$

Here, \mathbf{r}_i , \mathbf{s}_i , \mathbf{n}_i are respectively the $N \times 1$ received vector, the $L \times 1$ transmit vector by the L users and the $N \times 1$ additive noise. In the case of line of sight between the users and the base station, the model

$$\mathbf{V} = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & \dots & 1 \\ e^{-j2\pi\frac{d}{\lambda}\sin(\theta_1)} & \dots & e^{-j2\pi\frac{d}{\lambda}\sin(\theta_L)} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi(N-1)\frac{d}{\lambda}\sin(\theta_1)} & \dots & e^{-j2\pi(N-1)\frac{d}{\lambda}\sin(\theta_L)} \end{pmatrix} \quad (12)$$

applies. Here, θ_i is the angle of the user in the cell and is supposed to be uniformly distributed over $[-\alpha, \alpha]$. $\mathbf{P}^{\frac{1}{2}}$ is a $L \times 1$ power matrix due to the different distances from which the users emit.

In this case, we would like to derive the mean channel capacity (8) per dimension for random positions of the users of the system, which is given by

$$C = \frac{1}{N} \mathbb{E} \left(\log_2 \det \left(\mathbf{I} + \frac{1}{\sigma^2} \mathbf{V}\mathbf{P}\mathbf{V}^H \right) \right). \quad (13)$$

Note that the moments of $\mathbf{P}\mathbf{V}^H\mathbf{V}$ were computed in [7]. When the phases of \mathbf{V} are uniformly distributed, theorem 3 in that paper expresses the first 7 such mixed moments in terms of so-called *Vandermonde expansion coefficients*, which also are computed. Also, theorem 5 of the same paper tells us how to get the expansion coefficients when the phase distribution has a continuous density: To compute the expansion coefficients for this case, we have to compute the density moments (4) numerically. We will therefore do the following:

- 1) Compute the first seven density moments (4), and from this get the the expansion coefficient up to order 7.
- 2) Estimate the first seven moments of $\mathbf{P}\mathbf{V}^H\mathbf{V}$ using the result in [7] (note that the moments of $\mathbf{P}\mathbf{V}^H\mathbf{V}$ are obtained from those of $\mathbf{P}\mathbf{V}^H\mathbf{V}$ by multiplication by c).
- 3) Substitute these in (10) to get a capacity estimate.

In the simulations, these steps are followed to obtained a capacity estimate. Different number of terms in the approximation (10) are used to see how fast the series converges to the true capacity. Note that there is no known formula for the true capacity in this case. The true capacity is instead also approximated, by taking 1000 samples of the random matrices involved in the expression for the capacity. In figure 8, the SNR-values where the Vandermonde convolution capacity estimates begin to converge to the actual value are clearly seen. It is perhaps surprising that the approximation with fewer terms is more accurate for higher SNR values. This is the case since the power series approximation is better only close to the origin when more terms are added: Far away from the origin, the approximation can be worse when more terms are added (but will eventually improve when even more terms are added).

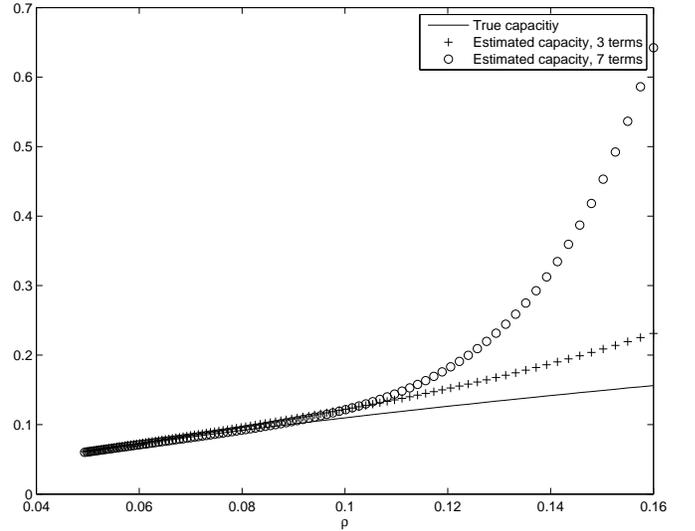


Fig. 8. Estimation of channel capacity for the three and seven terms approximations (10). The actual capacity is also shown. SNR up to 0.16 have been tested. $L = N = 36$, $d = 1$, $\lambda = 10$.

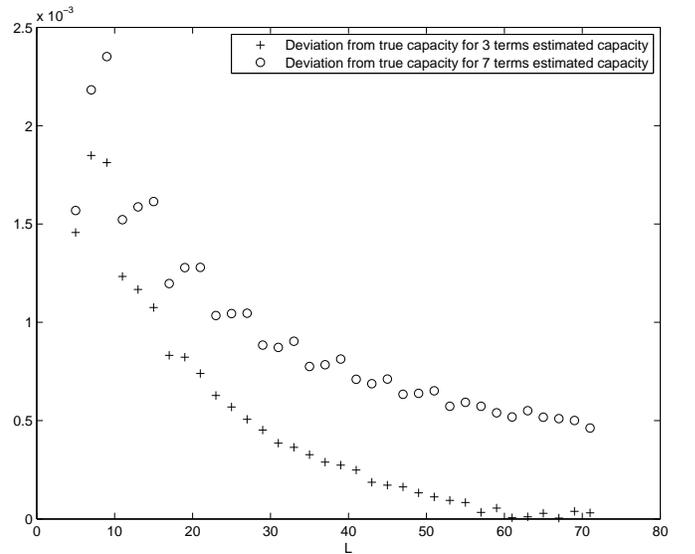


Fig. 9. Deviation from the true capacity for the three and seven terms approximations (10). Various L have been tested. $\rho = 0.04$, $N = L$, $d = 1$, $\lambda = 10$.

In figure 9, ρ has been fixed at 0.04, and different values of L have been tried in the capacity estimation. It is seen that the capacity estimated is better for higher L . The explanation is that the Vandermonde convolution formulas are asymptotic, so that they give better approximations for higher L and N . Also here it is seen that the 7-term approximation is worse. We can't expect that the deviation goes to zero for higher L , since the error not only lies in the (asymptotic) Vandermonde convolution formulas, but also in the approximation (10) (and the error from this part does not go away when we increase L).

C. Multifold scattering

We consider a MIMO system (between two users equipped with N Uniform Linear Array antennas and L scatterers in between). The received signal in one end of the MIMO system is given by

$$\mathbf{r}_i = \mathbf{V}_1^H \mathbf{P}^{\frac{1}{2}} \mathbf{V}_2 \mathbf{s}_i + \mathbf{n}_i \quad (14)$$

where

$$\mathbf{V}_1 = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & \cdots & 1 \\ e^{-j2\pi \frac{d}{\lambda} \sin(\theta_1)} & \cdots & e^{-j2\pi \frac{d}{\lambda} \sin(\theta_L)} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi(N-1) \frac{d}{\lambda} \sin(\theta_1)} & \cdots & e^{-j2\pi \frac{d}{\lambda} \sin(\theta_L)} \end{pmatrix}$$

and

$$\mathbf{V}_2 = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & \cdots & 1 \\ e^{-j2\pi \frac{d}{\lambda} \sin(\phi_1)} & \cdots & e^{-j2\pi \frac{d}{\lambda} \sin(\phi_L)} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi(N-1) \frac{d}{\lambda} \sin(\phi_1)} & \cdots & e^{-j2\pi \frac{d}{\lambda} \sin(\phi_L)} \end{pmatrix},$$

where all θ_i and ϕ_i are independent.

The scatterers distort the signal with attenuation P_i . This model has already been studied in [14] using an approximation of the Vandermonde matrix by an i.i.d zero mean random matrix. As shown in [7], this is not the case as the limiting eigenvalues of the Gram matrix associated to the Vandermonde matrix are quite different from the Marčenko Pastur law [9]. The mean capacity (8) per received dimension is given by

$$C = \frac{1}{N} \mathbb{E} \left(\log_2 \det \left(\mathbf{I} + \frac{1}{\sigma^2} \mathbf{V}_1^H \mathbf{P}^{\frac{1}{2}} \mathbf{V}_2 \mathbf{V}_2^H \mathbf{P}^{\frac{1}{2}} \mathbf{V}_1 \right) \right). \quad (15)$$

We will assume that $\mathbf{P} = \mathbf{I}$ (other case are more involved). In other words, we need the moments of $\mathbf{V}_1^H \mathbf{V}_2 \mathbf{V}_2^H \mathbf{V}_1$. To get these, we can use theorem 9 of [7]. Also here we are restricted to getting capacity estimates away from 0, since we only have the lower order moments available. In this case, it is also expected that we would need more observations to get good capacity estimates: Figure 10 shows the approximation of the true capacity (15) by taking many samples of model (14), similarly to figure 7 for Vandermonde matrices with uniform phase distribution. 36×36 matrices were used, and $\sigma = 3$. The values are seen to be between 0.35 and 0.36. For figure 7, one could infer yet another decimal.

D. Capacity of Gaussian versus Vandermonde models

If \mathbf{X} is an $N \times N$ standard, complex, Gaussian matrix, then an explicit expression of the capacity exists [24]

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} \log_2 \det \left(\mathbf{I} + \rho \left(\frac{1}{N} \mathbf{X} \mathbf{X}^H \right) \right) &= \\ 2 \log_2 \left(1 + \rho - \frac{1}{4} (\sqrt{4\rho + 1} - 1)^2 \right) & \quad (16) \\ - \frac{\log_2 e}{4\rho} (\sqrt{4\rho + 1} - 1)^2 & . \end{aligned}$$

In figure 11, several realizations of the capacity are computed for Gaussian matrix samples of size 36×36 . The asymptotic capacity (16) is also shown. In figure 12, several realizations of the capacity are computed for Vandermonde samples of the same size, with the phase distributions given by (5) and (3). Realizations of (15) are also shown. It is seen that

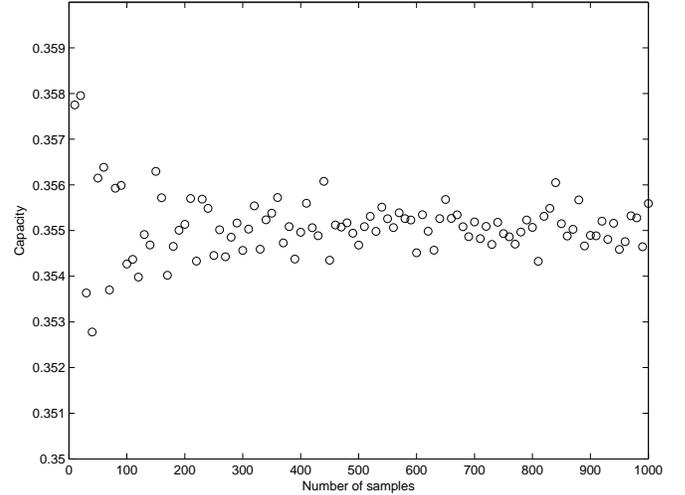


Fig. 10. Estimation of the true capacity obtained from different number of samples approximating the expectation in (15). 36×36 matrices were used, and $\sigma = 3$.

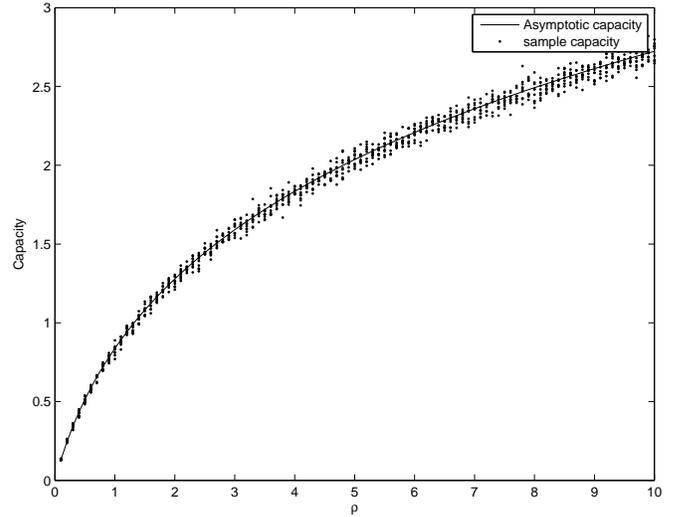
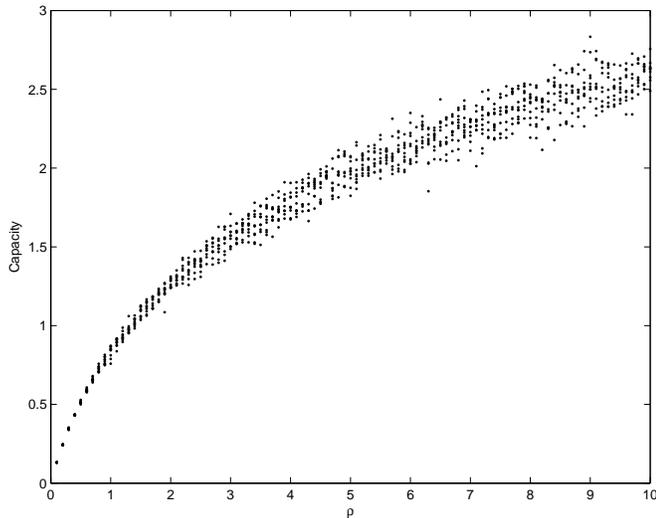
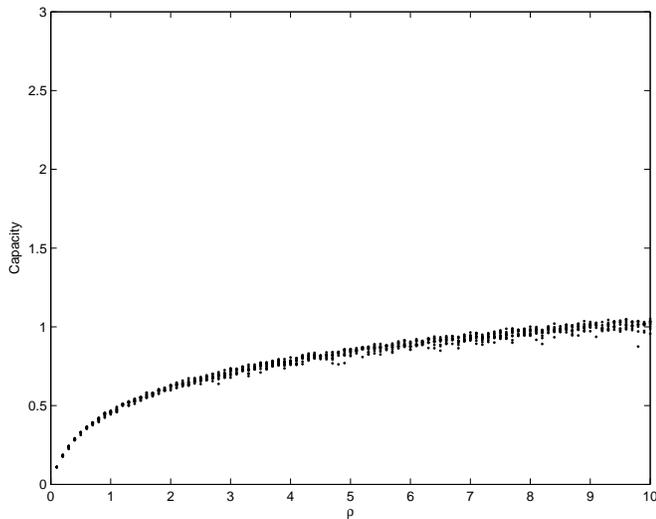


Fig. 11. Several realizations of the capacity $\frac{1}{N} \log_2 \det \left(\mathbf{I} + \rho \frac{1}{N} \mathbf{X} \mathbf{X}^H \right)$ when \mathbf{X} is standard, complex, Gaussian. Matrices of size 36×36 were used. The asymptotic capacity (16) is also shown.

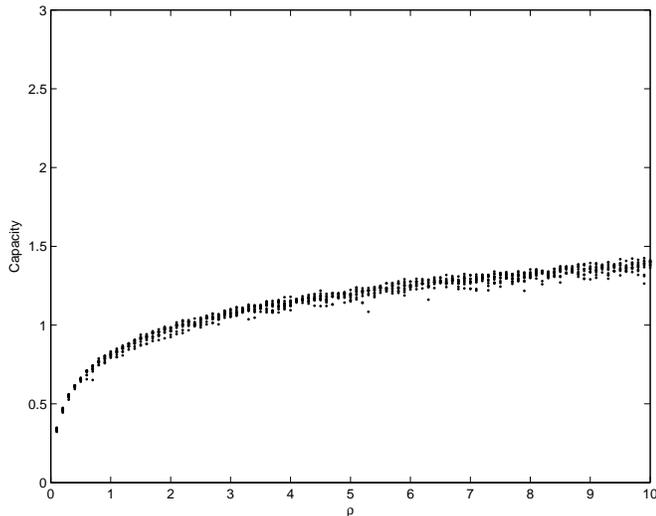
the uniform phase distribution gives highest values for the capacity. The effect of using two independent Vandermonde matrices is seen to increase the capacity somewhat. It is seen that the variance of the Vandermonde capacities is higher than for the Gaussian counterparts. This should come as no surprise, due to the slower convergence to the asymptotic limits for Vandermonde matrices [7]. Although the capacities of Vandermonde matrices with uniform phase distribution and Gaussian matrices seem to be close, we have actually no proof that the capacities of Vandermonde matrices are finite, since it is unknown whether the Vandermonde matrix (1) has a compactly supported limiting eigenvalue distribution [7], although extensive simulations seem to confirm this fact. In any case, the results show that a structured ULA form of the antenna geometry provides a decrease of capacity with respect



(a) Realizations of $\frac{1}{N} \log_2 \det(\mathbf{I} + \rho \mathbf{V}\mathbf{V}^H)$ when ω has uniform phase distribution (5).



(b) Realizations of $\frac{1}{N} \log_2 \det(\mathbf{I} + \rho \mathbf{V}\mathbf{V}^H)$ when ω has the phase distribution (3).



(c) Realizations of (15).

Fig. 12. Several realizations of the capacity for Vandermonde matrices for the two phase distributions used in this paper.

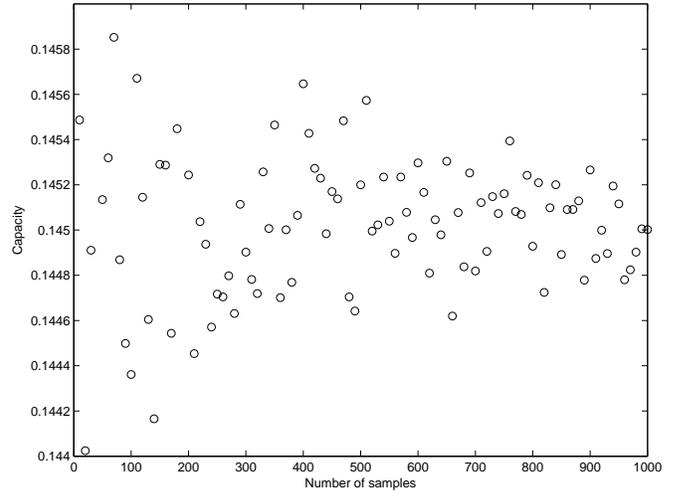


Fig. 13. Estimation of the true capacity $\frac{1}{N} \log_2 \det(\mathbf{I} + \rho \frac{1}{N} \mathbf{X}\mathbf{X}^H)$ when \mathbf{X} is complex standard, Gaussian. 36×36 matrices were used, and $\sigma = 3$. The estimate is obtained from different number of samples approximating the true capacity.

to having a more random like geometry in the case of line of sight environments.

Figure 13 shows the number of samples needed to get accurate estimation of the capacity for Gaussian matrices, similarly to how this was done for Vandermonde matrices with uniform phase distribution in figure 7. The estimated values are seen to lie between 0.144 and 0.146, which is very close to the values we obtained for Vandermonde matrices. Same matrix sizes (36×36) and value for σ (3) was used.

The fact that the capacity of line of sight Vandermonde matrices is in simulation lower than the capacity of Gaussian matrices is a very interesting issue and permits to understand the differences between line of sight and non-line of sight environments. Interestingly, the moments of structured Vandermonde matrices with uniformly distributed phases are always larger than the moments of Gaussian i.i.d. matrices, see corollary 2 in [7]. When the phase distribution is given by (3), a similar result holds due to theorem 5 in [7]. However, one can not prove from these results any effective relation between the capacities.

IV. SIGNAL RECONSTRUCTION

Several works have investigated how irregular sampling affects the performance of signal reconstruction in the presence of noise in different fields namely sensor networks [25], [26], image processing [27], [28], geophysics [29], compressive sampling [30]. The usual Nyquist theorem states that for a signal with maximum frequency f_{\max} , one needs to sample the signal at a rate which is at least twice this number. However, in many cases, this can not be performed or one has an observation of a signal at only a subset of the frequencies. Moreover, one feels that if the signal has a sparse spectrum, one can take fewer samples and still have the same information on the original signal. One of the central motivations of sparse sampling is exactly to understand under which condition one can still have less samples and recover the original signal up

to an error of ϵ [31]. Let us consider the signal of interest as a superposition of its frequency components (this is also the case for a unidimensional bandlimited physical signal) i.e

$$x(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} a_k e^{-\frac{j2\pi kt}{N}}$$

and suppose that the signal is sampled at various instants $[t_1, \dots, t_L]$ with $t_i \in [0, 1]$. This can be identically written as

$$x(\omega) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} a_k e^{-jk\omega}.$$

In the presence of noise, one can write it in the following manner:

$$\mathbf{x} = \mathbf{V}^T \mathbf{a} + \mathbf{n}$$

with $\mathbf{x} = [x(\omega_1), \dots, x(\omega_L)]^T$, $\mathbf{a} = [a_1, \dots, a_N]^T$, $\mathbf{n} = [n_1, \dots, n_L]$ and

$$\mathbf{V} = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & \dots & 1 \\ e^{-j\omega_1} & \dots & e^{-j\omega_L} \\ \vdots & \ddots & \vdots \\ e^{-j(N-1)\omega_1} & \dots & e^{-j(N-1)\omega_L} \end{pmatrix}. \quad (17)$$

We define c as the ratio of observations to the number of complex harmonics i.e $c = \frac{L}{N}$

A. Performance analysis of the reconstruction algorithm

The task of the reconstruction algorithm is to calculate an estimate $\hat{\mathbf{a}}$ of the spectrum \mathbf{a} . The usual reconstruction metric is the minimum mean square error which is defined as

$$\text{MMSE} = \frac{1}{N} \mathbb{E} \|\hat{\mathbf{a}} - \mathbf{a}\|^2.$$

The linear filter which minimizes the MMSE or maximizes the Signal to Interference plus noise ratio is known to be the MMSE filter. Results on the interplay between information theory and estimation theory [32], [33] show that the MMSE is strongly related to the eigenvalues of $\mathbf{V}\mathbf{V}^T$.

In particular, one has that:

$$\text{MMSE} = \frac{dC}{\rho},$$

where $C = \frac{1}{N} \log \det (\mathbf{I} + \rho \mathbf{V}\mathbf{V}^T)$ with ρ being the SNR.

Asymptotically, when $N \rightarrow \infty$ and $L \rightarrow \infty$ such as $\frac{L}{N} \rightarrow c$, the MMSE depends only on the SNR ρ , the ratio c and the probability distribution of the sampling p_w . The MMSE can be computed in the same vein as previously with a Taylor approximation for a given distribution of the sampling in the low SNR regime.

B. Estimation of the sampling distribution

In the following, we suppose that one has K observations of the received sampled vector \mathbf{x} :

$$\mathbf{Y} = [\mathbf{x}_1, \dots, \mathbf{x}_K] = \mathbf{V}^T [\mathbf{a}_1, \dots, \mathbf{a}_K] + [\mathbf{n}_1, \dots, \mathbf{n}_K]$$

The vector \mathbf{x} is the discrete output of the sampled continuous signal $x(w)$ for which the distribution is unknown (however, c is known). This case happens when one has an observation without the knowledge of the sampling rate for example. As recalled in section IV of [7], the problem falls in the realm of deconvolution and one is able to infer on the density of the density of w in other words finding the moment of

$$\int_0^{2\pi} p_w(x)^k dx.$$

If we define

$$\hat{\mathbf{R}} = \frac{1}{K} \mathbf{Y}\mathbf{Y}^H, \quad (18)$$

and I_n by (4), then by successive Gaussian deconvolution as detailed in section II, we have

$$\begin{aligned} m_{\hat{\mathbf{R}}}^1 &= c_2 + \sigma^2 \\ m_{\hat{\mathbf{R}}}^2 &= c_2 + (c_2^2 I_2 + c_2 c_3) \\ &\quad + 2\sigma^2(c_2 + c_3) + \sigma^4(1 + c_1) \\ m_{\hat{\mathbf{R}}}^3 &= c_2 + (3c_2^2 I_2 + 3c_2 c_3) \\ &\quad + (c_2^3 I_3 + 3c_2^2 c_3 I_2 + c_2 c_3^2) \\ &\quad + 3\sigma^2(1 + c_1)c_2 \\ &\quad + 3\sigma^2((1 + c_1)c_2^2 I_2 + c_3(c_3 + 2c_2)) \\ &\quad + 3\sigma^4(c_1^2 + 3c_1 + 1)c_2 \\ &\quad + \sigma^6(c_1^2 + 3c_1 + 1), \end{aligned} \quad (19)$$

where $m_{\hat{\mathbf{R}}}^i = E \left[\text{tr}_N(\hat{\mathbf{R}}^i) \right]$, $\lim_{N \rightarrow \infty} \frac{N}{K} = c_1$, $\lim_{N \rightarrow \infty} \frac{L}{N} = c_2$, and $\lim_{N \rightarrow \infty} \frac{L}{K} = c_3$. The estimation of I_n is averaged on the various set of K observations.

We have tested (19) the following way: We have taken a phase distribution ω which is uniform on $[0, \alpha]$, and 0 elsewhere. The density is thus $\frac{2\pi}{\alpha}$ on $[0, \alpha]$, and 0 elsewhere. In this case we can compute that

$$\begin{aligned} I_2 &= \frac{2\pi}{\alpha} \\ I_3 &= \left(\frac{2\pi}{\alpha} \right)^2. \end{aligned}$$

The first of these equations, combined with (19), enables us to estimate α from the observations (18). This is tested in figure 15 for various number of observations. In figure 14 we have also tested estimation of I_2, I_3 from the observations using the same equations. When one has a distribution which is not uniform, the integrals I_3, I_4, \dots would also be needed in finding the characteristics of the underlying phase distribution. Figure 14 shows that the estimation of I_2 requires far fewer observation than the estimation of I_3 . In both figures, the values $K = 10, L = 36, N = 100$, and $\sigma = \sqrt{0.1}$ were used and α was $\frac{\pi}{4}$.

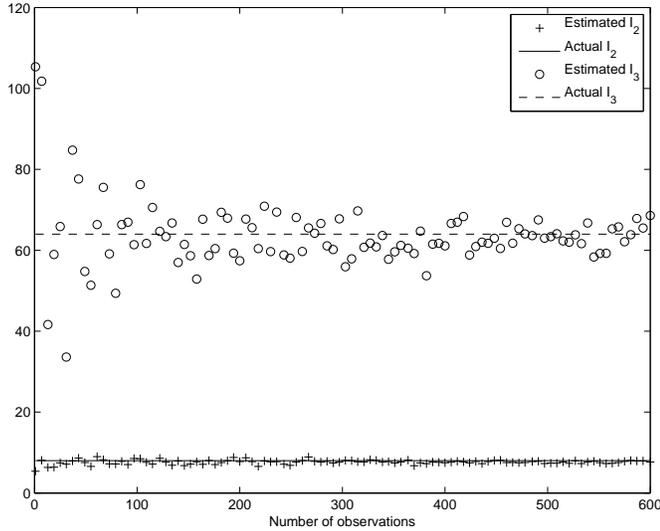


Fig. 14. Estimated values of I_2 and I_3 using (19), for various number of observations, and for $K = 10, L = 36, N = 100, \sigma = \sqrt{0.1}$. The actual value of α was $\frac{\pi}{4}$.

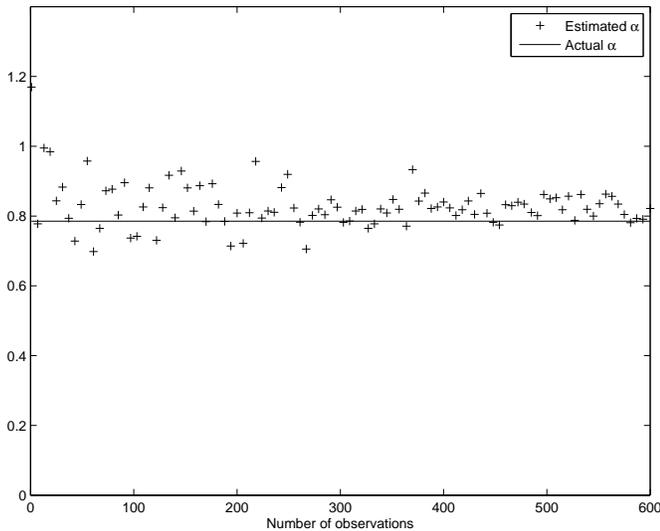


Fig. 15. Estimated values of α using (19), for various number of observations, and for $K = 10, L = 36, N = 100, \sigma = \sqrt{0.1}$. The actual value of α was $\frac{\pi}{4}$.

V. CONCLUSION

In this part, we have provided some useful applications of random Vandermonde matrices. The applications concentrated on wireless capacity analysis, source separation and signal sampling analysis. As shown, many useful system models use independent Vandermonde matrices and Gaussian matrices combined in some way. The presented examples show how random Vandermonde matrices applied for such systems can be handled in practice to obtain estimates on quantities such as the number of paths in channel modeling, the transmission powers of the users in wireless transmission or the sampling distribution for signal recovery. The paper has only touched upon a limited number of applications but the results already

provide benchmark figures in the non-asymptotic regime.

APPENDIX A

THE PROOFS OF PROPOSITION 1 AND 2

The moments and fluctuations of

$$\mathbf{\Gamma} = \frac{1}{K} \mathbf{R} \mathbf{R}^H = \mathbf{V} \mathbf{P}^{\frac{1}{2}} \left(\frac{1}{K} \alpha \alpha^H \right) \mathbf{P}^{\frac{1}{2}} \mathbf{V}^H.$$

are related to the moments of \mathbf{W} through the formulas [16]

$$\begin{aligned} E[tr_n(\mathbf{W})] &= E[tr_N(\mathbf{\Gamma})] + \sigma^2 \\ E[tr_n(\mathbf{W}^2)] &= E[tr_N(\mathbf{\Gamma}^2)] \\ &\quad + 2\sigma^2(1 + c_1)E[tr_N(\mathbf{\Gamma})] + \sigma^4(1 + c_1) \\ E[tr_n(\mathbf{W}^3)] &= E[tr_N(\mathbf{\Gamma}^3)] \\ &\quad + 3\sigma^2(1 + c_1)E[tr_N(\mathbf{\Gamma}^2)] \\ &\quad + 3\sigma^2 c_1 E[(tr_N(\mathbf{\Gamma}))^2] \\ &\quad + 3\sigma^4 \left(c_1^2 + 3c_1 + 1 + \frac{1}{K^2} \right) E[tr_N(\mathbf{\Gamma})] \\ &\quad + \sigma^6 \left(c_1^2 + 3c_1 + 1 + \frac{1}{K^2} \right), \end{aligned} \quad (20)$$

where $c_1 = \frac{N}{K}$. Define the matrix

$$\mathbf{S} = \left(\frac{1}{K} \alpha \alpha^H \right) \mathbf{P}^{\frac{1}{2}} \mathbf{V}^H \mathbf{V} \mathbf{P}^{\frac{1}{2}},$$

and note that

$$\begin{aligned} E[tr_N(\mathbf{\Gamma}^k)] &= c_2 E[tr_L(\mathbf{S}^k)], \text{ and} \\ E[(tr_N(\mathbf{\Gamma}))^k] &= c_2^k E[(tr_L(\mathbf{S}))^k], \end{aligned} \quad (21)$$

where $c_2 = \frac{L}{N}$. We can now use the formulas [7]

$$\begin{aligned} c_3 E[tr_L(\mathbf{S})] &= c_3 E[tr_L(\mathbf{T})] \\ c_3 E[tr_L(\mathbf{S}^2)] &= c_3 E[tr_L(\mathbf{T}^2)] \\ &\quad + c_3^2 E[(tr_L(\mathbf{T}))^2] \\ c_3 E[tr_L(\mathbf{S}^3)] &= (1 + K^{-2}) c_3 E[tr_L(\mathbf{T}^3)] \\ &\quad + 3c_3^2 E[(tr_L(\mathbf{T})) tr_L(\mathbf{T}^2)] \\ &\quad + c_3^3 E[(tr_L(\mathbf{T}))^3] \\ E[(tr_L(\mathbf{S}))^2] &= E[(tr_L(\mathbf{T}))^2] \\ &\quad + \frac{1}{KL} E[tr_L(\mathbf{T}^2)] \end{aligned} \quad (22)$$

where $c_3 = \frac{L}{K}$, and $\mathbf{T} = \mathbf{P} \mathbf{V}^H \mathbf{V}$. (20), (21), and (22) can be combined to the following form (m_W^i and m_T^i are the i th moments of \mathbf{W} and \mathbf{T} respectively):

$$\begin{aligned} m_W^1 &= c_2 m_T^1 + \sigma^2 \\ m_W^2 &= c_2 m_T^2 + c_2 c_3 E[(tr_L(\mathbf{T}))^2] \\ &\quad + 2\sigma^2(c_2 + c_3)E[tr_L(\mathbf{T})] + \sigma^4(1 + c_1) \\ m_W^3 &= c_2 \left(1 + \frac{1}{K^2} \right) E[tr_L(\mathbf{T}^3)] \\ &\quad + 3c_2 c_3 E[(tr_L(\mathbf{T})) (tr_L(\mathbf{T}^2))] \\ &\quad + c_2 c_3^2 E[(tr_L(\mathbf{T}))^3] \\ &\quad + 3\sigma^2 \left((1 + c_1)c_2 + \frac{c_1 c_2^2}{KL} \right) E[tr_L(\mathbf{T}^2)] \\ &\quad + 3\sigma^2 c_3 (c_3 + 2c_2) E[(tr_L(\mathbf{T}))^2] \\ &\quad + 3\sigma^4 \left(c_1^2 + 3c_1 + 1 + \frac{1}{K^2} \right) c_2 E[tr_L(\mathbf{T})] \\ &\quad + \sigma^6 \left(c_1^2 + 3c_1 + 1 + \frac{1}{K^2} \right). \end{aligned} \quad (23)$$

Up to now, all formulas have provided exact expressions for the expectations. For the next step, exact expressions for the

expectations are only known when the phases are uniformly distributed, in which case the formulas are [7]

$$\begin{aligned}
c_2 E[tr_L(\mathbf{T})] &= c_2 tr_L(\mathbf{P}) \\
c_2 E[tr_L(\mathbf{T}^2)] &= (1 - N^{-1}) c_2 tr_L(\mathbf{P}^2) + c_2^2 (tr_L(\mathbf{P}))^2 \\
c_2 E[tr_L(\mathbf{T}^3)] &= (1 - 3N^{-1} + 2N^{-2}) c_2 tr_L(\mathbf{P}^3) \\
&+ 3(1 - N^{-1}) c_2^2 tr_L(\mathbf{P}) tr_L(\mathbf{P}^2) + c_2^3 (tr_L(\mathbf{P}))^3 \quad (24) \\
E\left[(tr_L(\mathbf{T}))^2\right] &= tr_L(\mathbf{P})^2 \\
E\left[(tr_L(\mathbf{T}))^3\right] &= tr_L(\mathbf{P})^3 \\
E\left[(tr_L(\mathbf{T})) (tr_L(\mathbf{T}^2))\right] &= (1 - N^{-1}) tr_L(\mathbf{P}) tr_L(\mathbf{P}^2) + c_2 (tr_L(\mathbf{P}))^3.
\end{aligned}$$

If the phase distribution ω is not uniformly distributed, we have the following approximations [7]:

$$\begin{aligned}
c_2 E[tr_L(\mathbf{T})] &= c_2 tr_L(\mathbf{P}) \\
c_2 E[tr_L(\mathbf{T}^2)] &\approx c_2 tr_L(\mathbf{P}^2) + c_2^2 I_2(tr_L(\mathbf{P}))^2 \\
c_2 E[tr_L(\mathbf{T}^3)] &\approx c_2 tr_L(\mathbf{P}^3) + 3c_2^2 I_2 tr_L(\mathbf{P}) tr_L(\mathbf{P}^2) + c_2^3 I_3 (tr_L(\mathbf{P}))^3 \\
E\left[(tr_L(\mathbf{T}))^2\right] &= (tr_L(\mathbf{P}))^2 \\
E\left[(tr_L(\mathbf{T}))^3\right] &= (tr_L(\mathbf{P}))^3 \\
E\left[(tr_L(\mathbf{T})) (tr_L(\mathbf{T}^2))\right] &\approx tr_L(\mathbf{P}) tr_L(\mathbf{P}^2) + c_2 I_2 (tr_L(\mathbf{P}))^3, \quad (25)
\end{aligned}$$

where the approximation is $O(N^{-1})$, and where I_k is defined by (4).

Proposition 2 is proved by combining (23) with (24), while proposition 1 is proved by combining (23) with (25).

Matlab code for implementing the different steps here (like (20), (22), and (24)) can be found in [34].

REFERENCES

- [1] C. S. Burrus and T. W. Parks, *DFT/FFT and Convolution Algorithms*. New York: John Wiley, 1985.
- [2] G. H. Golub and C. F. V. Loan, *Matrix Computations*. John Hopkins University Press, 1983.
- [3] R. Norberg, "On the Vandermonde matrix and its application in mathematical finance," *working paper no. 162, Laboratory of Actuarial Mathematics, Univ. of Copenhagen*, 1999.
- [4] L. Sampaio, M. Kobayashi, Ø. Ryan, and M. Debbah, "Vandermonde matrices for security applications," *work in progress*, 2008.
- [5] J. J. Waterfall, J. Joshua, F. P. Casey, R. N. Gutenkunst, K. S. Brown, C. R. Myers, P. W. Brouwer, V. Elser, and J. P. Sethna, "Sloppy-model universality class and the Vandermonde matrix," *Physical Review Letters*, vol. 97, no. 15, 2006.
- [6] L. Sampaio, M. Kobayashi, Ø. Ryan, and M. Debbah, "Vandermonde frequency division multiplexing," *9th IEEE Workshop on Signal Processing Advances for wireless applications, Recife, Brazil*, 2008.
- [7] Ø. Ryan and M. Debbah, "Random Vandermonde matrices-part I: Fundamental results," *Submitted to IEEE Trans. on Information Theory*, 2008.
- [8] V. Girko, *Theory of Random Determinants*. Kluwer Academic Publishers, 1990.
- [9] V. Marchenko and L. Pastur, "Distribution of Eigenvalues for Some Sets of Random Matrices," *Math USSR Sb.*, vol. 1, pp. 457–483, 1967.
- [10] S. Shamai and S. Verdú, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. Information Theory*, pp. 1302–1326, May 2001.
- [11] E. Telatar, "Capacity of multi-antenna gaussian channels," *Eur. Trans. Telecomm. ETT*, vol. 10, no. 6, pp. 585–596, Nov. 1999.
- [12] M. Debbah, W. Hachem, P. Loubaton, and M. de Courville, "MMSE Analysis of Certain Large Isometric Random Precoded Systems," *IEEE Transactions on Information Theory, Volume: 49 Issue: 5, Page(s): 1293–1311*, May 2003.
- [13] Ø. Ryan and M. Debbah, "Free deconvolution for signal processing applications," *Submitted to IEEE Trans. on Information Theory*, 2007, <http://arxiv.org/abs/cs.IT/0701025>.
- [14] R. R. Muller, "A random matrix model of communication via antenna arrays," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2495–2506, 2002.
- [15] P. Comon, "Independent Component Analysis, a new concept?" *Signal Processing Elsevier, special issue on Higher-Order Statistics*, vol. 36, pp. 287–314, 1994.
- [16] Ø. Ryan and M. Debbah, "Channel capacity estimation using free probability theory," *Submitted to IEEE Trans. Signal Process.*, 2007, <http://arxiv.org/abs/0707.3095>.
- [17] K. Rottmann, *Mathematische Formelsammlung*. B.I. Wissenschaftsverlag, 1991.
- [18] R. Seroul and D. O'Shea, *Programming for Mathematicians*. Springer, 2000.
- [19] A. Nica and R. Speicher, *Lectures on the Combinatorics of Free Probability*. Cambridge University Press, 2006.
- [20] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*. American Mathematical Society, 2000.
- [21] T. Rappaport, *Wireless Communications, Principles and Practice*. New Jersey: Prentice-Hall, 1996.
- [22] L. S. Cardoso, M. Debbah, P. Bianchi, and J. Najim, "Cooperative Spectrum Sensing Using Random Matrix Theory," *International Symposium on Wireless Pervasive Computing*, 2008.
- [23] S. Haykin, "Cognitive Radio: Brain Empowered Wireless Communications," *Journal on Selected Areas in Communications*, vol. 23, pp. 201–220, 2005.
- [24] A. M. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*. www.nowpublishers.com, 2004.
- [25] D. Ganesan, S. Ratnasamy, H. Wang, and D. Estrin, "Coping with Irregular Spatio-temporal Sampling in Sensor Networks," *ACM SIGCOMM*, vol. 34, pp. 125–130, 2004.
- [26] A. Nordio, C.-F. Chiasserini, and E. Viterbo, "Performance of Field Reconstruction Techniques with Noise and Uncertain Sensor Locations," *IEEE Transactions on Signal Processing, submitted*, 2006.
- [27] T. Strohmer, T. Binder, and M. Süßner, "How to Recover Smooth Object Boundaries in Noisy Medical Images," *IEEE ICIP'96, Lausanne*, pp. 331–334, 1996.
- [28] D. S. Early and D. G. Long, "Image Reconstruction and Enhanced Resolution Imaging from Irregular Samples," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, pp. 291–302, 2001.
- [29] M. Rauth and T. Strohmer, "A frequency Domain Approach to the Recovery of Geophysical Potentials," *Proc. Conf. SampTA'97, Aveiro/Portugal*, pp. 109–114, 1997.
- [30] E. J. Candès and J. Romberg, "Sparsity and Incoherence in Compressive Sampling," *Inverse Problems* 23, pp. 969–985, 2007.
- [31] H. G. Feichtinger, K. Gröchenig, and T. Strohmer, "Efficient Numerical Methods in Non-Uniform Sampling Theory," *Numerische Mathematik*, vol. 69, pp. 423–440, 1995.
- [32] D. Guo, S. Shamai, and S. Verdú, "Mutual Information and Minimum Mean-Square Error in Gaussian Channels," *IEEE Trans. Information Theory*, vol. 51, pp. 1261–1283, 2005.
- [33] D. P. Palomar and S. Verdú, "Gradient of Mutual Information in Linear Vector Gaussian Channels," *IEEE Trans. Information Theory*, vol. 52, pp. 141–154, 2006.
- [34] Ø. Ryan, *Tools for convolution with Vandermonde matrices*, 2008, <http://ifi.uio.no/~oyvindry/vandermonde/>.