ANR-07-MDCO-012-01 Project SESAME

conSistent EStimation and lArge random MatricEs

# Deliverable D4.1

# Structured Random Matrices
# Intermediate Report

December 2010

Content:

I. "Convolution operations arising from Vandermonde matrices", Øyvind Ryan and Mérouane Debbah (Supelec)
This first part deals with large Vandermonde matrices with random phases. Various composite matrices composed from Vandermonde matrices are considered. Related Toeplitz and Hankel matrices are also discussed.

II. "Structured Spatio-temporal Sample Covariance Matrix Enhancement with Application to Blind Channel Estimation in Cyclic Prefix systems", Samir-Mohamad Omar, Dirk T.M. Slock (EURECOM)
This second part deals with block Toeplitz and block circulant covariance matrices. An algorithm is proposed for improving estimates of such matrices, enforcing bandedness and low rank structure. Perspectives here include a performance analysis of such algorithms using large system techniques from the first part.

# Chapter I

# Convolution operations arising from Vandermonde matrices


### Abstract

Different types of convolution operations involving large Vandermonde matrices are considered. The convolutions parallel those of large Gaussian matrices and additive and multiplicative free convolution. First additive and multiplicative convolution of Vandermonde matrices and deterministic diagonal matrices are considered. After this, several cases of additive and multiplicative convolution of two independent Vandermonde matrices are considered. It is also shown that the convergence of any combination of Vandermonde matrices is almost sure. We will divide the considered convolutions into two types: those which depend on the phase distribution of the Vandermonde matrices, and those which depend only on the spectra of the matrices. A general criterion is presented to find which type applies for any given convolution. A simulation is presented, verifying the results. Implementations of all considered convolutions are provided and discussed, together with the challenges in making these implementations efficient. The implementation is based on the technique of Fourier-Motzkin elimination, and is quite general as it can be applied to virtually any combination of Vandermonde matrices. Generalizations to related random matrices, such as Toeplitz and Hankel matrices, are also discussed.


## I.1 Introduction

Certain random matrices have in the large dimensional limit a deterministic behavior of the eigenvalue distributions, meaning that one can compute the eigenvalue distributions of $\mathbf{AB}$ and $\mathbf{A}+\mathbf{B}$ based only on the individual eigenvalue distributions of $\mathbf{A}$ and $\mathbf{B}$, when the matrices are independent and large.

The process of computing theses eigenvalues is called **convolution**, or **deconvolution** when one would like to compute the inverse operation. Gaussian-like matrices fit into this setting, and the concept which can be used to find the eigenvalue distribution from that of the component matrices in this case is called freeness [1]. Free probability theory [1], which uses the concept of freeness, is not a new tool but has grown into an entire field of research since the pioneering work of Voiculescu in the 1980's ([2, 3, 4, 5]). However, the basic definitions of free probability are quite abstract and this has hinged a burden on its actual practical use. The original goal was to introduce an analogy to independence in classical probability that can be used for non-commutative random variables like matrices. These more general random variables are elements of what is called a *noncommutative probability space*. The convolution/deconvolution techniques used are various. The classical ones are either analytic (using $R$ and $S$ transforms [6, 1]) or based on moments [**?**, 8, 9, 10]. Recent deconvolution techniques based on statistical eigen-inference methods using large Wishart matrices [**?**], random Matrix theory [**?**] or other deterministic equivalents *à la* Girko [13, 14, 15, 16] were proposed and are possible alternatives. Each one has its advantages and drawbacks. Unfortunately, although successfully applied [17, 18], all these techniques can only treat very simple models i.e. the case where one of the considered matrices is unitarily invariant. This invariance has a special meaning in wireless networks and supposes that there is some kind of symmetry in the problem to be analyzed. The moments technique, which will be the focus of this work, is very appealing and powerful in order to derive the exact asymptotic moments of "non-free matrices", for which we still do not have a general framework. It requires combinatorial skills and can be used for a large class of random matrices. The main drawback of the technique (compared to other tools such as the Stieltjes transform method [19]) is that it can rarely provide the exact eigenvalue distribution. However, in many applications, one needs only a subset of the moments depending on the number of parameters to be estimated.

Recently [20], Vandermonde matrices (which do not fall within the free probability framework) were shown to be a case of high interest in wireless communications. Such matrices have various applications in signal reconstruction [21], cognitive radio [22], physical layer security [23], and MIMO channel modeling [24]. A Vandermonde matrix with entries on the unit circle is on the form

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

$\mathbf{V}$ will in this paper always denote a Vandermonde matrix, and its dimension will be denoted $N \times L$. The $\omega_1$,...,$\omega_L$, also called phase distributions, will be assumed i.i.d., taking values in $[0, 2\pi)$. We will also assume, as in many applications, that $N$ and $L$ go to infinity at the same rate, and write $c = \lim_{N\to\infty} \frac{L}{N}$ for the aspect ratio. If necessary, we will write $\mathbf{V}_\omega$ to emphasize the actual phase

distribution, or $\mathbf{V}_{\omega,c}$ to also emphasize the aspect ratio. In [20], the limit eigenvalue distributions of combinations of $\mathbf{V}^H\mathbf{V}$ and diagonal matrices $\mathbf{D}(N)$ were shown to be dependent on only the limit eigenvalue distributions of the two matrices. Important combinations are the multiplicative and additive models,

$$\mathbf{D}(N)\mathbf{V}^H\mathbf{V} \text{ and } \mathbf{D}(N) + \mathbf{V}^H\mathbf{V}. \tag{I.2}$$

In the large $N$-limit, (I.2) thus gives rise to two convolution operations,

**1)** $\lim_{N\to\infty} \mathbf{D}(N)\mathbf{V}^H\mathbf{V}$ and $\lim_{N\to\infty}(\mathbf{D}(N) + \mathbf{V}^H\mathbf{V})$,

which thus depend only on the input spectra. Here lim is used to denote the limit of the eigenvalue distribution of the considered matrix, in an appropriate metric. However, it is not clear from [20] how 1) can be computed algorithmically, as only sketches for this were provided. We also have the operations

**2)** $\lim_{N\to\infty} \mathbf{D}(N)\mathbf{V}\mathbf{V}^H$ and $\lim_{N\to\infty}(\mathbf{D}(N) + \mathbf{V}\mathbf{V}^H)$,

for which it is unknown whether the result only depends on the spectra. This case happens in practical scenarios (for cognitive applications [22] as well as secure transmissions [23]) when a Vandermonde precoder $\mathbf{V}$ is used in a given Toeplitz channel matrix $\mathbf{D}(N)$ independent from $\mathbf{V}$. One can then compute cognitive and secrecy rates. When we replace with independent Vandermonde matrices $\mathbf{V}_1$ and $\mathbf{V}_2$ which may or may not have the same phase distributions, it is also unknown if the convolution operations

**3)** $\lim_{N\to\infty} \mathbf{V}_1^H\mathbf{V}_1\mathbf{V}_2^H\mathbf{V}_2$ and $\lim_{N\to\infty}(\mathbf{V}_1^H\mathbf{V}_1 + \mathbf{V}_2^H\mathbf{V}_2)$,

**4)** $\lim_{N\to\infty} \mathbf{V}_1\mathbf{V}_1^H\mathbf{V}_2\mathbf{V}_2^H$ and $\lim_{N\to\infty}(\mathbf{V}_1\mathbf{V}_1^H + \mathbf{V}_2\mathbf{V}_2^H)$,

only depend on the spectra of $\mathbf{V}_1$ and $\mathbf{V}_2$. These cases are important for the recovery of the distribution of sensors (which are deployed in a clustered manner with different mean positions) and in the case of MIMO multi-fold scattering [25].

Expressions such as 4), when different types of matrices are multiplied, will in the following be called *mixed moments*.

In this contribution we explain which of the above operations depend only on the spectra of the matrices, state expressions for those convolutions (in fact, we also state expressions for the cases where the result can not be written in terms of the spectra), explain how these expressions have been obtained algorithmically, and explain an accompanying software implementation [26, 27] of the corresponding algorithms. We also attempt to complete the analysis started in [20], by stating a very general criterion for when the mixed moments of (many) Vandermonde matrices and deterministic matrices depend only on the input spectra:

*If there are no terms on the form* $\mathbf{V}_1^H\mathbf{V}_2$ *in a mixed moment, with* $\mathbf{V}_1$ *and* $\mathbf{V}_2$ *independent and with different phase distributions, the mixed moment will depend only on the spectra of the input matrices. In all other cases, we*

*can't expect dependence on just the spectra of the input matrices, and the mixed moment can depend on the entire phase distributions of the input matrices.*

The software implementation can in fact be extended to handle all cases which meet this criterion, as well as cases where knowledge of the phase distribution also is required. In this way it is an indispensable tool, as it automates the very tedious computations inherent in the presented formulas, for which no simple expressions are known.

Concluding from the criterion, 1) will depend only on the spectra (as shown in [20]), as does 3). 4) may not depend on only the spectra when the two phase distributions are different. Despite this, 4) is interesting in its own right, since it has a geometric interpretation in terms of phase distributions, and is therefore handled separately. For case 2), we state more generally that when the pattern $\mathbf{D}(N)\mathbf{V}$ appears in a mixed moment, we can not expect dependence only on the spectrum.

It turns out that other types of random matrices can use the same methods as for Vandermonde matrices to compute their moments, such as Toeplitz matrices and Hankel matrices. We will explain how the software implementation has been extended to handled these matrices as well.

The paper is organized as follows. Section I.2 provides background essentials on random matrix theory needed for the main results, which are stated in I.3. The results include the precise statement of the criterion above for when we only have dependence on the spectra of the matrices, results on the convolution operations 1)-4), and extensions to related random matrices such as Toeplitz and Hankel matrices. A generalization of our results to almost sure convergence of matrices is also made. All presented formulas are obtained from the implementation, and the major pieces in this implementation are gone through in Section I.4, such as partition iteration, and Fourier-Motzkin elimination [28]. Section I.5 presents a simulation which verifies the results.

## I.2 Random matrix Background Essentials

In the following, upper (lower boldface) symbols will be used for matrices (column vectors), whereas lower symbols will represent scalar values, $(.)^T$ will denote the transpose operator, $(.)^\star$ conjugation, and $(.)^H = \left((.)^T\right)^\star$ hermitian transpose. $\mathbf{I}_L$ will represent the $L \times L$ identity matrix. We let Tr be the (non-normalized) trace for square matrices, defined by,

$$\mathrm{Tr}(\mathbf{A}) = \sum_{i=1}^{L} a_{ii},$$

where $a_{ii}$ are the diagonal elements of the $L \times L$ matrix $\mathbf{A}$. We also let tr be the normalized trace, defined by $\mathrm{tr}(\mathbf{A}) = \frac{1}{L}\mathrm{Tr}(\mathbf{A})$.

In the following we will implicitly assume that $L$ and $N$ go to infinity in such a way that $\frac{L}{N} \to c$. $\mathbf{D}_r(N), 1 \leq r \leq n$ will denote non-random diagonal $L \times L$ matrices. We will have use for the following definition:

**Definition 1.** *We will say that the* $\{\mathbf{D}_r(N)\}_{1\leq r\leq n}$ *have a joint limit distribution as* $N\to\infty$ *if the limit*

$$D_{i_1,\ldots,i_s}=\lim_{N\to\infty}\operatorname{tr}\left(\mathbf{D}_{i_1}(N)\cdots\mathbf{D}_{i_s}(N)\right) \tag{I.3}$$

*exists for all choices of* $i_1,\ldots,i_s\in\{1,..,n\}$.

A joint limit distribution for the $\mathbf{D}_r(N)$ will always be assumed in the following. The corresponding concept for random matrices is the following:

**Definition 2.** *Let* $\{\mathbf{A}_n\}_{n=1}^{\infty}$ *be an ensemble of (square) random matrices. We say that* $\{\mathbf{A}_n\}_{n=1}^{\infty}$ *converge in distribution if the limit*

$$\lim_{n\to\infty}\mathbb{E}[\operatorname{tr}((\mathbf{A}_n)^r)] \tag{I.4}$$

*exists for all* $r$. *We will say that ensembles* $\{\mathbf{A}_{1n},\mathbf{A}_{2n},\ldots\}_{n=1}^{\infty}$ *of random matrices converge in distribution if the limit*

$$\lim_{n\to\infty}\mathbb{E}[\operatorname{tr}(\mathbf{A}_{i_1n}\mathbf{A}_{i_2n}\cdots\mathbf{A}_{i_sn})] \tag{I.5}$$

*exists whenever the matrix product* $\mathbf{A}_{i_1n}\mathbf{A}_{i_2n}\cdots\mathbf{A}_{i_sn}$ *is well-defined, and square.*

When we refer to moments, we will generally mean (I.4), while mixed moments refer to (I.5). A stronger form of convergence, which we will generalize our results to, is *almost sure convergence in distribution*. This type of convergence requires that (I.4), (I.5) are replaced with

$$\begin{aligned}&\operatorname{tr}\left((\mathbf{A}_n)^r\right)\overset{\text{a.s.}}{\to}C_r\\&\operatorname{tr}(\mathbf{A}_{i_1n}\mathbf{A}_{i_2n}\cdots\mathbf{A}_{i_sn})\overset{\text{a.s.}}{\to}C_{i_1,\ldots,i_s},\end{aligned}$$

where $C_r, C_{i_1,\ldots,i_s}$ are constants.

We will also need some basic concepts from partition theory. $\mathcal{P}(n)$ will denote the partitions of $\{1,\ldots,n\}$. For a partition $\rho=\{W_1,\ldots,W_r\}\in\mathcal{P}(n)$, $W_1,\ldots,W_r$ denote its blocks, while $|\rho|=r$ denotes the number of blocks, $\|\rho\|=n$ the number of elements in the partition. We will write $k\sim_\rho l$ when $k$ and $l$ belong to the same block of $\rho$. We will also write $b(i)$ for the index of the block in $\rho$ $i$ belongs to. Partition notation is adapted to the mixed moment (I.3) in the following way:

**Definition 3.** *For* $\rho=\{W_1,\ldots,W_k\}$, *with* $W_i=\{w_{i1},\ldots,w_{i|W_i|}\}$, *we define*

$$D_{W_i} = D_{i_{w_{i1}},\ldots,i_{w_{i|W_i|}}} \tag{I.6}$$

$$D_\rho = \prod_{i=1}^{k}D_{W_i}. \tag{I.7}$$

The set of partitions is a partially ordered set under the refinement order, i.e. $\rho_1 \leq \rho_2$ whenever any block of $\rho_1$ is contained within a block of $\rho_2$. By $\rho_1 \vee \rho_2$ we will mean the smallest partition (w.r.t. the refinement order) which is larger than both $\rho_1$ and $\rho_2$. $\vee$ will in our results be used in conjunction with the partition $[0,1]_n \in \mathcal{P}(2n)$, defined by

$$[0,1]_n = \{\{1,2\},\{3,4\},...,\{2n-1,2n\}\}.$$

$[0,1]_n$ is an example of what is called an interval partition, meaning that each block consists solely of successive numbers. We will also write $[\cdot,\cdot]$ for the intervals in an interval partition, so that we could also have written

$$[0,1]_n = \{[1,2],[3,4],...,[2n-1,2n]\}.$$

We will in the following consider the trace of a general mixed moment of Vandermonde matrices and deterministic matrices, the only requirement being that matrices and their adjoints appear in alternating order so that the resulting matrix is square:

$$\operatorname{tr}\left(\mathbf{D}_1(N)\mathbf{V}_{i_1}^H\mathbf{V}_{i_2}\cdots\mathbf{D}_n(N)\mathbf{V}_{i_{2n-1}}^H\mathbf{V}_{i_{2n}}\right), \tag{I.8}$$

where $\mathbf{V}_1, \mathbf{V}_2, ...$ are assumed independent and with phase distributions $\omega_1, \omega_2, ....$ In particular, we assume that $N_{i_{2k}} = N_{i_{2k-1}}$ when the $\mathbf{V}_i$ are $N_i \times L$, in order for the dimensions of the matrices in (I.8) to match. It turns out we can obtain the asymptotic behavior of (I.8) for arbitrary continuous phase distributions $\omega_i$. For (I.8) we will let $\sigma$ be the partition in $\mathcal{P}(2n)$ defined by equality of the phase distributions, i.e. $j \sim_\sigma k$ if and only if $\omega_{i_j} = \omega_{i_k}$ ($i_j$ and $i_k$ may or may not be different for this). Similarly we will let $\sigma_1$ be the partition in $\mathcal{P}(2n)$ defined by dependence of the Vandermonde matrices, i.e. $j \sim_{\sigma_1} k$ if and only if $i_j = i_k$. Obviously, $\sigma_1 \leq \sigma$.

## I.3 Statement of main results

The main result of the paper addresses moments on the form (I.8), and goes as follows.

**Theorem 1.** *Let $\mathbf{V}_i$ be independent $N_i \times L$ Vandermonde matrices with aspect ratios $c_i = \lim_{N_i\to\infty} \frac{L}{N_i}$ and phase distributions $\omega_i$ with continuous densities on $[0, 2\pi)$. The mixed moment*

$$\lim_{N\to\infty} \operatorname{tr}\left(\mathbf{D}_1(N)\mathbf{V}_{i_1}^H\mathbf{V}_{i_2}\cdots\mathbf{D}_n(N)\mathbf{V}_{i_{2n-1}}^H\mathbf{V}_{i_{2n}}\right). \tag{I.9}$$

*always exists when $\mathbf{D}_i(N)$ have a joint limit distribution. When $\sigma \geq [0,1]_n$ (i.e. there are no terms on the form $\mathbf{V}_i^H\mathbf{V}_j$, with $\mathbf{V}_i$ and $\mathbf{V}_j$ independent and with different phase distributions), (I.9) depends only on the moments*

$$\begin{aligned} V_n^{(i)} &= \lim_{N\to\infty} E\left[\operatorname{tr}\left(\left(\mathbf{V}_i^H\mathbf{V}_i\right)^n\right)\right] \\ D_{i_1,...,i_s} &= \lim_{N\to\infty} \operatorname{tr}\left(\mathbf{D}_{i_1}(N)\cdots\mathbf{D}_{i_s}(N)\right), \end{aligned}$$

*the aspect ratios $c_i$, and $\sigma$, and assumes the form*

$$\sum_{s,r,i_t,j_t,k_t} a_{i_1,\ldots,i_s,j_1,\ldots,j_r,k_1,\ldots,k_r} D_{i_1,\ldots,i_s} \prod_{t=1}^{r} V_{j_t}^{(k_t)}, \tag{I.10}$$

*where the $a_{i_1,\ldots,i_s,j_1,\ldots,j_r,k_1,\ldots,k_r}$ are rational numbers.*

Theorem 1 is proved in Appendix A, and states exactly when we can hope for performing deconvolution, either by inferring on the spectrum of $\mathbf{D}_i(N)$, or on the spectrum or the phase distribution of $\mathbf{V}_i$ from (I.9). The proof will also state concrete expressions for the mixed moments which parallel the expressions of [20], and also summarize the algorithm needed to compute these expressions, as performed by the implementation. The implementation is thus *moment-based*, in that it computes the moments as defined in (I.4), from the moments of the input matrices. We do not know any other methods than that of moments to infer on the spectra of such matrices, since other analytical tools have not been developed yet.

As an example, Theorem 1 states that

$$\operatorname{tr}\left(\left((\mathbf{V}_1 + \mathbf{V}_2 + \cdots)^H (\mathbf{V}_1 + \mathbf{V}_2 + \cdots)\right)^p\right), \tag{I.11}$$

which characterize the singular law of a sum of independent Vandermonde matrices, depend only on the moments when the $\mathbf{V}_i$ are independent with the same phase distribution. When the phase distributions are different, however, the same can not be said. The final observation in Theorem 1 about the polynomial form of the mixed moment is also important, since it is a property shared with freeness. Although (I.10) is seen not to be multi-linear in the moments in general, several of the particular convolutions we consider will be seen to have such a multi-linearity property.

In the following, we state expressions for the convolutions 1)-4) on the form (I.10). Their proofs will be apparent from the proof of Theorem 1, and can be found in Appendix B. The aspect ratio $c$ will be handled in a particular way in these results, so that it is applied outside the algorithm itself. The results are stated so that it is possible to turn them around for "deconvolution": for instance, from the moments of $\mathbf{D}(N)\mathbf{V}^H\mathbf{V}$, one can infer on the moments of $\mathbf{D}(N)$. The application of the theorems in terms of deconvolution is certainly as important as the limit results themselves, since it enables us to infer on the parameters in an underlying model (here represented by $\mathbf{D}(N)$ and $\mathbf{V}$). The accompanying implementation of this paper also supports deconvolution.

As for the convolutions 2), this form is not compatible with the form (I.9) due to the placement of the $\mathbf{D}(N)$. We will therefore not handle this operation, only state in Appendix B why one in this case can't expect that the result only depends on the spectra of $\mathbf{D}(N)$ and $\mathbf{V}$.

All formulas in the following are generated by the accompanying software implementation, which is gone through in Section I.4. Implementation details pertaining to the different convolutions are gone through in Appendix B. Note

that the software implementation is capable not only of generating the listed mathematical formulas for the convolutions, but also to perform the computations numerically, as would be needed in real-time applications.

### I.3.1 The convolutions $\lim_{N\to\infty} \mathbf{D}(N)\mathbf{V}^H\mathbf{V}$ and $\lim_{N\to\infty}(\mathbf{D}(N)+\mathbf{V}^H\mathbf{V})$

In Theorem 1 of [20], the moments $\lim_{N\to\infty} \operatorname{tr}\left(\left(\mathbf{D}(N)\mathbf{V}^H\mathbf{V}\right)^n\right)$ were expressed in terms of the integrals

$$I_{k,\omega} = (2\pi)^{k-1}\int_0^{2\pi} p_\omega(x)^k, \tag{I.12}$$

$p_\omega$ being the density of the phase distribution. These again determine the moments of $\mathbf{V}^H\mathbf{V}$ uniquely ((13) and (20) in [20]), so that, indeed, the moments of the matrices (I.2) depend only on the spectra of the input matrices. This gives the following result for the multiplicative convolution in 1):

**Theorem 2.** *Assume that* $\mathbf{V}$ *has a phase distribution with continuous density,*

$$V_n = \lim_{N\to\infty} \operatorname{tr}\left(\left(\mathbf{V}^H\mathbf{V}\right)^n\right) \tag{I.13}$$

$$D_n = c\lim_{N\to\infty} \operatorname{tr}\left(\mathbf{D}(N)^n\right) \tag{I.14}$$

$$M_n = c\lim_{N\to\infty} \operatorname{tr}\left(\left(\mathbf{D}(N)\mathbf{V}^H\mathbf{V}\right)^n\right), \tag{I.15}$$

*where* $c = \lim_{N\to\infty}\frac{L}{N}$. *Then we have that*

$$\begin{aligned}
M_1 &= D_1\\
M_2 &= D_2 - D_1^2 + D_1^2V_2\\
M_3 &= D_3 - 3D_2D_1 + 3D_2D_1V_2\\
&\quad +2D_1^3 - 3D_1^3V_2 + D_1^3V_3\\
M_4 &= D_4 - \frac{8}{3}D_2^2 + \frac{8}{3}D_2^2V_2 - 4D_3D_1\\
&\quad +4D_3D_1V_2 + 12D_2D_1^2 - 18D_2D_1^2V_2\\
&\quad +6D_2D_1^2V_3 - \frac{19}{3}D_1^4 + \frac{34}{3}D_1^4V_2\\
&\quad -6D_1^4V_3 + D_1^4V_4
\end{aligned}$$

*where all coefficients are rational numbers. Also, whenever* $\{M_n\}_{1\le n\le k}$ *are known, and* $\{V_n\}_{1\le n\le k}$*(or* $\{D_n\}_{1\le n\le k}$*) also are known, then* $\{D_n\}_{1\le n\le k}$*(or* $\{V_n\}_{1\le n\le k}$*) are uniquely determined.*

The proof of Theorem 2 can be found in Appendix B. Restricting to uniform phase distribution we get the following result, also generated by the implementation.

**Corollary 1.** *When* $\mathbf{V}$ *has uniform phase distribution, we have that*

$$\begin{aligned} M_1 &= D_1 \\ M_2 &= D_2 + D_1^2 \\ M_3 &= D_3 + 3D_2D_1 + D_1^3 \\ M_4 &= D_4 + \frac{8}{3}D_2^2 + 4D_3D_1 + 6D_2D_1^2 + D_1^4 \end{aligned}$$

The additive convolution in 1) can be split into sums of many terms similar to (I.15), and for each term, the results of [20] can be applied. We obtain the following result, also proved in Appendix B:

**Theorem 3.** *Assume that has a phase distribution with continuous density,*

$$M_n = c \lim_{N\to\infty} \operatorname{tr}\left(\left(\mathbf{D}(N) + \mathbf{V}^H\mathbf{V}\right)^n\right),$$

*where* $c = \lim_{N\to\infty} \frac{L}{N}$*. With* $V_n$ *as in (I.13) and* $D_n$ *as in (I.14), we have that*

$$\begin{aligned} M_1 &= D_1 + 1 \\ M_2 &= D_2 + 2D_1 + V_2 \\ M_3 &= D_3 + 3D_2 + 3D_1V_2 + V_3 \\ M_4 &= D_4 + 4D_3 + 2D_2 + 4D_2V_2 \\ &\quad -2D_1^2 + 2D_1^2V_2 + 4D_1V_3 + V_4 \end{aligned}$$

*where all coefficients are rational numbers. Also, whenever* $\{M_n\}_{1\le n\le k}$ *are known, and* $\{V_n\}_{1\le n\le k}$*(or* $\{D_n\}_{1\le n\le k}$*) also are known, then* $\{D_n\}_{1\le n\le k}$ *(or* $\{V_n\}_{1\le n\le k}$*) are uniquely determined.*

Restricting to uniform phase distribution we get another specialized result:

**Corollary 2.** *When* $\mathbf{V}$ *has uniform phase distribution, we have that*

$$\begin{aligned} M_1 &= D_1 + 1 \\ M_2 &= D_2 + 2D_1 + 2 \\ M_3 &= D_3 + 3D_2 + 6D_1 + 5 \\ M_4 &= D_4 + 4D_3 + 10D_2 + 2D_1^2 + 20D_1 + \frac{44}{3} \end{aligned}$$

### I.3.2 The convolutions $\lim_{N\to\infty} \mathbf{V}_1^H\mathbf{V}_1\mathbf{V}_2^H\mathbf{V}_2$ and $\lim_{N\to\infty}(\mathbf{V}_1^H\mathbf{V}_1 + \mathbf{V}_2^H\mathbf{V}_2)$

The following result says that the convolution 3) only depends on the spectra of the input matrices:

**Theorem 4.** *Assume that* $\mathbf{V}_1$ *and* $\mathbf{V}_2$ *are independent Vandermonde matrices where the phase distributions have continuous densities, and set*

$$\begin{aligned} V_1^{(n)} &= \lim_{N\to\infty} \operatorname{tr}\left(\left(\mathbf{V}_1^H\mathbf{V}_1\right)^n\right) \\ V_2^{(n)} &= \lim_{N\to\infty} \operatorname{tr}\left(\left(\mathbf{V}_2^H\mathbf{V}_2\right)^n\right) \\ M_n &= \lim_{N\to\infty} \operatorname{tr}\left((\mathbf{V}_1^H\mathbf{V}_1\mathbf{V}_2^H\mathbf{V}_2)^n\right) \qquad \text{(I.16)} \\ N_n &= \lim_{N\to\infty} \operatorname{tr}\left((\mathbf{V}_1^H\mathbf{V}_1+\mathbf{V}_2^H\mathbf{V}_2)^n\right) \qquad \text{(I.17)} \end{aligned}$$

*$M_n, N_n$ are completely determined by $V_2^{(i)}, V_3^{(i)}, ...$, and the aspect ratios $c_1 = \lim_{N_1\to\infty}\frac{L}{N_1}, c_2 = \lim_{N_2\to\infty}\frac{L}{N_2}$. Moreover, $M_n, N_n$ are higher degree polynomials in the $V_2^{(i)}, V_3^{(i)}, ...$ on the form (I.10). Also, whenever $\{M_n\}_{1\le n\le k}$ (or $\{N_n\}_{1\le n\le k}$) are known, and $\{V_1^{(n)}\}_{1\le n\le k}$ also are known, then $\{V_2^{(n)}\}_{1\le n\le k}$ are uniquely determined.*

The proof can be found in Appendix B. Due to the complexity in the expressions , we do not state formulas for the first moments in Theorem 4.

Interestingly, since the joint distribution of $\{\mathbf{V}^H\mathbf{V}, \mathbf{D}(N)\}$ is not multi-linear in the moments of $\mathbf{D}(N)$, while the joint distribution of $\{\mathbf{V}_1^H\mathbf{V}_1, \mathbf{V}_2^H\mathbf{V}_2\}$ is, it is seen that the joint distributions are different in the two cases, even if the moments of the component matrices are the same.

### I.3.3 The convolution $\lim_{N\to\infty}\mathbf{V}_1\mathbf{V}_1^H\mathbf{V}_2\mathbf{V}_2^H$ when the matrices have equal phase distribution

When the phase distributions are different, Theorem 1 explains that the moments of $\mathbf{V}_1\mathbf{V}_1^H\mathbf{V}_2\mathbf{V}_2^H$ are not necessarily expressible in terms of the moments of the component matrices. This is, however, the case when the phase distributions are equal. We thus have the following result, which proof can be found in Appendix B:

**Theorem 5.** *Assume that* $\mathbf{V}_1$ *and* $\mathbf{V}_2$ *are independent Vandermonde matrices with the same phase distribution, and that this has a continuous density, and set*

$$\begin{aligned} V_n &= \lim_{N\to\infty} \operatorname{tr}\left(\left(\mathbf{V}_i^H\mathbf{V}_i\right)^n\right) \\ M_n &= \lim_{N\to\infty} \operatorname{tr}\left((\mathbf{V}_1^H\mathbf{V}_2\mathbf{V}_2^H\mathbf{V}_1)^i\right). \end{aligned}$$

*Then we have that*

$$\begin{aligned} M_1 &= -1+V_2 \\ M_2 &= -3+6V_2-4V_3+V_4 \\ M_3 &= -58+123V_2-96V_3+39V_4-9V_5+V_6 \\ M_4 &= -\frac{21532}{5}+\frac{410726}{45}V_2-\frac{321191}{45}V_3+\frac{44516}{15}V_4 \\ &\quad -772V_5+136V_6-16V_7+V_8 \end{aligned}$$

Restricting to uniform phase distribution we get another specialized result:

**Corollary 3.** *When* $\mathbf{V}_1$ *and* $\mathbf{V}_2$ *have uniform phase distribution, we have that*

$$\begin{aligned} M_1 &= 1 \\ M_2 &= 2 \\ M_3 &= 5 \\ M_4 &= \frac{44}{3} \end{aligned}$$

### I.3.4 The convolution $\lim_{N\to\infty}\left(\mathbf{V}_{\omega_1}^{(1)}\left(\mathbf{V}_{\omega_1}^{(1)}\right)^H + \mathbf{V}_{\omega_2}^{(2)}\left(\mathbf{V}_{\omega_2}^{(2)}\right)^H\right)$

$\mathbf{V}^H\mathbf{V}$ can be viewed as the sample covariance matrix of the random vector $(1, e^{-j\omega}, ..., e^{-j(N-1)\omega})$. A similar interpretation of the convolution $\left(\mathbf{V}_{\omega_1}^{(1)}\left(\mathbf{V}_{\omega_1}^{(1)}\right)^H + \mathbf{V}_{\omega_2}^{(2)}\left(\mathbf{V}_{\omega_2}^{(2)}\right)^H\right)$ is thus as a sample covariance matrix of a random vector of the same type, but where the phase distribution is $\omega_1$ parts of the time, and $\omega_2$ the rest of the time. This convolution does not satisfy the requirement $\sigma \geq [0,1]_n$ from Theorem 1, so there is no guarantee that the result only depends on the spectra of the input matrices. It will be apparent from Theorem 6 below that the dependence is, indeed, on more than just these spectra: Knowledge about the phase distributions is also required, and we will in fact interpret this convolution instead as an operation on phase distributions.

Consider first two independent Vandermonde matrices $\mathbf{V}_{\omega,c_1}^{(1)}$, $\mathbf{V}_{\omega,c_2}^{(2)}$ with an equal number of rows $N$ and with a common phase distribution $\omega$. By stacking $\mathbf{V}_{\omega,c_1}^{(1)}$, $\mathbf{V}_{\omega,c_2}^{(2)}$ horizontally into one larger matrix, it is straightforward to show that the distribution of

$$\mathbf{V}_{\omega,c_1}^{(1)}\left(\mathbf{V}_{\omega,c_1}^{(1)}\right)^H + \mathbf{V}_{\omega,c_2}^{(2)}\left(\mathbf{V}_{\omega,c_2}^{(2)}\right)^H \tag{I.18}$$

equals that of $\mathbf{V}_{\omega,c_1+c_2}\mathbf{V}_{\omega,c_1+c_2}^H$. This case when the phase distributions are equal is therefore trivial.

When $\mathbf{V}_{\omega_1,c}^{(1)}$, $\mathbf{V}_{\omega_2,c}^{(2)}$ are independent with the same number of rows, but with different phase distributions, computing the distribution of

$$\mathbf{V}_{\omega_1,c_1}^{(1)}\left(\mathbf{V}_{\omega_1,c_1}^{(1)}\right)^H + \mathbf{V}_{\omega_2,c_2}^{(2)}\left(\mathbf{V}_{\omega_2,c_2}^{(2)}\right)^H \tag{I.19}$$

seems, however, to be more complex. The following result explains that, at least in the limit, the situation is simpler. There the sum can be replaced by another Vandermonde matrix, whose phase distribution can be constructed in a particular way from the original ones:

**Theorem 6.** *Let* $\mathbf{V}_{\omega_1,c_1}$ *and* $\mathbf{V}_{\omega_2,c_2}$ *be independent* $N \times L_1$, $N \times L_2$ *random Vandermonde matrices with phase distributions* $\omega_1$, $\omega_2$, *respectively, and with*

*aspect ratios* $c_1 = \lim_{N\to\infty} \frac{L_1}{N}$, $c_2 = \lim_{N\to\infty} \frac{L_2}{N}$, *respectively. Then the limit distribution of*

$$\mathbf{V}_{\omega_1,c_1}\mathbf{V}_{\omega_1,c_1}^H + \mathbf{V}_{\omega_2,c_2}\mathbf{V}_{\omega_2,c_2}^H \tag{I.20}$$

*equals that of*

$$\mathbf{V}_{\omega_1 *_{c_1,c_2} \omega_2, c_1+c_2}\mathbf{V}_{\omega_1*\omega_2,c_1+c_2}^H, \tag{I.21}$$

*where* $\omega_1 *_{c_1,c_2} \omega_2$ *denotes the phase distribution with density* $\frac{1}{c_1+c_2}(c_1 p_{\omega_1} + c_2 p_{\omega_2})$, *where* $p_{\omega_1}, p_{\omega_2}$ *are the densities of the phase distributions* $\omega_1, \omega_2$.

The proof of Theorem 6 can be found in Appendix C. The result is only asymptotic, meaning that the mean eigenvalue distribution for finite $N$ of the two mentioned matrices are in fact different. This can be seen by setting $L = N = 2$, and observing that the distribution of $\frac{1}{2}\left(e^{j\omega_1} + e^{j\omega_2}\right)$ is in general different from that of $e^{\omega_1 *_{1,1} \omega_2}$. No trivial proof for Theorem 6 is thus known, since the strategy of stacking the Vandermonde matrices (from the reasoning for (I.18)) will not work.

Theorem 6 says that one depends on knowledge about the phase distributions for Convolution 4). To verify this, set $\omega_1$ and $\omega_2$ equal to the uniform distributions on $[0, \pi)$, and then change $\omega_2$ to the uniform distribution on $[\pi, 2\pi)$. The phase distributions here give the same moments (since they are shifted versions). However, the two versions of $\frac{1}{2}(p_{\omega_1} + p_{\omega_2})$ give phase distributions with different moments, since we get the uniform distribution on $[0, \pi)$ in the first case, and the uniform distribution on $[0, 2\pi)$ in the second case: the moments of these are different, since the uniform distribution on $[0, 2\pi)$ minimizes the moments of Vandermonde matrices [20]. For the same reason, Theorem 6 says that the moments of (I.20) are minimized when $\omega_1 *_{c_1,c_2} \omega_2$ equals the uniform distribution.

### I.3.5 Hankel and Toeplitz matrices

[20] states that the moments of $\mathbf{V}^H\mathbf{V}$ can be expressed in terms of volumes of certain convex polytopes. It turns out that the moments of Hankel, Markov and Toeplitz matrices can be expressed in terms of a subset of these polytopes [29], so that we can use the same strategy to compute the moments of these matrices also. The proof of the following theorem relating to the moments of Toeplitz matrices is therefore explained in Appendix B.

**Theorem 7.** *Define the Toeplitz matrix*

$$\mathbf{T}_n = \frac{1}{\sqrt{n}} \begin{pmatrix} X_0 & X_1 & X_2 & \cdots & X_{n-2} & X_{n-1} \\ X_1 & X_0 & X_1 & & & X_{n-2} \\ X_2 & X_1 & X_0 & & \ddots & \vdots \\ \vdots & & & \ddots & & X_2 \\ X_{n-2} & & & & X_0 & X_1 \\ X_{n-1} & X_{n-2} & \dots & X_2 & X_1 & X_0 \end{pmatrix},$$

*where $X_i$ are i.i.d., real-valued random variables with variance 1. Let $M_i$ be the $2i$'th asymptotic moment of $\mathbf{T}_n$ (the odd moments vanish). These moments are given by*

$$\begin{aligned} M_1 &= 1 \\ M_2 &= \frac{8}{3} \\ M_3 &= 11 \\ M_4 &= \frac{1435}{24} \end{aligned}$$

A similar result for Hankel matrices also holds:

**Theorem 8.** *Define the Hankel matrix*

$$\mathbf{H}_n = \frac{1}{\sqrt{n}} \begin{pmatrix} X_1 & X_2 & \cdots & \cdots & X_{n-1} & X_n \\ X_2 & X_3 & & & X_n & X_{n+1} \\ \vdots & & & & X_{n+1} & X_{n+2} \\ & & & \ddots & & \\ X_{n-2} & X_{n-1} & & & & \vdots \\ X_{n-1} & X_n & & & X_{2n-3} & X_{2n-2} \\ X_n & X_{n+1} & \cdots & \cdots & X_{2n-2} & X_{2n-1}. \end{pmatrix},$$

*where $X_i$ are i.i.d., real-valued random variables with variance 1. Let $M_i$ be the $2i$'th asymptotic moment of $\mathbf{H}_n$ (the odd moments vanish). These moments are given by*

$$\begin{aligned} M_1 &= 1 \\ M_2 &= \frac{8}{3} \\ M_3 &= 14 \\ M_4 &= 100 \end{aligned}$$

Similar results can also be written down for Markov matrices, but these expressions are skipped. It seems that expressions for the joint distribution of Hankel and Toeplitz matrices and matrices $\mathbf{D}(N)$ on the same form as before do not exist, meaning that the mixed moments may not exist, or that they depend on more than the spectra of the component matrices. The details of this are also skipped.

### I.3.6 Generalizations to almost sure convergence

Up to now, we have only shown convergence in distribution for the different convolutions and mixed moments. The same results also hold when we replace convergence in distribution with almost sure convergence in distribution. We summarize this in the following result:

**Theorem 9.** *Assume that the matrices $\mathbf{D}_i(N)$ have a joint limit distribution as $N \to \infty$, and that $\mathbf{V}_1, \mathbf{V}_2, ...$ are independent, with continuous phase distributions. Any combination of matrices on the form (I.8) converges almost surely in distribution, whenever the matrix product is well-defined and square.*

The proof of Theorem 9 can be found in Appendix D. In particular, the matrices we have considered in our convolution operations, such as $\mathbf{V}_1^H \mathbf{V}_1 \mathbf{V}_2^H \mathbf{V}_2$, $\mathbf{V}_1^H \mathbf{V}_1 + \mathbf{V}_2^H \mathbf{V}_2$, all converge almost surely in distribution.

### I.3.7 Generalized Vandermonde matrices

We have not considered generalized Vandermonde matrices up to now, i.e. matrices were the columns in $\mathbf{V}$ are not uniform distributions of powers [30, 20]. Although similar results can also be stated for these matrices, we only explain how they will differ.

In case of uniform power distribution, the column sum of (I.1) is

$$\frac{1 - e^{jNx}}{1 - e^{jx}}, \tag{I.22}$$

and this is substituted into the integrand of the expression defining the Vandermonde mixed moment expansion coefficients (see Appendix A). For generalized Vandermonde matrices, one can also define these coefficients [20], the difference being that one replaces the sum of the powers (I.22) with a different function, and requires that the function has the property proved in Lemma 2 in Appendix A. The details for computing the mixed moments (I.9) go otherwise the same way as the expressions in Appendix A, with the exception that we have different values for the Vandermonde mixed moment expansion coefficients. However, the integrals defining these coefficients may be hard to compute for a non-uniform power distribution, even for the case of uniform phase distribution, since Fourier-Motzkin elimination (see Section I.4) can be applied only in the case of uniform power- and phase distribution.

We conjecture that Theorem 6 holds also for general power distributions. It is likely that a similar calculation as in Appendix C can prove this, but we do not go into details on this.

## I.4 Software implementation

In this section, we will repeatedly refer to the implementation [26], which contains all code needed to verify all results in this paper. Implementations therein have two purposes:

1. to generate the exact coefficients in the formulas in this paper (generated directly in latex),

2. to compute the convolution with a given set of moments numerically.

In Appendix A, we explain why iteration through partitions and Fourier-Motzkin elimination are two main things needed in the implementation. In this section, we will explain how these tasks can be implemented efficiently.

### I.4.1 Reducing the complexity in iterating over partitions

Formulas in [20] and in this paper sum over sets of partitions. Iterating over partitions is very time-consuming, and must therefore be performed efficiently. There are several ways how this can be performed[1]. It turns out that one can reduce the number of partitions needed for computations considerably.

Assume that $\mathbf{V}$ has uniform phase distribution, and consider

$$\operatorname{tr}\left(\mathbf{V}^H\mathbf{V}\cdots\mathbf{V}^H\mathbf{V}\right). \tag{I.23}$$

To compute (I.23), we traverse all partitions. For each partition an equation system is constructed, and the partition contributes with the volume of the corresponding solution set to the equation system in (I.23). The following observations [31, 20] simplifies this computation:

- If a block is a singleton, then the corresponding volume is the same as that of the partition with that block removed. By using this observation repeatedly, we obtain that any noncrossing partition gives 1 in volume contribution.
- If a block contains two successive elements, then the corresponding volume is the same as that of the partition with any one of the two elements removed
- If a partition is a cyclic shift of another, then the corresponding volumes are the same.

These observations can reduce the number of the computations dramatically. To make precise how these observations can be used, we state two definitions:

**Definition 4.** *A partition $\pi$ is said to be alternating if $i$ and $i+1$ (where the sum is taken cyclically mod $n$) are in different blocks for all $i$, and no blocks in $\pi$ are singletons. The alternating partition obtained by removing all singleton blocks and all successive elements in all blocks incrementally is called the standard form of the partition. The set of alternating partitions of $\{1,...,n\}$ with $k$ blocks is denoted $\mathcal{A}(n,k)$.*

**Definition 5.** *We say that two partitions are equivalent whenever one is a cyclic shift (with a fixed number of elements) of the other.*

Note that the standard form of any noncrossing partitions is the empty partition. The first two observations above say that computations only need

[1] The implementation in this paper uses an implementation [26] which lists all partitions of $n$ elements with a given number of blocks

to be performed for alternating partitions (since any partition can be reduced to an alternating one in standard form), while the third observation says that computations are only needed for one representative in each equivalence class, with equivalence defined as in Definition 5. For instance, there are 678570 partitions of $\{1, ..., 11\}$. The number of alternating partitions of the same set is 4427. The number of equivalence classes of alternating partitions is 715.

The moments of Vandermonde matrices can thus be computed by iterating over the smaller set of cyclic equivalence classes of alternating partitions. This iteration can be accomplished with a computer program[2]. We also need to keep track of the size of each equivalence class of alternating partitions. This is done by a program which efficiently hashes all partitions. This is a computationally intensive process, but which needs to be done only once for the required number of moments.

### I.4.2 Constructing linear equation systems

For a partition $\rho = \{W_1, ..., W_r\} \in \mathcal{P}(n)$, [20] relates (I.23) to the corresponding volume of the solution set of the equations

$$\begin{aligned} \sum_{k \in W_1} x_{k-1} &= \sum_{k \in W_1} x_k \\ \sum_{k \in W_2} x_{k-1} &= \sum_{k \in W_2} x_k \\ \vdots \quad & \quad \vdots \\ \sum_{k \in W_r} x_{k-1} &= \sum_{k \in W_r} x_k, \end{aligned} \tag{I.24}$$

where all variables are constrained to lie between 0 and 1. $\rho$ reflects how the $\omega_i$ are grouped into independent sets of variables: The left sides in (I.24) represent the $\mathbf{V}^H$-terms in the entries of the matrix product $\mathbf{V}^H\mathbf{V}$, whereas the right sides represent the $\mathbf{V}$-terms in the same matrix product. Equations of the form (I.24) also apply to the more general form [20]

$$\operatorname{tr}\left(\mathbf{V}_{i_1}\mathbf{V}_{i_1}^H \cdots \mathbf{V}_{i_k}\mathbf{V}_{i_k}^H\right), \tag{I.25}$$

where $\mathbf{V}_1, \mathbf{V}_2, ...$ are independent and with uniform phase distribution.

In Appendix A, it is shown that in order to express the arbitrary mixed moments of Vandermonde matrices (independent or not), we need to solve systems similar to (I.24), with the difference that different number of variables may appear on the left and right hand sides. Note that the volume of the solution set of (I.24) is always a rational number. This enables our implementation to generate exact formulas. For Toeplitz and Hankel matrices, it turns out that a subset of these equation systems serve the same role in order to compute their moments.

[2]It is not obvious how the observations can be applied in an efficient implementation. The implementation [26] first generates all partitions, and then picks out those which have the alternating property and no singleton blocks

### I.4.3 Solving the linear equation systems

In all cases of Toeplitz, Hankel, and Vandermonde, the coefficient matrix of the equations we construct has rank $r-1$ ($r$ being the number of blocks), and we need to find the number of solutions. Since we also have the constraints that $0 \leq x_i \leq 1$, this really corresponds to finding all solutions to a set of linear inequalities. A much preferred method for doing so is Fourier-Motzkin elimination [28]. The first step before we perform this elimination would be to bring the equations into a standard form. We do this by expressing the $r-1$ pivot variables (after row reduction) by means of the free variables. Since all variables are between 0 and 1 (which are split into two inequalities), our equations are

$$\begin{array}{rcl}
\sum_{j=1}^{n-r+1} a_{1j}x_j & \leq & 1 \\
\sum_{j=1}^{n-r+1} -a_{1j}x_j & \leq & 0 \\
\sum_{j=1}^{n-r+1} a_{2j}x_j & \leq & 1 \\
\sum_{j=1}^{n-r+1} -a_{2j}x_j & \leq & 0 \\
\vdots & & \vdots \\
\sum_{j=1}^{n-r+1} a_{(r-1)j}x_j & \leq & 1 \\
\sum_{j=1}^{n-r+1} -a_{(r-1)j}x_j & \leq & 0 \\
x_1 & \leq & 1 \\
-x_1 & \leq & 0 \\
x_2 & \leq & 1 \\
-x_2 & \leq & 0 \\
\vdots & & \vdots \\
x_{n-r+1} & \leq & 1 \\
-x_{n-r+1} & \leq & 0,
\end{array} \tag{I.26}$$

where we have re-indexed the variables so that $x_1, ..., x_{n-r+1}$ are the free variables, $x_{n-r+2}, ..., x_n$ are the pivot variables. The coefficients $a_{ij}$ are taken from $-1, 0, 1$, and are the coefficients we obtain when the pivot variables are expressed in terms of the free variables. By reordering the equations, we get what we call the standard form (where the equations are sorted by the first coefficient):

$$\begin{array}{rcl}
x_1 + \sum_{j=1}^{n-r} b_{1j}x_{j+1} & \leq & e_1 \\
\vdots \qquad\qquad\qquad\quad \vdots & & \vdots \\
x_1 + \sum_{j=1}^{n-r} b_{r_1 j}x_{j+1} & \leq & e_{r_1} \\
\sum_{j=1}^{n-r} c_{1j}x_{j+1} & \leq & f_1 \\
\vdots \qquad\qquad\qquad\quad \vdots & & \\
\sum_{j=1}^{n-r} c_{r_2 j}x_{j+1} & \leq & f_{r_2} \\
-x_1 + \sum_{j=1}^{n-r} d_{1j}x_{j+1} & \leq & g_1 \\
\vdots \qquad\qquad\qquad\quad \vdots & & \vdots \\
-x_1 + \sum_{j=1}^{n-r} d_{r_3 j}x_{j+1} & \leq & g_{r_3}.
\end{array} \tag{I.27}$$

Fourier-Motzkin elimination now consists of eliminating the first variable, and working on the remaining equations to eliminate variables iteratively. Most of the coefficient matrices here are *combinatorial matrices* on the same form as those in [28].

Fourier-Motzkin elimination is computationally intensive, in the sense that the number of inequalities grow rapidly during elimination. Our aim is to compute the volume of the solution set rather than finding specific solutions. The volume can be split into many smaller disjoint parts, each part corresponds to a choice of minimum (`min`) for the first equations, and a choice of maximum (`max`) for the last equations Each part corresponds to the solution of a set of equations with one less variable. More precisely, let the equations in (I.27) have coefficient vectors $B_1, ..., B_{r_1}, C_1, ..., C_{r_2}, D_1, ..., D_{r_3}$, so that

$$\begin{array}{rcll} B_i & = & (1, b_{i1}, ..., b_{i(n-r)}, e_i) & 1 \leq i \leq r_1 \\ C_i & = & (0, c_{i1}, ..., c_{i(n-r)}, f_i) & 1 \leq i \leq r_2 \\ D_i & = & (-1, d_{i1}, ..., d_{i(n-r)}, g_i) & 1 \leq i \leq r_3. \end{array}$$

Each choice of `min`, `max`, with $1 \leq \texttt{min} \leq r_1$, $1 \leq \texttt{max} \leq r_3$ gives rise to a volume described by the solution to the set of equations

$$\begin{array}{lll} B_k - B_{\texttt{min}} & 1 \leq k \leq r_1 & k \neq \texttt{min} \\ D_k - D_{\texttt{max}} & 1 \leq k \leq r_3 & k \neq \texttt{max} \\ D_{\texttt{max}} + B_{\texttt{min}} & & \\ C_k & 1 \leq k \leq r_2, & \end{array}$$

where the equations are described by row vectors as above. There are $r_1 - 1 + r_3 - 1 + 1 + r_2 = r_1 + r_2 + r_3 - 1$ equations here, which is one less equation than what we started with. Note that the first element is zero in all these equations, so that the first column can be removed in the coefficient matrix. Therefore, the original system has been reduced to one with one equation less and one less variable. There may be more zero leading columns also, and all these can be removed. When the leading column is nonzero, the rows are sorted so that we get a new system on the form (I.27), and the procedure continues. In the process, the choice of `max` and `min` have decided the lower and upper integral bounds for the $x_1$-variable. These are stored, and after all Fourier-Motzkin elimination we have a full set of integral bounds, and the corresponding volume is computed by integrating over these bounds. This can be implemented easily [26], since integration over a volume with integral bounds which are linear in the variables can be defined in terms of simple row operations and integration by parts.

### I.4.4 Optimizations for Fourier-Motzkin elimination

The challenge in computing the volumes of the solution sets in Fourier-Motzkin elimination lies in that there are many eliminations which need to be peformed, and we do this for every partition in a large set of partitions. The Fourier-Motzkin elimination steps themselves can be stored and reused, but this of little help since we have to keep track of the corresponding integral bounds for

the solution sets. There are however, a couple of optimizations which can be used during elimination:

- if both row and the negative of that row are present as an equation, the solution set is empty, so that we can stop elimination
- Duplicate rows can be deleted
- Rows where only the last elements differ can be merged.

## I.5 Simulations

Results in this paper have been concerned with finding the spectral limit distribution from those of the input matrices. However, in practice, one has a certain model where one or more parameters are unknown, one observes output from that model, and would like to infer on the parameters of the model. The strengths in the results of this paper lie in that this kind of "deconvolution" is made possible to infer on the parameters of various models. As an example,

1. From observations of the form $\mathbf{D}(N)\mathbf{V}^H\mathbf{V}$ or $\mathbf{D}(N)+\mathbf{V}^H\mathbf{V}$, one can infer on either the spectrum of $\mathbf{D}(N)$, or the spectrum or phase distribution of $\mathbf{V}$, when exactly one of these is unknown.
2. From observations of the form $\mathbf{V}_1^H\mathbf{V}_1\mathbf{V}_2^H\mathbf{V}_2$ or $\mathbf{V}_1^H\mathbf{V}_1+\mathbf{V}_2^H\mathbf{V}_2$, one can infer on the spectrum or phase distribution of one of the Vandermonde matrices, when one of the Vandermonde matrices is known.

Moreover, the complexity in this inference is dictated by the number of moments considered. We do not go into depths on all the different types of deconvolutions made possible, only sketch a very simple example of inference as in 1). The other types of deconvolution go similarly, since the implementation supports each of them through functions with similar signatures. The example only makes an estimate of the first lower order moments of the component matrix $\mathbf{D}(N)$. These moments can give valuable information: in cases where it is known that there are few distinct eigenvalues, and the multiplicities are known, only some lower order moments are needed in order to get an estimate of these eigenvalues. We remark that this kind of deconvolution can be improved by further development of a second order theory for Vandermonde matrices.

In Figure I.1, we have, for Vandermonde matrices of size $N \times L$ with $L = N$, and for increasing $N$, formed 10 observations of the form $\mathbf{D}(N)\mathbf{V}^H\mathbf{V}$. The average of the moments of these observations are then taken, and a method in the framework [26] is applied to get an estimate of the moments of $\mathbf{D}(N)$. In the simulation, we have compared the estimate for the second and third moment of $\mathbf{D}(N)$ obtained by the implementation, with the actual second and third moments. The diagonal matrix $\mathbf{D}(N)$ is chosen so that the distribution of its eigenvalues is $\frac{1}{3}\delta_{0.5} + \frac{1}{3}\delta_1 + \frac{1}{3}\delta_{1.5}$, i.e. $0.5, 1, 1.5$ are the only eigenvalues, and they have equal probability. The simulation seems to indicate that the

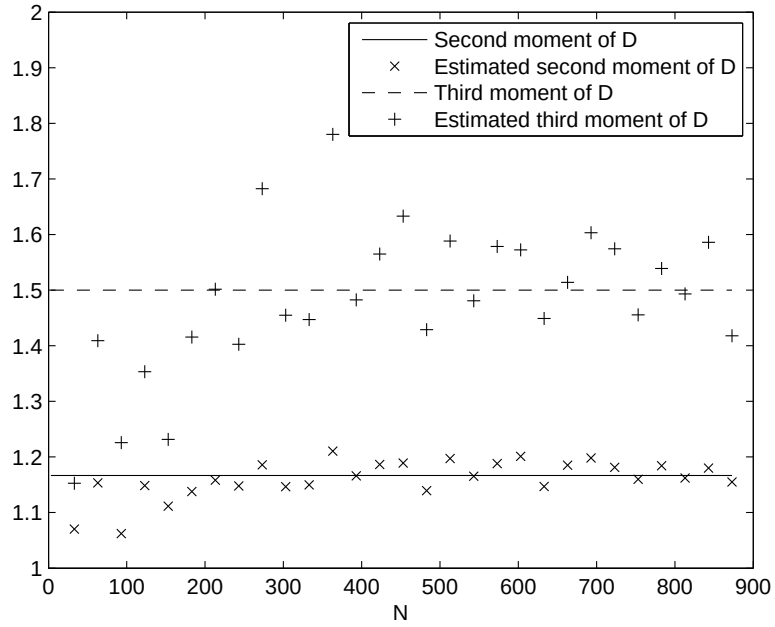


Figure I.1: Estimation of the second and third moment of $\mathbf{D}(N)$ from the average of 10 observations of the form $\mathbf{D}(N)\mathbf{V}^H\mathbf{V}$, for increasing values of $N$. $\mathbf{V}$ has dimensions $N \times N$.

implementation performs better estimation when the matrices grow large, in accordance with the fact that only an asymptotic result is applied.

Although it is difficult to make a full picture of the spectral distribution of $\mathbf{V}_1^H\mathbf{V}_1$ (or the phase distribution of $\mathbf{V}_1$) from deconvolution on models such as $\mathbf{V}_1^H\mathbf{V}_1\mathbf{V}_2^H\mathbf{V}_2$ (although the moments in many cases determine the distribution of the eigenvalues [**?**]), such deconvolution can still be useful. For instance, from the lower order moments one can to a certain amount say "how far away $\mathbf{V}_1$ is from having uniform phase distribution", since the uniform phase distribution achieves the lowest moments of all Vandermonde matrices [20].

## I.6 Conclusion and further directions

This contribution has explained how all types of moments in Vandermonde-type expressions can be obtained, and when one can expect that the moments/spectrum of the result only depend on the moments/spectrum of the input matrices (which is a requirement for performing deconvolution). The results can be used to compute the moments of any singular law involving a combination of many

independent matrices. An implementation which is capable of performing these moment computations is also presented, and moment formulas generated by the implementation were presented. The applications to wireless communications are still under study [32]. We have also described convolution operations on Vandermonde matrices which can not be performed in terms of the spectrum, but rather in terms of the phase distributions. We have also expanded known results on convergence of Vandermonde matrices to almost sure convergence.

Interestingly, Vandermonde matrices fit into a framework similar to that of freeness. Future papers will address a unified framework, where a more general theory which addresses when deconvolution is possible is presented.

It is still an open problem to find exact formulas for any moment of a Vandermonde matrix. The same applies to identifying these moments as the moments of a certain density. Future papers may also address how the implementation presented here can be made more efficient.

# Appendices

# A The proof of Theorem 1

Let us first assume that all phase distributions are uniform. Writing out the matrix product in (I.9) we get

$$\begin{aligned}\sum_{(i_1,\ldots,i_n)}\sum_{(j_1,\ldots,j_n)} & \\ &\prod_{i=1}^{|\sigma_1|} N_i^{-|\sigma_{1i}|/2} L^{-1} \\ &\times \mathbb{E}(e^{i_2(\omega_{\sigma_1(1),j_1}-\omega_{\sigma_1(2),j_2})}\times\cdots \\ &\qquad \times e^{i_1(\omega_{\sigma_1(2n-1),j_n}-\omega_{\sigma_1(2n),j_1})}) \\ &\times \mathbf{D}_1(N)(j_1,j_1)\times\cdots\times\mathbf{D}_n(N)(j_n,j_n), \end{aligned} \tag{I.28}$$

where

1. $1 \le j_1, ..., j_n \le L$ (as in [20])
2. $0 \le i_1, ..., i_n \le N_l - 1$ for appropriate $l$ (as in [20]),
3. $\sigma_1 = \{\sigma_{11}, ..., \sigma_{1|\sigma_1|}\}$ with $\sigma_{1k} = \{j | i_j = k\}$,
4. $\omega_{\sigma_1(i),j_i}$ is the phase for column $j_i$ in the $i$'th matrix entry.

Define the partition $\pi = \pi(j_1, ..., j_n) \in \mathcal{P}(n)$ by equality of the $j_i$, i.e. $k \sim_\pi l$ if and only if $j_k = j_l$. Noting that $\omega_{\sigma_1(k),j_k}$, $\omega_{\sigma_1(l),j_l}$ are equal if and only if $\sigma_1(k) = \sigma_1(l)$ and $j_k = j_l$ (if not they are independent), we define $\rho(\pi) \le \sigma_1 \in \mathcal{P}(2n)$ as the partition in $\mathcal{P}(n)$ generated by the relations:

$$k \sim_{\rho(\pi)} l \text{ if } \begin{cases} \lfloor k/2 \rfloor + 1 \sim_\pi \lfloor l/2 \rfloor + 1 \text{ and} \\ k \sim_{\sigma_1} l \end{cases}$$

Here $\lfloor x \rfloor$ means the largest whole number less than $x$. In other words, $k$ and $l$ are in the same block of $\rho(\pi)$ if and only if the corresponding phases $\omega_{\sigma_1(k),j_k}$ and $\omega_{\sigma_1(l),j_l}$ from the $k$'th and $l$'th matrix entries are dependent. We will have use for the following relation between $\rho(\pi)$ and $\pi$, which will help us to limit our calculations to a certain class of partitions.

**Lemma 1.** *The following holds:*

$$|\pi| \leq |\rho(\pi)| - r(\pi) + 1, \tag{I.29}$$

*Moreover, both equality and strict inequality can occur in (I.29).*

*Proof.* Since each block in $\pi$ is associated with at least one block in $\rho(\pi)$ by definition, we have that $|\pi| \leq |\rho(\pi)|$. Moreover, if $\rho_1$ is adjacent to $\rho_2$, they have a $j$-value common at their border, so that $|\pi| \leq |\rho(\pi)| - 1$. If $\rho_3$ is adjacent to $\{\rho_1, \rho_2\}$, they also have a $j$-value common at their border, so that also $|\pi| \leq |\rho(\pi)| - 2$. We can continue in this way for $\rho_4, \rho_5, ..., \rho_r$, and we obtain in the end that $|\pi| \leq |\rho(\pi)| - r(\pi) + 1$. It is also clear from this construction, by considering different border possibilities for the $\rho_1, \rho_2, ...$, that both equality and strict inequality can occur. □

In the following we will denote the set of partitions where (I.29) holds by $\mathcal{B}(n)$ (note that $\mathcal{B}(n)$ will also depend on $\sigma_1$, but this dependency will be implicitly assumed, and will thus not be mentioned in the following). Writing $\rho(\pi) = \{W_1, ..., W_{|\rho(\pi)|}\}$, there are $|\rho(\pi)|$ independent phases in the corresponding term, which we denote $\omega_{W_1}, ..., \omega_{W_{|\rho(\pi)|}}$. Write $W_j = W_j^{\cdot} \cup W_j^H$, where $W_j^{\cdot}$ consists of the even elements of $W_j$ (corresponding to the $\mathbf{V}$-terms), $W_j^H$ consists of the odd elements of $W_j$ (corresponding to the $\mathbf{V}^H$-terms). (I.9) can now be written (computations are similar to Appendix 1 in [20])

$$\begin{aligned}
&\sum_{\pi \in \mathcal{P}(n)} \sum_{(i_1,...,i_n)} \sum_{\substack{(j_1,...,j_n) \\ \pi(j_1,...,j_n)=\pi}} \\
&\quad \prod_{i=1}^{|\sigma_1|} N_i^{-|\sigma_{1i}|/2} L^{-1} \\
&\quad \times \prod_{r=1}^{|\rho(\pi)|} \mathbb{E}\left( e^{j\left(\sum_{k \in W_r^H} i_{(k+1)/2+1} - \sum_{k \in W_r^{\cdot}} i_{k/2+1}\right)\omega_{W_r}} \right) \\
&\quad \times \mathbf{D}_1(N)(j_1, j_1) \times \cdots \times \mathbf{D}_n(N)(j_n, j_n).
\end{aligned} \tag{I.30}$$

Since $\mathbb{E}\left(e^{jn\omega}\right) = 0$ when $\omega$ is uniform and $n \neq 0$, we get that the $i_1, ..., i_n$ contribute in (I.30) only if

$$\sum_{k \in W_r^H} i_{(k+1)/2+1} = \sum_{k \in W_r^{\cdot}} i_{k/2+1} \tag{I.31}$$

for $1 \leq r \leq |\rho(\pi)|$. The coefficient matrix of this system, denoted $A$, is a $|\rho(\pi)| \times n$ with entries from $\{-1, 0, 1\}$. The rank of $A$ is at most $k - 1$, since

the sum of all rows is 0. Note that the number of solutions to (I.31) can also be written

$$\int_{[0,2\pi)^{|\rho(\pi)|}} F(x)dx_1 \cdots dx_{|\rho(\pi)|}, \tag{I.32}$$

where

$$F(x) = \prod_{k=1}^{n} \frac{1 - e^{jN_{i_{2k}}(x_{b(2k-1)} - x_{b(2k)})}}{1 - e^{j(x_{b(2k-1)} - x_{b(2k)})}}, \tag{I.33}$$

where $b(k)$ means the block in $\rho(\pi)$ which $k$ belongs to. This follows (as in [20]) from summing over all possible choices of $i_1, ..., i_n$ in (I.30), and using the formula for the sum of a geometric series.

It is easily seen that the rank of $A$ is exactly $k-1$ when $\rho(\pi) \vee [0,1]_n = 1_{2n}$. More generally, if $\rho(\pi) \vee [0,1]_n = \{\rho_1, ..., \rho_r\}$ with each $\rho_i \geq [0,1]_{\|\rho_i\|/2}$, the rank of the system is $|\rho(\pi)| - r$. This follows since the equations corresponding to each $\rho_i$ have no variables in common with those from other $\rho_j$, and since the sum of the equations corresponding to $\rho_i$ is 0, so that the coefficient matrix corresponding to $\rho_i$ has one less than full rank. Also, note that $N_{i_{s+2}} = N_{i_{t+2}}$ whenever $2s+1, 2s+2, 2t+1, 2t+2$ all belong to the same such $\rho_i$, and denote this common value by $N_{\rho_i}$ (meaning that there is a common upper limit $N_{\rho_i}$ to all variables $i_l$ occurring in connection with the same block $\rho_i$). This means that the number of solutions to (I.31) is of order

$$\begin{aligned} & O\left(\prod_{i=1}^{r} N_{\rho_i}^{\|\rho_i\|/2-|\rho_i|+1}\right) \\ = \quad & O\left(\prod_{i=1}^{r} L^{\|\rho_i\|/2-|\rho_i|+1}\right) = O\left(L^{n-|\rho|+r}\right). \end{aligned}$$

Since $r$ depends only on $\pi$, we will also write $r = r(\pi)$. Since the number of solutions to (I.31) is given by (I.32), the limit

$$\lim_{L\to\infty} \frac{1}{\prod_{i=1}^{r} N_{\rho_i}^{\|\rho_i\|/2-|\rho_i|+1}} \int_{[0,2\pi)^{|\rho(\pi)|}} F(x)dx_1 \cdots dx_{|\rho(\pi)|} \tag{I.34}$$

exists (here $u$ denotes the uniform distribution), and we will denote this limit by $K_{\rho(\pi),u}$. Moreover $K_{\rho(\pi),u} = \prod_{i=1}^{r} K_{\rho_i,u}$, since the splitting $\rho(\pi) \vee [0,1]_n = \{\rho_1, ..., \rho_r\}$ actually splits the equations into $r$ sets where each set has no variables in common with other sets. This definition extends that of Vandermonde mixed moment expansion coefficients from [20] to the case where the equations (I.31) may have an unequal number of variables on each side. Note that in [20], those coefficients were defined in terms of $\pi \in \mathcal{P}(n)$, while here they are defined in terms of $\rho(\pi)$, which captures any $\sigma_1$, which is new in the analysis given here. Also in accordance with [20], we will denote by $K_{\rho(\pi),u,L}$ the quantity inside the limit of (I.34), so that the number of solutions to (I.31) is

$$\prod_{i=1}^{r} N_{\rho_i}^{\|\rho_i\|/2-|\rho_i|+1} K_{\rho(\pi),u,L}. \tag{I.35}$$

The number of blocks in the partitions $\rho(\pi), \pi$ say how many distinct choices from $(i_1, ..., i_n)$ and $(j_1, ..., j_n)$, respectively, contribute in (I.30). By substituting

$$\sum_{j_k} \mathbf{D}_i(N)(j_k, j_k) = L\text{tr}(\mathbf{D}(N)),$$

and using (I.29), we see that (I.30) is

$$\begin{aligned} & O(L^{-n-1+n-|\rho(\pi)|+r+|\pi|}) \\ \leq \quad & O(L^{-n-1+n-|\rho(\pi)|+r+|\rho(\pi)|-r+1}) = O(1), \end{aligned}$$

with equality if and only if $\pi \in \mathcal{B}(n)$ by Lemma 1. To check if $\pi$ belongs to $\mathcal{B}(n)$, $\rho(\pi)$ needs to be computed, and it is checked if equality in (I.29) holds. If so, the corresponding equation system (I.31) is constructed, and solved using Fourier-Motzkin elimination. Adding contributions for all partitions, we obtain (I.9).

For $\pi \in \mathcal{B}(n)$, noting that we can write $\prod_{i=1}^{|\sigma_1|} N_i^{-|\sigma_{1i}|/2} = \prod_{i=1}^{r} N_{\rho_i}^{-\|\rho_i\|/2}$, and using (I.35), we can write the contribution from $\pi$ in (I.30) as

$$\begin{aligned} & \prod_{i=1}^{r} N_{\rho_i}^{-\|\rho_i\|/2} L^{-1} L^{|\pi|} \prod_{i=1}^{r} N_{\rho_i}^{\|\rho_i\|/2-(|\rho_i|-1)} K_{\rho(\pi),u,L} D_\pi \\ = \quad & L^{|\rho(\pi)|-r(\pi)} \prod_{i=1}^{r} N_{\rho_i}^{-(|\rho_i|-1)} K_{\rho(\pi),u,L} D_\pi \\ = \quad & \prod_{i=1}^{r} L^{|\rho_i|-1} \prod_{i=1}^{r} N_{\rho_i}^{-(|\rho_i|-1)} K_{\rho(\pi),u,L} D_\pi \\ = \quad & \prod_{i=1}^{r} \left( \frac{L}{N_{\rho_i}} \right)^{|\rho_i|-1} K_{\rho(\pi),u,L} D_\pi. \end{aligned}$$

Thus, if $\omega$ is uniform, taking limits in (I.30) gives

$$\begin{aligned} & \sum_{\pi \in \mathcal{B}(n)} \prod_{i=1}^{r} \left( \frac{L}{N_{\rho_i}} \right)^{|\rho_i|-1} K_{\rho(\pi),u,L} D_\pi \\ \to \quad & \sum_{\pi \in \mathcal{B}(n)} \prod_{i=1}^{r} c_{\rho_i}^{|\rho_i|-1} K_{\rho(\pi),u} D_\pi \\ = \quad & \sum_{\pi \in \mathcal{B}(n)} \prod_{i=1}^{r} \left( c_{\rho_i}^{|\rho_i|-1} K_{\rho_i,u} \right) D_\pi, \end{aligned}$$

where we have substituted $c_{\rho_i} = \lim_{L\to\infty} \frac{L}{N_{\rho_i}}$. When the $\omega_i$ are not uniform, we can still in (I.30) sum over the different $i_1, ..., i_n$ to factor out the term

$$\int_{[0,2\pi)^{|\rho(\pi)|}} F(\omega) d\omega_1 \cdots d\omega_{|\rho(\pi)|}, \tag{I.36}$$

where $F$ is defined by (I.33), and where the only difference from (I.32) is that the uniform distribution $u$ has been replaced with $\omega$. The analysis is otherwise the same as in the uniform case, the major issue being the existence of the limits

$$\lim_{L\to\infty} \frac{1}{\prod_{i=1}^r N_{\rho_i}^{\|\rho_i\|/2-|\rho_i|+1}} \int_{[0,2\pi)^{|\rho(\pi)|}} F(\omega) d\omega_1 \cdots d\omega_{|\rho(\pi)|}, \tag{I.37}$$

which thus also will be called Vandermonde mixed moment expansion coefficients, and denoted $K_{\rho(\pi),\omega}$. As in the uniform case, note that $K_{\rho(\pi),\omega} = \prod_{i=1}^r K_{\rho_i,\omega}$, and if these limits exist, we get as in the uniform case a limit on the form

$$\sum_{\pi\in\mathcal{B}(n)} \prod_{i=1}^r \left( c_{\rho_i}^{|\rho_i|-1} K_{\rho_i,\omega} \right) D_\pi.$$

In [20], it was shown that the limits $K_{\pi,\omega}$ exist when $\omega$ has a continuous density. We will show that the same holds for $K_{\rho(\pi),\omega}$, and the proof of this will follow from the following lemma, which is a generalization of Lemma 2 in [20]:

**Lemma 2.** *Let $\rho(\pi) \leq \sigma_1 \in \mathcal{P}(2n)$ be any partition such that $\rho(\pi) \vee [0,1]_n = 1_{2n}$. For any $\epsilon > 0$,*

$$\lim_{N\to\infty} \frac{1}{N^{n+1-|\rho(\pi)|}} \int_{B_{\epsilon,k}} F(\omega) d\omega = 0, \tag{I.38}$$

*where*

$$B_{\epsilon,k} = \{(\omega_1, ..., \omega_{|\rho(\pi)|}) | |\omega_{b(2k-1)} - \omega_{b(2k)}| > \epsilon\}, \tag{I.39}$$

*and where $b(k)$ denotes the block in $\rho(\pi)$ which $k$ belongs to.*

*Proof.* The condition $\rho(\pi) \vee [0,1]_n = 1_{2n}$ implies that when $\omega \in B_{\epsilon,k}$, $\omega_i - \omega_j < 2n\epsilon$ for all $i, j$, which means that the definition of $B_{\epsilon,k}$ is similar to the definition of $B_{\epsilon,r}$ in Lemma 2 in Appendix H in [20]. The proof otherwise follows the same lines as [20]. □

Using Lemma 2 repeatedly, we see also that

$$\lim_{N\to\infty} \frac{1}{N^{n+r-|\rho(\pi)|}} \int_{\cup_i B_{\epsilon,k_i}} F(\omega) d\omega = 0$$

Letting $N \to \infty$, we obtain as in Appendix H of [20] in the limit

$$\begin{aligned}
&\lim_{N\to\infty} \frac{1}{N^{n+r-|\rho(\pi)|}} \int F(\omega) d\omega \\
&= K_{\rho(\pi),u} \prod_{i=1}^r (2\pi)^{|\rho_i\cap\sigma_{1j}|-1} \int \prod_j p_{\omega_j}(x)^{|\rho_i\cap\sigma_{1j}|} dx \\
&= K_{\rho(\pi),u} \prod_{i=1}^r (2\pi)^{|\rho_i\cap\sigma_j|-1} \int \prod_j p_{\omega_j}(x)^{|\rho_i\cap\sigma_j|} dx,
\end{aligned}$$

where $\sigma_j$ are the blocks of $\sigma$. Things have now been reduced to the case of uniform phase distribution. In summary, (I.30) can be written

$$\sum_{\pi\in\mathcal{B}(n)} D_\pi \prod_{i=1}^{r} \left((2\pi c_{\rho_i})^{|\rho_i|-1} K_{\rho_i,u}\right) \prod_{i=1}^{r} \int \prod_j p_{\omega_j}(x)^{|\rho_i\cap\sigma_j|} dx, \tag{I.40}$$

This is the standard form which the implementation uses, where the output of Fourier-Motzkin elimination is substituted into $K_{\rho(\pi),u}$. In (I.40), we recognize the integrals $I_{k,\omega}$ in (I.12). We will therefore substitute $I_{k,\omega}$ in the following.

The requirement from Theorem 1 that $\sigma \geq [0,1]_n$ (which happens whenever terms of the form $\mathbf{V}_{\omega_1}^H \mathbf{V}_{\omega_2}$ (with $\omega_1, \omega_2$ different and $\mathbf{V}_{\omega_1}, \mathbf{V}_{\omega_2}$ independent) do not occur) translates to the fact that, for any $\pi \in \mathcal{P}(n)$, in all $\rho_i$ the corresponding random matrices have equal phase distributions (with no assumptions on whether the random matrices are independent or not). From this it follows from (I.40) that no integrand in (I.40) will contain two different densities. Therefore, the mixed moment is completely determined from the integrals $I_{k,\omega}$. To make the connection between these quantities and the moments, we need the following lemma, compiled from [20]:

**Lemma 3.** *Let* $\mathbf{V}$ *be a Vandermonde matrix with phase distribution* $\omega$ *and aspect ratio* $c$*. For each* $n$ *there exists an invertible* $n \times n$ *matrix* $A_n$ *so that*

$$\begin{pmatrix} 1 \\ cI_{2,\omega} \\ c^2 I_{3,\omega} \\ \vdots \\ c^{n-1} I_{n,\omega} \end{pmatrix} = A_n \begin{pmatrix} 1 \\ V_2 \\ V_3 \\ \vdots \\ V_n \end{pmatrix} \tag{I.41}$$

Inserting (I.41) in (I.40) when each $\rho_i$ consists of equal phase distributions, we obtain that (I.9) is completely determined from the moments $V_n^{(i)}$. Since there are $r$ integrals multiplied together in (I.40), its general form is seen to coincide with that of (I.10). We have thus proved Theorem 1. When terms of the form $\mathbf{V}_{\omega_1}^H \mathbf{V}_{\omega_2}$ occur, Section I.3.4 shows that we can't expect dependence on only the moments. Instead the mixed moment depends on the entire phase distribution.

## A Handling the aspect ratio

We will finally comment on appropriate forms of (I.40) which are useful in implementations of convolution. We first turn to the case when there are deterministic matrices present. Assume that all matrix aspect ratios are equal to $c$, so that $\prod_{i=1}^{r} \left(c_{\rho_i}^{|\rho_i|-1}\right) = c^{|\rho|-r} = c^{|\pi|-1}$ when $\pi \in \mathcal{B}(n)$. Defining $m_n = cM_n$ and $d_n = cD_n$ as in [20]) in this case, (I.40) can also be written

$$m_n = \sum_{\substack{\rho(\pi)\leq\sigma_1 \\ \pi\in\mathcal{B}(n)}} d_\pi K_{\rho(\pi),u} \prod_{i=1}^{r} \int \prod_j p_{\omega_j}(x)^{|\rho_i\cap\sigma_j|} dx, \tag{I.42}$$

i.e. the aspect ratio $c$ can be handled as in [20], providing a clear parallel with Proposition 3 in that paper.

When there are no deterministic matrices present, and $\sigma \geq [0,1]_n$, the right hand side in (I.40) is

$$\prod_{i=1}^{r} c_{\rho_i}^{|\rho_i|-1} I_{|\rho_i|,\omega_i} K_{\rho_i,u}, \tag{I.43}$$

where we recognize the elements in the vector on the left hand side in (I.41). Therefore, an implementation of convolution would first compute the $c^{k-1} I_{k,\omega_i}$ using Lemma 3, and substitute these directly into (I.43). For deconvolution, (I.43) would be computed first for one of the unknown component matrices, and then the moments would be recovered in a second step using Lemma 3.

In summary, in order to compute the mixed moments (I.9) of Vandermonde matrices, we need to

1. iterate through partitions $\pi \in \mathcal{P}(n)$, compute $\rho(\pi)$, and determine whether $\pi \in \mathcal{B}(n)$,
2. perform Fourier-Motzkin elimination in order to solve the set of equations given by (I.31),
3. compute the quantities in (I.43), either from direct knowledge of the phase distribution, or by computing them from the moments using (I.41),
4. compute the final result by inserting the results from 1), 2), and 3) into (I.40).

This explains why Section I.4 focuses on the implementation perspectives of these tasks.

# B The proofs of the convolution formulas

In this appendix, we provide additional remarks, which together with the proof in Appendix A will suffice to prove the different convolution formulas. We first provide a short explanation why convolution 2) does not depend only on the spectra of the component matrices. When $\mathbf{D}_i(N)$ only occurs in patterns of the form $\mathbf{V}\mathbf{D}(N)\mathbf{V}^H$, we factored out the moments of $\mathbf{D}(N)$ in (I.30) in Appendix A. For other patterns, one ends instead up with integral expressions along the diagonal of $\mathbf{D}(N)$ (the diagonal elements of $\mathbf{D}(N)$ are multiplied with different complex exponentials), which are hard to express in terms of the moments of $\mathbf{D}(N)$.

## A The proof of Theorem 2

We can sum over all $\pi$ in (I.40) for these convolutions, since $r(\pi) = 1$ and $|\rho(\pi)| = |\pi|$ for all $\pi$ whenever $\sigma = \sigma_1 = 1_{2n}$. The implementation has obtained the result by inserting (I.41) into (I.40). $D_\pi$ in (I.40) can be handled in the following way: Let $R_n$ be the set of multi-indices $r = (r_1, ..., r_s)$ such that

- The $r_i$ are decreasing, and all are integers $> 0$.
- $\sum r_i = n$,

and set $d_r = \prod_{i=1}^{s} d_{r_i}$ for $r = (r_1, ..., r_s) \in R_n$. Set also

$$\begin{aligned} D_1 &= (d_1) \\ D_2 &= (d_2, d_1^2) \\ D_3 &= (d_3, d_2 d_1, d_1^3) \\ D_4 &= (d_4, d_2^2, d_3 d_1, d_2 d_1^2, d_1^4) \\ D_5 &= (d_5, d_3 d_2, d_4 d_1, d_2^2 d_1, d_3 d_1^2, d_2 d_1^3, d_1^5), \end{aligned}$$

and so on. It is clear from Appendix A that we can find a vector $K_n$ such that

$$\begin{pmatrix} M_1 \\ M_2 \\ M_3 \\ \vdots \\ M_n \end{pmatrix} = D_n^T K_n A_n \begin{pmatrix} 1 \\ V_2 \\ V_3 \\ \vdots \\ V_n \end{pmatrix}. \tag{I.44}$$

Moreover, the matrices $K_n$ and $A_n$ can be computed once and for all.

We see that there is only one term on the right hand side in (I.44) here containing $d_n$, so that this term can be found once $d_1, ..., d_{n-1}$ have been found. This enables us to perform deconvolution.

## B The proof of Theorem 3

Write $\operatorname{tr}\left((\mathbf{D} + \mathbf{V}^H \mathbf{V})^n\right)$ as

$$\begin{aligned} &= \sum_{k,s} \sum_{\substack{(r_1, ..., r_k) \\ \sum r_i = n-k-s}} \\ &\quad \operatorname{tr}\left( \underbrace{\mathbf{D} \cdots \mathbf{D}}_{r_1 \text{ times}} \mathbf{V}^H \mathbf{V} \underbrace{\mathbf{D} \cdots \mathbf{D}}_{r_2 \text{ times}} \mathbf{V}^H \mathbf{V} \underbrace{\mathbf{D} \cdots \mathbf{D}}_{s \text{ times}} \right) \\ &= \sum_{k,s} \sum_{\substack{(r_1, ..., r_k) \\ \sum r_i = n-k \\ r_1 \geq s}} \operatorname{tr}\left( \mathbf{D}^{r_1} \mathbf{V}^H \mathbf{V} \cdots \mathbf{D}^{r_k} \mathbf{V}^H \mathbf{V} \right). \end{aligned}$$

Each summand here can be computed by inserting (I.41) into (I.40) as above. Also, the multi-indices $(r_1, ..., r_k)$ are easily traversed. It is clear that one can generalize (I.44) to compute each summand (the vector $K_n$ is simply expanded to handle more mixed moments). This explains how the implementation computes the formulas for Theorem 3.

Deconvolution for Theorem 3 follows the exact same argument as for Theorem 2.

## C The proof of Theorem 4

(I.16) corresponds to the case where

$$\sigma = \sigma_1 = \{\{1, 2, 5, 6, 9, 10, ...\}, \{3, 4, 7, 8, 11, 12, ...\}\}.$$

Following the notation in (I.30) in Appendix A, when $\pi = 1_{2n}$ in (I.16), $\rho(\pi) = [0, 1]_{2n}$, so that the contribution is

$$k \int p_{\omega_1}^n(x)dx \int p_{\omega_2}^n(x)dx$$

contributes, where $k$ is a scalar. Also, for all other choices of $\pi$, the integral $I_{n,\omega_2}$ does not contribute, so that the equation for th $n$'th moment uniquely determines $I_{n,\omega_2}$, when the lower order integrals $\{I_{k,\omega_2}\}_{k<n}$ are known. Due to Lemma 3, the same can be said for the moments, so that it is possible to perform deconvolution.

Similarly, the contribution from $\pi = 1_n$ in (I.17) for the term when the second summand is always chosen is $kI_{n,\omega_2}$, where $k$ is a scalar. Moreover, $I_{n,\omega_2}$ contributes only for this term and this $\pi$, so that the equation for th $n$'th moment uniquely determines $I_{n,\omega_2}$. It follows as above that it is possible to perform deconvolution.

## D The proof of Theorem 5

This case corresponds to $\sigma = 1_{2n}$, and

$$\sigma_1 = \{[2, 3], [4, 5], ..., [2n - 2, 2n - 1], [2n, 1]\}.$$

From (I.40) it is clear that we can write

$$\begin{pmatrix} M_1 \\ M_2 \\ \vdots \\ M_n \end{pmatrix} = K_n \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_{2n} \end{pmatrix}, \tag{I.45}$$

where $K_n$ is an $n \times 2n$ matrix, depending only on the values computed from Fourier-Motzkin elimination.

Deconvolution in general for (Theorem 5) is impossible, since the equation system (I.45) has twice as many unknowns as equations. So, in this case, we need some prior knowledge about the phase distribution in order to perform deconvolution.

## E The proofs of Theorem 7 and Theorem 8

For Toeplitz matrices, [29] shows that we can compute the moments in the same way as for Vandermonde matrices, but that we need only consider equations on the form (I.24) with all blocks of $\rho$ of cardinality two. The case of Hankel

matrices is similar, however here the variables in (I.24) are placed differently on the left and right sides[3].

# C The proof of Theorem 6

Assume first that all aspect ratios are equal to $c$. In (25) in Theorem 7 in [20], set $\mathbf{D}_i(N) = I_L$, and place the last matrix $\mathbf{V}_{i_1}$ in front instead to obtain

$$\begin{array}{l} \lim_{N\to\infty} \mathbb{E}[\mathrm{tr}(\mathbf{V}_{i_1}\mathbf{V}_{i_1}^H\mathbf{V}_{i_2}\mathbf{V}_{i_2}^H \times \cdots \times \mathbf{V}_{i_n}\mathbf{V}_{i_n}^H)] \\ = \sum_{\rho\le\sigma\in\mathcal{P}(n)} K_{\rho,\omega} c^{|\rho|} \end{array} \tag{I.46}$$

(note that $c^{|\rho|}$ appears instead of $c^{|\rho|-1}$ since $\mathbf{V}_{i_1}$ is moved to front, and thus the additional $c$-factor is due to the fact that we take the trace of a matrix with different dimensions). As in Appendix A it is straightforward to generalize this to the case where the independent Vandermonde matrices have different aspect ratios, i.e. (I.46) is

$$\sum_{\rho\le\sigma\in\mathcal{P}(n)} K_{\rho,\omega} \prod_{i=1}^{|\sigma|} c_i^{|\rho\cap\sigma_i|},$$

where $\rho\cap\sigma_i$ is the partition consisting of the blocks of $\rho$ contained in $\sigma_i$. Using Theorem 8 in [20] (i.e. we also assume that the phase distributions are different, with $\mathbf{V}_i$ having phase distribution $\omega_i$), we thus generalize (25) to

$$\begin{aligned} & \sum_{\sigma\ge\rho} K_{\rho,u}(2\pi)^{|\rho|-1} \int_0^{2\pi} \prod_{i=1}^{s} p_{\omega_i}(x)^{|\rho\cap\sigma_i|} dx \prod_{i=1}^{|\sigma|} c_i^{|\rho\cap\sigma_i|} \\ = & \sum_{\sigma\ge\rho} K_{\rho,u}(2\pi)^{|\rho|-1} \int_0^{2\pi} \prod_{i=1}^{s} (c_i p_{\omega_i}(x))^{|\rho\cap\sigma_i|} dx \\ = & K_{\rho,u}(2\pi)^{|\rho|-1} \int_0^{2\pi} \prod_{i=1}^{s} (c_1 p_{\omega_1}(x) + c_2 p_{\omega_1}(x))^{|\rho|} dx \\ = & (c_1+c_2)^{|\rho|} K_{\rho,u}(2\pi)^{|\rho|-1} \\ & \times \int_0^{2\pi} \prod_{i=1}^{s} \left( \frac{1}{c_1+c_2}(c_1 p_{\omega_1}(x) + c_2 p_{\omega_1}(x)) \right)^{|\rho|} dx \\ = & \lim_{N\to\infty} \mathbb{E}[\mathrm{tr}(\mathbf{V}_{\omega_1 *_{c_1,c_2} \omega_2, c_1+c_2} \mathbf{V}^H_{\omega_1 *_{c_1,c_2} \omega_2, c_1+c_2})^n], \end{aligned}$$

where we have used (I.46) on the density $\frac{1}{c_1+c_2}(c_1 p_{\omega_1}(x) + c_2 p_{\omega_1}(x))$.

[3] In the software described for this paper, Toeplitz, Hankel, and Vandermonde matrices all reuse the same code, but different sets of partitions are considered, depending on the type of the matrix. Also, the way the corresponding equation is constructed from the partition depends on the type of the matrix

# D The proof of Theorem 9

We will first concentrate on the proof for almost sure convergence for a single Vandermonde matrix, as this case is the simplest. This proof will follow the same lines as that of almost sure convergence in [29], in that one uses Chebyshev's inequality, the Borel Cantelli lemma, and the following result:

**Lemma 4.** *Assume that* $\mathbf{V}$ *is an ensemble of random Vandermonde matrices with a continuous phase distribution, such that* $\frac{L}{N} \to c$. *For any* $r \geq 1$ *there exists a constant* $C_r$ *such that, for all* $L$,

$$\mathbb{E}\left[\left(\mathrm{tr}\left(\left(\mathbf{V}^H\mathbf{V}\right)^r\right) - \mathbb{E}\left[\mathrm{tr}\left(\left(\mathbf{V}^H\mathbf{V}\right)^r\right)\right]\right)^4\right] \leq C_r L^{-3}. \tag{I.47}$$

Comparing with [29, 1], Lemma 4 suggests that Vandermonde matrices converge somewhat faster than Hankel- and Toeplitz matrices, but somewhat slower than Gaussian matrices.

*Proof.* We can write

$$\begin{aligned}
&\mathbb{E}\left[\left(\mathrm{tr}\left(\left(\mathbf{V}^H\mathbf{V}\right)^r\right) - \mathbb{E}\left[\mathrm{tr}\left(\left(\mathbf{V}^H\mathbf{V}\right)^r\right)\right]\right)^4\right] \\
&= \mathbb{E}\left[\left(\mathrm{tr}\left(\left(\mathbf{V}^H\mathbf{V}\right)^r\right)\right)^4\right] \\
&\quad -4\mathbb{E}\left[\mathrm{tr}\left(\left(\mathbf{V}^H\mathbf{V}\right)^r\right)\right]\mathbb{E}\left[\left(\mathrm{tr}\left(\left(\mathbf{V}^H\mathbf{V}\right)^r\right)\right)^3\right] \\
&\quad +6\left(\mathbb{E}\left[\mathrm{tr}\left(\left(\mathbf{V}^H\mathbf{V}\right)^r\right)\right]\right)^2\mathbb{E}\left[\left(\mathrm{tr}\left(\left(\mathbf{V}^H\mathbf{V}\right)^r\right)\right)^2\right] \\
&\quad -3\left(\mathbb{E}\left[\mathrm{tr}\left(\left(\mathbf{V}^H\mathbf{V}\right)^r\right)\right]\right)^4.
\end{aligned} \tag{I.48}$$

We use certain interval partitions to define the following classes of partitions in $\mathcal{P}(4r)$:

- $\mathcal{P}_0$: partitions $\pi$ such that

$$\pi \leq \{[1,r],[r+1,2r],[2r+1,3r],[3r+1,4r]\},$$

- $\mathcal{P}_{1,2}$: partitions $\pi \notin \mathcal{P}_0$ such that

$$\pi \leq \{[1,2r],[2r+1,3r],[3r+1,4r]\},$$

- $\mathcal{P}_{2,3}$: partitions $\pi \notin \mathcal{P}_0$ such that

$$\pi \leq \{[1,r],[r+1,3r],[3r+1,4r]\}$$

(all other $\mathcal{P}_{i,j}$ are defined similarly),

- $\mathcal{P}_{1,2,3}$: partitions $\pi \notin \mathcal{P}_0 \cup \mathcal{P}_{1,2} \cup \mathcal{P}_{1,3} \cup \mathcal{P}_{1,4} \cup \mathcal{P}_{2,3} \cup \mathcal{P}_{2,4} \cup \mathcal{P}_{3,4}$ such that

$$\pi \leq \{[1, 3r], [3r + 1, 4r]\}$$

  (all other $\mathcal{P}_{i,j,k}$ are defined similarly),

- $\mathcal{P}_{1,2,3,4}$: partitions which are in none of the sets $\mathcal{P}_0, \mathcal{P}_{i,j}, \mathcal{P}_{i,j,k}$.

These classes of partitions are indexed by which intervals in $\{[1, r], [r+1, 2r], [2r+1, 3r], [3r + 1, 4r]\}$ are joined (in the sense that at least one block in a partition $\pi$ in $\mathcal{P}_{1,2}$ should contain elements from both the first and second interval in $\{[1, r], [r + 1, 2r], [2r + 1, 3r], [3r + 1, 4r]\}$), and we can write $\mathcal{P}(4r)$ as a disjoint union:

$$\begin{aligned} \mathcal{P}(4r) &= \mathcal{P}_0 \cup \mathcal{P}_{1,2} \cup \mathcal{P}_{1,3} \cup \mathcal{P}_{1,4} \cup \mathcal{P}_{2,3} \cup \mathcal{P}_{2,4} \cup \mathcal{P}_{3,4} \\ &\quad \cup \mathcal{P}_{1,2,3} \cup \mathcal{P}_{1,2,4} \cup \mathcal{P}_{1,3,4} \cup \mathcal{P}_{2,3,4} \\ &\quad \cup \mathcal{P}_{1,2,3,4}. \end{aligned} \tag{I.49}$$

We will denote the set of sets on the right hand side in (I.49) by $\mathcal{S}$. Write

$$\begin{aligned} S_{T,\pi} &= \sum_{\substack{(j_1,\cdots,j_{4r}) \\ \pi(j_1,\ldots,j_{4r})=\pi}} \sum_{(i_1,\ldots,i_{4r})} N^{-4r} L^{-4} \\ &\quad \times \mathbb{E}_T \left( \prod_{k=1}^{n} \left( e^{\jmath(\omega_{b(k-1)} - \omega_{b(k)})i_k} \right) \right), \end{aligned} \tag{I.50}$$

where

1. $\pi = \pi(j_1, ..., j_{4r})$ is defined as in Appendix A,
2. $T$ is a subset of $\{1, 2, 3, 4\}$,
3. $\mathbb{E}_T(x_1 \cdots x_4) = \mathbb{E}(\prod_{i \in T} x_i) \prod_{i \in T^c} \mathbb{E}(x_i)$ (i.e. $T$ dictates which random variables $x_i$ are grouped within the same expectation),
4. $b(k)$ means the block of $\pi$ which $k$ belongs to (with $\omega_{W_1}, ..., \omega_{W_s}$ independent when $W_1, ...W_s$ are the blocks of $\pi$).
5. $k-1$ is formed modulo $\{[1, r], [r+1, 2r], [2r+1, 3r], [3r+1, 4r]\}$, meaning that the values of $k \to k - 1$ actually takes the form

$$\begin{aligned} 1 &\to r \\ r+1 &\to 2r \\ 2r+1 &\to 3r \\ 3r+1 &\to 4r \\ k &\to k-1,\ k \notin \{1, r+1, 2r+1, 3r+1\}, \end{aligned}$$

6. $N^{-4r}$ are all normalizing factors in (I.1), $L^{-4}$ are the normalizing factors which come from taking the four traces for each term in (I.48).

When we write out (I.48) (by writing out the matrix product as in Appendix A, we end up with sums of the form $\sum_\pi S_{T,\pi}$, with various values for $T$. We have in particular

$$\begin{aligned}\sum_{\pi\in\mathcal{P}(4n)} S_{\{1,2,3,4\},\pi} &= \mathbb{E}\left[\left(\mathrm{tr}\left(\left(\mathbf{V}^H\mathbf{V}\right)^r\right)\right)^4\right]\\ \sum_{\pi\in\mathcal{P}(4n)} S_{\{\},\pi} &= \left(\mathbb{E}\left[\mathrm{tr}\left(\left(\mathbf{V}^H\mathbf{V}\right)^r\right)\right]\right)^4.\end{aligned}$$

However, since only one Vandermonde matrix appears here, the analysis from Appendix A simplifies to the case $\sigma_1 = \sigma = 1_{4r}$, for which the quantities can be expressed directly in terms of $\pi \in \mathcal{P}(4r)$ rather than $\rho(\pi) \in \mathcal{P}(8r)$ (as in Appendix A), so that the notation from [20]) can be followed more closely. As with (I.30), (I.48) thus becomes

$$\begin{aligned}&\sum_{\pi\in\mathcal{P}(4r)} \left(S_{\{1,2,3,4\},\pi} - 4S_{\{2,3,4\},\pi} + 6S_{\{3,4\},\pi} - 3S_{\{\},\pi}\right)\\ =\ &\sum_{S\in\mathcal{S}}\sum_{\pi\in S}(S_{\{1,2,3,4\},\pi} - 4S_{\{2,3,4\},\pi}\\ &\qquad +6S_{\{3,4\},\pi} - 3S_{\{\},\pi}),\end{aligned} \tag{I.51}$$

due to the ordering of the expectations in (I.48). We now consider all possibilities for $S \in \mathcal{S}$ in (I.51). For $\pi \in \mathcal{P}_0$ it is clear that one can split the expectations further to obtain

$$S_{\{1,2,3,4\},\pi} = S_{\{2,3,4\},\pi} = S_{\{3,4\},\pi} = S_{\{\},\pi},$$

and by adding up we see that the contribution from $S \in \mathcal{P}_0$ in (I.51) is 0. Similarly, by splitting up the expectations as much as possible, the contributions for other $\pi$ in (I.51) is seen to be

$$\begin{aligned}\pi \in \mathcal{P}_{1,2} &: S_{\{1,2\},\pi} - 4S_{\{\},\pi} + 6S_{\{\},\pi} - 3S_{\{\},\pi}\\ \pi \in \mathcal{P}_{1,3} &: S_{\{1,3\},\pi} - 4S_{\{\},\pi} + 6S_{\{\},\pi} - 3S_{\{\},\pi}\\ \pi \in \mathcal{P}_{1,4} &: S_{\{1,4\},\pi} - 4S_{\{\},\pi} + 6S_{\{\},\pi} - 3S_{\{\},\pi}\\ \pi \in \mathcal{P}_{2,3} &: S_{\{2,3\},\pi} - 4S_{\{2,3\},\pi} + 6S_{\{\},\pi} - 3S_{\{\},\pi}\\ \pi \in \mathcal{P}_{2,4} &: S_{\{2,4\},\pi} - 4S_{\{2,4\},\pi} + 6S_{\{\},\pi} - 3S_{\{\},\pi}\\ \pi \in \mathcal{P}_{3,4} &: S_{\{3,4\},\pi} - 4S_{\{3,4\},\pi} + 6S_{\{3,4\},\pi} - 3S_{\{\},\pi}\\ \pi \in \mathcal{P}_{1,2,3} &: S_{\{1,2,3\},\pi} - 4S_{\{2,3\},\pi} + 6S_{\{\},\pi} - 3S_{\{\},\pi}\\ \pi \in \mathcal{P}_{1,2,4} &: S_{\{1,2,4\},\pi} - 4S_{\{2,4\},\pi} + 6S_{\{\},\pi} - 3S_{\{\},\pi}\\ \pi \in \mathcal{P}_{1,3,4} &: S_{\{1,3,4\},\pi} - 4S_{\{3,4\},\pi} + 6S_{\{3,4\},\pi} - 3S_{\{\},\pi}\\ \pi \in \mathcal{P}_{2,3,4} &: S_{\{2,3,4\},\pi} - 4S_{\{2,3,4\},\pi} + 6S_{\{3,4\},\pi} - 3S_{\{\},\pi}\\ \pi \in \mathcal{P}_{1,2,3,4} &: S_{\{1,2,3,4\},\pi} - 4S_{\{2,3,4\},\pi} + 6S_{\{3,4\},\pi} - 3S_{\{\},\pi}.\end{aligned}$$

Adding everything here, and using that the contributions from $\mathcal{P}_{i,j,k}$ all are equal for different $i, j, k$, and that the contributions from $\mathcal{P}_{i,j}$ all are equal for

different $i, j$ (which is obvious by associating each interval $[kr+1,(k+1)r]$ with $r$ values on a circle, noting that the different classes of partitions can be viewed as different ways of connecting the circles, and that the actual circles being joined does not matter for the final value), we obtain that (I.48) equals

$$\sum_{\pi \in \mathcal{P}_{1,2,3,4}} \left( S_{\{1,2,3,4\},\pi} - 4S_{\{2,3,4\},\pi} + 6S_{\{3,4\},\pi} - 3S_{\{\},\pi} \right),$$

i.e. we need only sum over $\pi \in \mathcal{P}_{1,2,3,4}$ (all other terms cancel). If the phase distribution is uniform, we consider the coefficient matrix for the equation system corresponding to $\pi \in \mathcal{P}_{1,2,3,4}$ (formed as in Appendix A). This has rank $|\pi| - 1$, so that the number of solutions $(i_1, ..., i_{4r})$ solving the equation system has order $N^{4r-|\pi|+1}$. Since the number of $j_1, ..., j_{4r}$ such that $\pi(j_1, ..., j_{4r}) = \pi$ is of order $O\left(L^{|\pi|}\right)$, (I.50) is

$$O\left(N^{-4r} L^{-4} L^{|\pi|} N^{4r-|\pi|+1}\right) = O(L^{-3}). \tag{I.52}$$

This proves the claim for the uniform distribution. When the Vandermonde matrices do not have uniform phase distribution, as long as the phase distribution is continuous, we can reduce to the case of uniform phase distribution using Lemma 2 and the techniques in Appendix A. The constant $C_r$ needs only to be modified by taking into account the maximum of all Vandermonde mixed moment expansion coefficients of order $4r$. □

To prove the general case, we must in (I.47) replace $\mathbf{V}^H\mathbf{V}$ with the combination appearing in (I.9). One in this case instead considers the interval partition

$$\{[1, nr], [nr+1, 2nr], [2nr+1, 3nr], [3nr+1, 4nr]\}$$

instead of the interval partition $\{[1, r], [r+1, 2r], [2r+1, 3r], [3r+1, 4r]\}$. The sets of partitions $\mathcal{P}_0, \mathcal{P}_{1,2}, ...$ are defined similarly, and they are now sets in $\mathcal{P}(4rn)$. For a mixed moment as in (I.48) (with $\mathbf{V}^H\mathbf{V}$ replaced with combinations as in(I.9)), one shows as before that only partitions in $\mathcal{P}_{1,2,3,4}$ contribute (i.e. all other terms cancel as above). Since the form (I.9) is used, one needs to construct the partition $\rho(\pi) \in \mathcal{P}(8rn)$ from $\pi \in \mathcal{P}(4rn)$, and as in Appendix A, only partitions satisfying (I.29) contribute (i.e. $\pi \in \mathcal{B}(4rn)$), and (I.52) becomes in this case

$$\begin{aligned} & O\left(N^{-4rn} L^{-4} L^{|\pi|} L^{4rn-|\rho(\pi)|+r(\pi)}\right) \\ = \;& O\left(N^{-4rn} L^{-4} L^{|\pi|} L^{4rn+1-|\pi|}\right) = O(L^{-3}), \end{aligned}$$

and the result follows.

# Bibliography


[1] F. Hiai and D. Petz, *The Semicircle Law, Free Random Variables and Entropy*. American Mathematical Society, 2000.

[2] D. V. Voiculescu, "Addition of certain non-commuting random variables," *J. Funct. Anal.*, vol. 66, pp. 323–335, 1986.

[3] ——, "Multiplication of certain noncommuting random variables," *J. Operator Theory*, vol. 18, no. 2, pp. 223–235, 1987.

[4] D. Voiculescu, "Circular and semicircular systems and free product factors," vol. 92, 1990.

[5] ——, "Limit laws for random matrices and free products," *Inv. Math.*, vol. 104, pp. 201–220, 1991.

[6] H. Bercovici and D. Voiculescu, "Free convolution of measures with unbounded support," *Indiana Univ. Math. J.*, vol. 42, no. 3, pp. 733–773, 1993.

[7] A. Nica and R. Speicher, *Lectures on the combinatorics of free probability*, ser. London Mathematical Society Lecture Note Series. Cambridge: Cambridge University Press, 2006, vol. 335.

[8] D. V. Voiculescu, K. J. Dykema, and A. Nica, *Free random variables*, ser. CRM Monograph Series. Providence, RI: American Mathematical Society, 1992, vol. 1, a noncommutative probability approach to free products with applications to random matrices, operator algebras and harmonic analysis on free groups.

[9] F. Benaych-Georges, "Infinitely divisible distributions for rectangular free convolution: classification and matricial interpretation," *Probab. Theory Related Fields*, vol. 139, no. 1-2, pp. 143–189, 2007.

[10] D. Voiculescu, "Limit laws for random matrices and free products," *Invent. Math.*, vol. 104, no. 1, pp. 201–220, 1991.

[11] N. Rao, J. Mingo, R. Speicher, and A. Edelman, "Statistical eigen-inference from large wishart matrices," *http://front.math.ucdavis.edu/0701.5314*, Jan, 2007.

[12] N. E. Karoui, "Spectrum estimation for large dimensional covariance matrices using random matrix theory," *http://front.math.ucdavis.edu/0609.5418*, Sept, 2006.

[13] V. L. Girko, *An Introduction to Statistical Analysis of Random Arrays*. The Netherlands, VSP, 1998.

[14] ——, *Theory of Stochastic Canonical Equations, Volumes I and II*. Kluwer Academic Publishers, 2001.

[15] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functionals of large random matrices," *The Annals of Applied Probability*, 2007.

[16] X. Mestre, "Improved estimation of eigenvalues of covariance matrices and their associated subspaces using their sample estimates," *submitted to IEEE Transactions on Information Theory*, 2008.

[17] 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.

[18] ——, "Channel capacity estimation using free probability theory," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5654–5667, November 2008.

[19] B. Dozier and J. W. Silverstein, "On the empirical distribution of eigenvalues of large dimensional information-plus-noise type matrices," *J. Multivariate Anal.*, vol. 98, no. 4, pp. 678–694, 2007.

[20] Ø. Ryan and M. Debbah, "Asymptotic behaviour of random vandermonde matrices with entries on the unit circle," *IEEE Trans. on Information Theory*, vol. 55, no. 7, pp. 3115–3148, 2009.

[21] B. Khan, M. Debbah, T. Y. Al-Naffouri, and Ø. Ryan, "Estimation of the distribution deployment of sensor networks," in *Proceedings of the International Symposium on Information Theory, ISIT2009*, 2009.

[22] 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.

[23] M. Kobayashi and M. Debbah, "On the secrecy capacity of frequency-selective fading channels: A practical vandermonde precoding," in *PIMRC, Cannes, France*, September 2008.

[24] R. L. de Lacerda Neto, L. Sampaio, H. Hoffsteter, M. Debbah, D. Gesbert, and R. Knopp, "Capacity of MIMO systems: Impact of polarization, mobility and environment," in *IRAMUS Workshop, Val Thorens, France*, January 2007.

[25] R. R. Muller, "A random matrix model of communication via antenna arrays," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2495–2506, 2002.

[26] Ø. Ryan, *Tools for convolution operations arising from Vandermonde matrices*, 2008, http://ifi.uio.no/~oyvindry/vandermondeconv/.

[27] ——, *Documentation for the Random Matrix Library*, 2009, http://ifi.uio.no/~oyvindry/rmt/doc.pdf.

[28] G. Dahl, "Combinatorial properties of Fourier-Motzkin elimination," *Electronic Journal of Linear Algebra*, vol. 16, pp. 334–346, 2007.

[29] W. Bryc, A. Dembo, and T. Jiang, "Spectral measure of large random Hankel, Markov and Toeplitz matrices," *The Annals of Probability*, vol. 34, no. 1, pp. 1–38, 2006.

[30] R. Norberg, "On the Vandermonde matrix and its application in mathematical finance," *working paper no. 162, Laboratory of Actuarial Mathematics, Univ. of Copenhagen*, 1999.

[31] A. Nordio, C.-F. Chiasserini, and E. Viterbo, "Reconstruction of multidimensional signals from irregular noisy samples," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4274–4285, September 2008.

[32] Ø. Ryan and M. Debbah, "Applications of vandermonde matrices to wireless communications," *work in progress*, 2009.

# Chapter II

# Structured Spatio-temporal Sample Covariance Matrix Enhancement with Application to Blind Channel Estimation in Cyclic Prefix systems


### Abstract

Multichannel aspect allows the introduction of blind channel estimation techniques. Most existing such techniques for frequency-selective channels are quite complex. In this paper, we consider the blind channel estimation problem for Single Input Multi Output (SIMO) cyclic prefix (CP) systems. We have shown before [1] that blind channel estimation becomes computationally much more attractive and more straight forward to analyze in terms of performance in CP systems. Inspired by the iterative sample covariance matrix (SCM) structure enhacement techniques of Cadzow and others [2], we propose here an algorithm to structure the sample block circulant covariance matrix by enforcing two essential properties: rank and FIR structure. These two properties are exhibited by the true covariance matrix in the case of FIR SIMO channels with spatially white noise and CP transmission. The proposed enhancement procedure leads to an interesting enhanced SCM, even for the single CP symbol case. A simulation study for some classical channel estimators that depend on the SCM (with and without structuring) is presented, indicating that structuring allows for considerable performance gain in terms of the channel normalized mean square error (NMSE) over a wide SNR range.

## II.1 Introduction

A wealth of blind channel estimation techniques have been introduced for spatio-temporal channels over the past decade, based on the singularity of the received signal power spectral density matrix [3]. This singularity can be exploited to separate the white noise contribution. The main problem characteristic in fact that allows channel identifiability is the minimum phase characteristic of the Single-Input Multiple-Output (SIMO) or MIMO matrix channel transfer function of the spatio-temporal channel. Spatio-temporal channels arise in mobile communications when multiple antennas or polarizations or beams are used at the receiver. Physical multi-channels can also arise in xDSL systems when the receiver has access to a complete cable bundle. Other problem formulations that lead to multi-channel models are the use of oversampling at the receiver or the decoupling of inphase and in-quadrature components when real symbols get modulated or the reception of multiple signal copies in ARQ protocols. A variety of blind symbol/channel estimation strategies can be developed depending on the amount of a priori information that gets formulated on the unknown symbols. In general, the less structure that gets exploited about the symbol alphabet, the less problems tend to be encountered with local minima. Of course, more estimation accuracy is obtained by exploiting more information. A reasonable strategy is hence to exploit a progressive range of algorithms exploiting increasing a priori information levels. The algorithm at the next level can be initialized with the estimate obtained at the previous level of a priori information. The memory introduced by a convolutive channel leads to the requirement of having to treat all available data in a contiguous observation interval in one shot if no suboptimality is allowed. This leads to problem formulations with large convolution matrices, large covariance matrices and high complexity. Attempts have been made by our own group to introduce asymptotic approximations, by approximating large Toeplitz convolution matrices by circulant matrices, to allow transformation to the frequency domain, or by others by introducing approximate DFT operations. Cyclic prefixes have been introduced in a number of existing systems such as OFDM systems for ADSL and wireless LANs. Recently, Orthogonal Frequency Division Multiple Access (OFDMA) has been adopted as a multiple access scheme for the Frequency Domain Duplexing Down-Link (FDD-DL)in LTE (Long Term Evolution) systems. The introduction of a cyclic prefix renders the transformation to the frequency domain clean and exact even for a finite data length. The resulting algorithmic simplifications will be detailed for a number of classical blind channel estimation methods. Furthermore, the same framework can be used to analyze the performance of the algorithms and the algorithmic simplifications also translate into much simplified performance expressions, which allow a direct and insightful analytical performance comparison between a number of algorithms.

This paper is organized as follows. Section II starts with the description of the basic baseband SIMO-cyclic prefix model. In section III we develope a unified framework for cyclic prefix system channel estimators. Section IV defines some classical blind channel estimators within the framework introduced

in section III. The algorithm for structuring the covariance matrix is developed in section V. In section VI, we provide the experimental results and finally a conclusion is drawn in Section VII.

## II.2 SIMO Cyclic Prefix Block TX Systems

Consider a SIMO system with M outputs:

$$\begin{aligned} \underbrace{\mathbf{u}[m]}_{M\times 1} &= \sum_{j=0}^{L_h-1} \underbrace{\mathbf{h}[j]}_{M\times 1}\,\underbrace{a[m-j]}_{1\times 1} + \underbrace{\mathbf{w}[m]}_{M\times 1} \\ &= \underbrace{H(q)}_{M\times 1}\,\underbrace{a[m]}_{1\times 1} + \underbrace{\mathbf{w}[m]}_{M\times 1} \end{aligned} \tag{II.1}$$

where $H(q) = \sum_{j=0}^{L} \mathbf{h}[j]\, q^{-j}$ is the SIMO system transfer function corresponding to the $z$ transform of the impulse response $\mathbf{h}[.]$. Equation (II.1) mixes time domain and $z$ transform domain notations to obtain a compact representation. In $H(q)$, $z$ is replaced by $q$ to emphasize its function as an elementary time advance operator over one sample period. Its inverse corresponds to a delay over one sample period: $q^{-1}\mathbf{a}[n] = \mathbf{a}[n-1]$.

Consider a (OFDM or single-carrier) CP block transmission system with $N$ samples per block. The introduction of a cyclic prefix of $K$ samples means that the last $K$ samples of the current block (corresponding to $N$ samples) are repeated before the actual block. If we assume w.l.o.g. that the current block starts at time 0, then samples $\mathbf{a}[N-K]\cdots\mathbf{a}[N-1]$ are repeated at time instants $-K, \ldots, -1$. This means that the output at sample periods $0, \ldots, N-1$ can be written in matrix form as

$$\begin{bmatrix} \mathbf{u}[0] \\ \vdots \\ \mathbf{u}[N-1] \end{bmatrix} = \mathrm{U}[0] \;=\; \mathrm{H}\;\mathrm{A}[0] + \mathrm{W}[0] \tag{II.2}$$

where the matrix H is not only (block) Toeplitz but even (block) circulant: each row is obtained by a cyclic shift to the right of the previous row. Consider now applying an $N$-point FFT to both sides of (II.2) at block $m$:

$$F_{N,M}\mathrm{U}[m] = F_{N,M} H\, F_N^{-1} \quad F_N \mathrm{A}[m] + F_{N,M}\mathrm{W}[m] \tag{II.3}$$

or with new notations:

$$\mathrm{Y}[m] \quad = \quad \mathcal{H} \qquad \mathrm{X}[m] \quad + \quad \mathrm{V}[m] \tag{II.4}$$

where $F_{N,M} = F_N \otimes I_M$ (Kronecker product: $A \otimes B = [a_{ij}B]$), $F_N$ is the $N$-point $N \times N$ DFT matrix, $\mathcal{H} = \mathrm{diag}\{\mathbf{h}_0, \ldots, \mathbf{h}_{N-1}\}$ is a block diagonal matrix with diagonal blocks $\mathbf{h}_n = \sum_{l=0}^{L_h-1} \mathbf{h}[l]\; e^{-j2\pi\frac{1}{N}nl}$, the $M \times 1$ channel transfer

function at tone $n$ (frequency $= n/N$ times the sample frequency). In OFDM, the transmitted symbols are in X$[m]$ and hence are in the frequency domain. The corresponding time domain samples are in A$[m]$. The OFDM symbol period index is $m$. In Single-Carrier (SC) CP systems, the transmitted symbols are in A$[m]$ and hence are in the time domain. The corresponding frequency domain data are in X$[m]$. The components of W are considered white noise, hence the components of V are white also. At tone (subcarrier) $n \in \{0, \ldots, N-1\}$ we get the following input-output relation

$$\underbrace{\mathbf{y}_n[m]}_{M\times 1} = \underbrace{\mathbf{h}_n}_{M\times 1}\ \underbrace{x_n[m]}_{1\times 1} + \underbrace{\mathbf{v}_n[m]}_{M\times 1} \tag{II.5}$$

where the symbol $x_n[m]$ belongs to some finite alphabet (constellation) in the case of OFDM.

## II.3 Frequency domain Framework for CIR Estimation

The basic idea relies on the fact that to get the cost function or information for the temporal channel response it suffices to sum up the cost functions or information over the tones after transforming back to the time domain. To be a bit more explicit, let $\mathbf{h}$ be the vectorized channel impulse response then there exists transformation matrices $G_k$ (containing DFT portions)such that

$$\mathbf{h}_k \ = \ G_k\, \mathbf{h}\,. \tag{II.6}$$

To be more accurate, $G_k$ is of size $M \times ML_h$ such that it contains the first $ML_h$ elements of the kth block row of $F_{N,M}$. Now, if at tone $k$ we have a cost function of the form

$$\mathbf{h}_k^H\, Q_k\, \mathbf{h}_k \tag{II.7}$$

then this induces a cost function for the overall channel impulse response of the form

$$\mathbf{h}^H\ \left[\sum_{k=0}^{N-1} G_k^H\, Q_k\, G_k\right]\ \mathbf{h} \tag{II.8}$$

and similarly for Fisher information matrices. So in what follows, we shall concentrate on the cost function for a given tone.

## II.4 Blind SIMO Channel Estimation

### II.4.1 Subchannel Response Matching (SRM)/Cross Relation method (CR)

The subchannel response Matching (SRM) estimator which was (re)invented four times in [5],[6],[7],[8], is based on a linear parametrization of the noise

subspace in terms of the channel coefficients [9] so that $P^{\perp}_{\mathbf{h}_k} = P_{\mathbf{h}_k^{\perp H}}$ where $\mathbf{h}_k^{\perp}$ is given by

$$\mathbf{h}_{k\,bal,min}^{\perp} = \begin{bmatrix} -\mathbf{h}_{k,2} & \mathbf{h}_{k,1} & 0 & \cdots & 0 \\ 0 & -\mathbf{h}_{k,3} & \mathbf{h}_{k,2} & \cdots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \mathbf{h}_{k,M} & 0 & \cdots & 0 & -\mathbf{h}_{k,1} \end{bmatrix}. \tag{II.9}$$

In (II.9) we choose the minimum number of rows in $\mathbf{h}_k^{\perp}$ which has a size $(M - \delta_{M,2}) \times M$. This noise parametrization is balanced in the sense that every subchannel appears the same number of times, in this case twice. A balanced $\mathbf{h}_k^{\perp}$ leads to $tr\{\mathbf{h}_k^{\perp}\mathbf{h}_k^{\perp H}\} = \alpha||\mathbf{h}_k||^2$, where $\alpha = 2 - \delta_{M,2}$. In the noiseless case, $\mathbf{y}_k = \mathbf{h}_k x_k$ and we have $\mathbf{h}_k^{\perp}\mathbf{y}_k = \mathcal{Y}_k\mathbf{h}_k = 0$. Based on this relation the channel at tone k can be uniquely determined up to a scale factor [7],[8], as the unique right singular vector of $\mathcal{Y}_k$ corresponding to the singular value zero. When noise is present, $\mathcal{Y}_k\mathbf{h}_k \neq 0$ and the SRM criterion is solved in the least-squares sense $||\mathbf{h}_k^{\perp}\mathbf{y}_k||_2^2 = tr\{\mathbf{h}_k^{\perp}\mathbf{y}_k\mathbf{y}_k^H\mathbf{h}_k^{\perp H}\}$. By the law of large numbers, asymptotically this criterion can be replaced by its expected value: $tr\{\mathbf{h}_k^{\perp}S_{\mathbf{y_k y_k}}\mathbf{h}_k^{\perp H}\}$. Practically, $S_{\mathbf{y_k y_k}}$is not available so it is replaced by the sampled spectrum per each tone $\hat{S}_{\mathbf{y_k y_k}}$ which is computed directly from the fourier transformed version of the received data as we will show in the next section. Moreover, the SRM criterion can be written in the form shown in (II.8) where $Q_k = \sum_{i=1}^{M} D_i\hat{S}^*_{\mathbf{y_k y_k}}D_i^H$ where $\mathcal{D}_{i+1} = \mathcal{C}\mathcal{D}_i\mathcal{C}$.

$$\mathcal{D}_{\mathbf{1}} = \begin{bmatrix} 0 & 1 & 0 & \cdots \\ -1 & 0 & \cdots & \\ 0 & \vdots & \ddots & \\ \vdots & & & \end{bmatrix}. \tag{II.10}$$

$$\mathcal{C} = \begin{bmatrix} 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & 0 & 1 & 0 \end{bmatrix}. \tag{II.11}$$

Then, we attempt to minimize the sum of the SRM criteria (cost functions) over all tones jointly to obtain an estimate of the channel impulse response $\mathbf{h}$. The SRM cost function is shown below:

$$\min_h \mathbf{h}^H \left[\sum_{k=0}^{N-1} G_k^H \left\{\sum_{i=1}^{M} D_i\hat{S}^*_{\mathbf{y_k y_k}}D_i^H\right\} G_k\right] \mathbf{h} \tag{II.12}$$

We denote the matrix between the braces in (II.12) by $Q_{SRM}$, it has the size of $ML_h \times ML_h$. The estimated channel impulse response $\hat{\mathbf{h}}_{SRM}$ that is obtained by solving (II.12) is the eigen vector that corresponds to the minimum

eigen value of $Q_{SRM}$. This solution has a scalar ambiguity that can be solved by forcing a least square constraint as follows: $\min_{\alpha} ||\mathbf{h}^0 - \alpha\hat{\mathbf{h}}_{SRM}||^2$. This yields the following solution:

$$\hat{\hat{\mathbf{h}}}_{SRM} = \frac{\hat{\mathbf{h}}_{SRM}^H \mathbf{h}^0}{||\hat{\mathbf{h}}_{SRM}||^2}\hat{\mathbf{h}}_{SRM} \tag{II.13}$$

### II.4.2 Noise Subspace Fitting (NSS)

The sampled spectrum per each tone $\hat{S}_{\mathbf{y_k y_k}}$ can be decomposed into signal and noise subspace contributions:

$$\begin{aligned} \hat{S}_{\mathbf{y_k y_k}} &= \hat{S}_{\mathbf{y}_k,\mathcal{S}} + \hat{S}_{\mathbf{y}_k,\mathcal{N}} \\ &= \hat{E}_{s,k}\hat{\Lambda}_{s,k}\hat{E}_{s,k}^H + \hat{E}_{n,k}\hat{\Lambda}_{n,k}\hat{E}_{n,k}^H \end{aligned} \tag{II.14}$$

The basic idea of the NSF is to fit the estimated noise subspace that we obtain from the sampled spectrum to the true noise subspace which is spanned by the columns of $\mathbf{h}_k^{\perp H}$.

$$\min_{h_k,T} \; ||\mathbf{h}_k^{\perp H} - \hat{E}_{n,k}T||_F \tag{II.15}$$

where $||X||_F = tr\{X^H X\}$. This criterion differs from the original subspace fitting strategy proposed in [10], which would propose $\min_{h_k,T} \; ||\hat{E}_{n,k} - \mathbf{h}_k^{\perp H}T||_F$ as criterion. We propose (II.15) because it leads to a simpler optimization problem. Both approaches can be made to be equivalent by the introduction of column space weighting. The cost function in (II.15) is separable. In particular, it is quadratic in T. Minimization w.r.t. T leads to $T = \hat{E}_{n,k}^H \mathbf{h}_k^{\perp H}$ and $\mathbf{h}_k^{\perp H} - \hat{E}_{n,k}\hat{E}_{n,k}^H\mathbf{h}_k^{\perp} = P_{\hat{E}_{n,k}}^{\perp}\mathbf{h}_k^{\perp H}$ where $P_{\hat{E}_{n,k}}^{\perp} = I - P_{\hat{E}_{n,k}} = P_{\hat{E}_{s,k}}$ and $P_{\hat{E}_{n,k}}, P_{\hat{E}_{s,k}}$ denote respectively the projection matrix on the noise subspace ($\hat{E}_{n,k}$) and the signal subspace ($\hat{E}_{s,k}$).

Hence,

$$\begin{aligned} \min_{h_k,T} \; ||\mathbf{h}_k^{\perp H} - \hat{E}_{n,k}T||_F &= \min_{h_k} \; ||P_{\hat{E}_{s,k}}\mathbf{h}_k^{\perp H}||_F \\ &= \min_{h_k} \; tr\{\mathbf{h}_k^{\perp}\hat{E}_{s,k}\hat{E}_{s,k}^H\mathbf{h}_k^{\perp H}\} \end{aligned} \tag{II.16}$$

Similar to the case of SRM, the NSF criterion can be written in the form shown in (II.8) where $Q_k = \sum_{i=1}^{M} D_i\hat{E}_{s,k}\hat{E}_{s,k}^H D_i^H$ and $D_i$ is the same as for the SRM criterion. Again, we attempt to minimize the NSF jointly over all the tones subject to the least square constraint to avoid introducing N constraints and to exploit the correlation exists between the different tones. Therefore, the NSF cost functions takes the following form:

$$\min_{h}\mathbf{h}^H \left[\sum_{k=0}^{N-1} G_k^H \left\{\sum_{i=1}^{M} D_i\hat{E}_{s,k}\hat{E}_{s,k}^H D_i^H\right\} G_k\right] \mathbf{h} \tag{II.17}$$

Following the same discussion as in case of SRM we get the following solution:

$$\hat{\tilde{\mathbf{h}}}_{NSF} = \frac{\hat{\mathbf{h}}_{NSF}^{H}\mathbf{h}^{0}}{||\hat{\mathbf{h}}_{NSF}||^{2}}\hat{\mathbf{h}}_{NSF} \tag{II.18}$$

The substanial computational power saving offered by our farmework namely, working per tones instead of working in the time domain, is elucidated by remarking that we perform Eigen Value Decomposition (EVD) of N matrices $\hat{S}_{\mathbf{y_k y_k}}$ each of size $(M \times M)$ while working in the time domain requires the EVD of a huge matrix $\hat{S}_{UU}$of size $(MN \times MN)$. Knowing that the number of operations required to perform the EVD of a matrix is proportional to the cubic of its size and the number of tones N for some systems (eg. LTE downlink) may reach to 2048, then the great advantage of our framework in terms of computational power saving becomes evident.

## II.5 Block Toeplitz Covariance Matrix Enhancement

Here we go back to sample covariance refinements suggested by Cadzow in the eighties [2] and which we tried to exploit to enhance the dereverberation of the acoustic channel [4]. The idea is to iteratively reinforce several structural properties, the reinforcement of which consists of a projection onto a convex set. The iterations then converge to the joint reinforcement of all properties. Theoretically, the matrix valued vector signal spectrum is of the form

$$S_{\mathbf{yy}}(z) = \mathbf{h}(z)\, S_{xx}(z)\, \mathbf{h}^{\dagger}(z) + S_{\mathbf{vv}}(z) \tag{II.19}$$

where $.^{\dagger}$ denotes paraconjugate which is defined as $h^{\dagger}(z) = h(1/z^{*})^{H}$ and $S_{\mathbf{vv}}(z) = \sigma_v^2\, I$ is the white noise spectrum. Saking for the simplicity of notations, we omit the index k in $\mathbf{S_{yy}}$. The signal part of the spectrum, $\mathbf{h}(z)\, S_{xx}(z)\, \mathbf{h}^{\dagger}(z)$ is singular, not because of spectral poverty as in the SISO case, but because of limited rank in the matrix dimension. In the SISO case, a stationary signal covariance matrix can only be singular if the signal consists of a number of (complex) sinusoids, with their number being smaller than the covariance matrix dimension. Singularity in the MIMO case has nothing to do with spectral poverty but with matrix singularity of the matrix spectrum at every frequency.

Inspired by [2], (II.19) suggests the following procedure. First we start with the sample spectrum at each tone:

$$\begin{aligned} &\widehat{S}_{\mathbf{yy}}(z_n) = \tfrac{1}{P}\textstyle\sum_{m=1}^{P} \mathbf{y}_n[m]\mathbf{y}_n^{H}[m] \ \ , \\ &n = 0, \ldots, N-1 \end{aligned} \tag{II.20}$$

where P is the number of OFDM symbols over which we compute the sample spectrum and $z_n = e^{j2\pi n/2N}$, with the following properties: $\widehat{S}_{\mathbf{y}}^{\dagger}(z_n) = \widehat{S}_{\mathbf{y}}^{H}(z_n)$ (Hermitian transpose).

Now, at each frequency bin $n$, $S_{\mathbf{y}}(z_n)$ is of the form

$$\begin{aligned} S_{\mathbf{y}}(z_n) &= S_{\mathbf{y},\mathcal{S}}(z_n) + S_{\mathbf{y},\mathcal{N}}(z_n) \\ &= \mathbf{h}(z_n)\, S_x(z_n)\, \mathbf{h}^\dagger(z_n) + \sigma_v^2\, I_M \\ &= V_{max,n}\, (\lambda_{max,n} - \sigma_v^2)\, V_{max,n}^H + \sigma_v^2\, I_M \end{aligned} \tag{II.21}$$

where $S_{\mathbf{y},\mathcal{S}}(z_n)$, $S_{\mathbf{y},\mathcal{N}}(z_n)$ are the signal and noise components of $S_{\mathbf{y}}(z_n)$, and $\lambda_{max,n}$ and $V_{max,n}$ are its maximum eigenvalue and corresponding eigenvector. Now, the $\widehat{S}_{\mathbf{y}}(z_n)$ can be forced to the closest (in Frobenius norm) matrix of the form in (II.21) by computing its spatial eigen decomposition. Let $\widehat{\lambda}_{1,n} \geq \widehat{\lambda}_{2,n} \geq \cdots \geq \widehat{\lambda}_{M,n}$ be its eigenvalues, hence $\widehat{\lambda}_{max,n} = \widehat{\lambda}_{1,n}$, $\widehat{V}_{max,n} = \widehat{V}_{1,n}$. Then we get $\widehat{S}_{\mathbf{y}}(z_n) = \widehat{S}_{\mathbf{y},\mathcal{S}}(z_n) + \widehat{S}_{\mathbf{y},\mathcal{N}}(z_n) = \widehat{V}_{max,n}\, (\widehat{\lambda}_{max,n} - \widehat{\sigma}_v^2)\, \widehat{V}_{max,n}^H + \widehat{\sigma}_v^2\, I_M$ with $\widehat{\sigma}_v^2 = \frac{1}{N(M-1)} \sum_{n=0}^{N-1} \sum_{i=2}^{M} \widehat{\lambda}_{i,n}$ due to the spatiotemporal white noise assumption. Note that in fact at every frequency bin only $\lambda_{max,n}$ and $V_{max,n}$ need to be computed since $\sum_{i=2}^{M} \widehat{\lambda}_{i,n} = \mathrm{tr}\,\{\widehat{S}_{\mathbf{y}}(z_n)\} - \widehat{\lambda}_{max,n}$. Since the noise spectrum $\widehat{S}_{\mathbf{y},\mathcal{N}}(z_n) = \widehat{\sigma}_v^2\, I_M$ is fairly simple, there is no further structure to be imposed. The signal spectrum $\widehat{S}_{\mathbf{y},\mathcal{S}}(z_n) = \widehat{V}_{max,n}\, (\widehat{\lambda}_{max,n} - \widehat{\sigma}_v^2)\, \widehat{V}_{max,n}^H$ on the other hand is supposed to be spectrum of a FIR correlation sequence. This FIR character can be imposed by windowing in the time domain. The resulting source whitened signal spectrum $\widehat{S}_{\mathbf{y},\mathcal{S}}(z_n)$ then undergoes IFFT to obtain the corresponding matrix correlation sequence. The frequency-wise rank structure enforcement will have destroyed the FIR character of the correlation sequence, which can then simply be enforced in the time domain by proper windowing (without forgetting the symmetry structure of the first block column of the block circulant matrix). The operations of eigen structure enforcement in frequency domain and FIR structure enforcement in the time domain can then be iterated untill convergence. Typically a few iterations suffice. We are now ready to state the following iterative process:

1. Compute the matrix spectrum $\widehat{S}_{\mathbf{y}}(z_n)$ at each frequency bin as illustrated in (II.20).

2. Compute the eigendecomposition of the spectrum $\widehat{S}_{\mathbf{y}}(z_n)$ at each frequency bin $n = 0, 1, \ldots, N-1$. Determine the noise variance $\widehat{\sigma}_v^2 = \frac{1}{N(M-1)} \sum_{n=0}^{N-1} \sum_{i=2}^{M} \widehat{\lambda}_{i,n}$ and the signal part of the spectrum $\widehat{S}_{\mathbf{y},\mathcal{S}}(z_n) = \widehat{V}_{max,n}\, (\widehat{\lambda}_{max,n} - \widehat{\sigma}_v^2)\, \widehat{V}_{max,n}^H$.

3. Compute the acoustic channel correlations

$$\begin{bmatrix} \widehat{r}_{\mathbf{y}}(0) \\ \widehat{r}_{\mathbf{y}}(1) \\ \vdots \\ \widehat{r}_{\mathbf{y}}^H(1) \end{bmatrix} = \frac{1}{N} (F_N^* \otimes I_M) \begin{bmatrix} \widehat{S}_{\mathbf{y}}(z_0) \\ \widehat{S}_{\mathbf{y}}(z_1) \\ \vdots \\ \widehat{S}_{\mathbf{y}}(z_{N-1}) \end{bmatrix} \tag{II.22}$$

Put the correlations outside the range $n \in \{0, 1, \ldots, L_h-1\}$ to zero to obtain the Hermitian of the following block row
$[\widehat{r}_{\mathbf{y}}(0)\ \widehat{r}^H_{\mathbf{y}}(1) \cdots \widehat{r}^H_{\mathbf{y}}(L_h-1)\ 0\ \cdots\ 0\ \widehat{r}_{\mathbf{y}}(L_h-1) \cdots \widehat{r}_{\mathbf{y}}(1)]$.

4. Compute the spectrum of the thus windowed correlation sequence

$$\begin{bmatrix} \widehat{S}_{\mathbf{y}}(z_0) \\ \widehat{S}_{\mathbf{y}}(z_1) \\ \vdots \\ \widehat{S}_{\mathbf{y}}(z_{N-1}) \end{bmatrix} = (F_N \otimes I_M) \begin{bmatrix} \widehat{r}_{\mathbf{y}}(0) \\ \widehat{r}_{\mathbf{y}}(1) \\ \vdots \\ \widehat{r}^H_{\mathbf{y}}(1) \end{bmatrix} \tag{II.23}$$

Go back to step 2 untill convergence. Note that the IFFTs and FFTs in (II.22) and (II.23) can be carried out efficiently in Matlab by reshaping the $N \times 1$ vectors of $M \times M$ blocks into $N \times M^2$ matrices.

After convergence, we make use of the refined spectrum we get at step (4) to get an enhanced channel impulse response estimation within the framework described in the previous section.

## II.6 Experimental Results

We run our simulations within the framework of an SIMO-OFDM system where each OFDM symbol is composed of 128 tones. The performance of the different deterministic channel estimators (structured and non structured) are evaluated by means of the NMSE vs. SNR. The NMSE is defined as $\frac{||\mathbf{h}^0-\hat{\mathbf{h}}||^2}{||\mathbf{h}^0||^2}$ and the SNR is defined as $\frac{\sum_{k=0}^{N-1} ||\mathbf{h}_k||^2 \sigma^2_{x_k}}{\sigma^2_v MN}$. The symbols are drawn from QPSK constellation and the NMSE is averaged over 10000 Monte-Carlo runs of the noise, symbols and the channel. We consider Rayleigh fading channel realizations where each one is composed of five i.i.d. channel coefficients. It is worthy to note that for SNR less than 20 dB the algorithm always converges typically after three or four iterations while at higher SNR the convergence is guranteed at no more than ten iterations. However, we consider a convergence is achieved when the following condition is fulfilled: $\frac{\widehat{\sigma}^2_{v,i}-\widehat{\sigma}^2_{v,i-1}}{\widehat{\sigma}^2_{v,i}} \leq 0.1$ where i denotes the number of the current iteration at which the convergence is checked. Figure II.1 shows the performance of both SRM and NSF with and without structuring where three antennas have been utilized at the receiver and the sampled spectrum has been computed from just one OFDM symbol. We remark that SRM yields better performance than NSF. This is due to the fact that when we work with one OFDM symbol then SRM is a weighted version of NSF with the weight being the largest eigen value of the sampled spectrum at each tone. However, when structuring is used both estimators show at least 3 dB gain even at very high SNR. It is also obvious that after structuring the performance of both estimators is congruent whatever is the SNR. To elaborate more the advantage of our structuring algorithm we plot in Figure II.2 the BER versus SNR where we have used the estimated channels

by various algorithms to equalize the received signal using MMSE equalizer and a hard decision decoding to extract the received bits. This result shows that our algorithm outperforms the non-structured ones by more than 2 dB at BER $= 10^{-2}$.

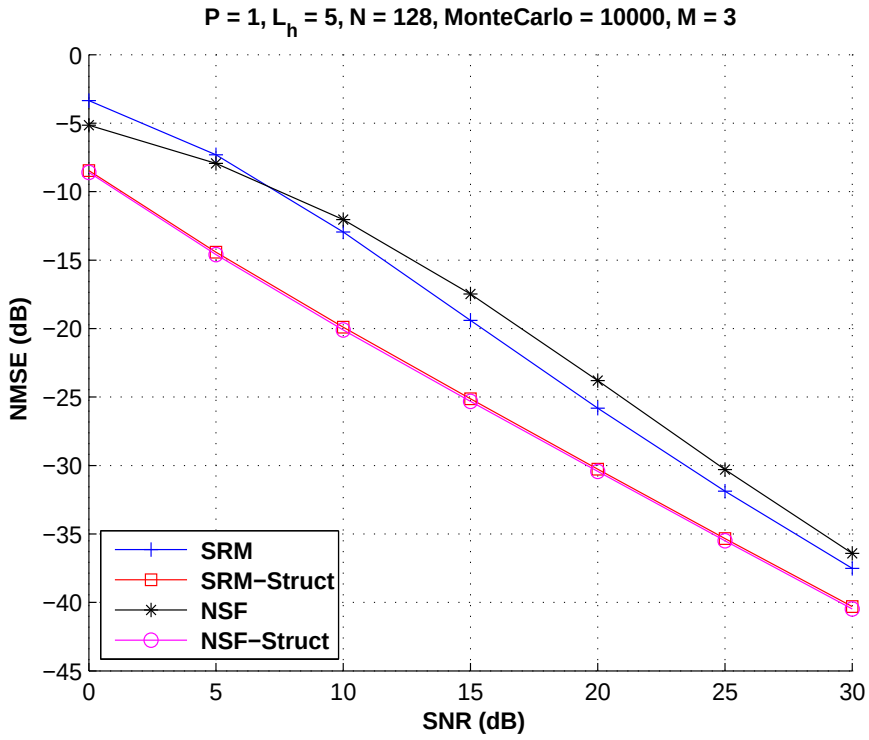


Figure II.1: The NMSE versus SNR for structured and non-structured estimators.

## II.7 Conclusions

To sum up, we have shown in this article the capabiltiy to exploit the classical blind deterministic channel estimators with a great computational power saving within the cyclic prefix systems. This is accomplished by minimizing the sum of the cost functions at different tones instead of minimizing the ordinary cost function in the time domain. Moreover, we propose a spatio-temporal based algorithm to enhance the sample covariance matrix upon which a class of well-known estimators rely. The enhancement is achieved by enforcing both the rank and the FIR structure properties. The simulations show that the proposed algorithm has the potential to provide a 5 dB gain (in terms of NMSE) at low

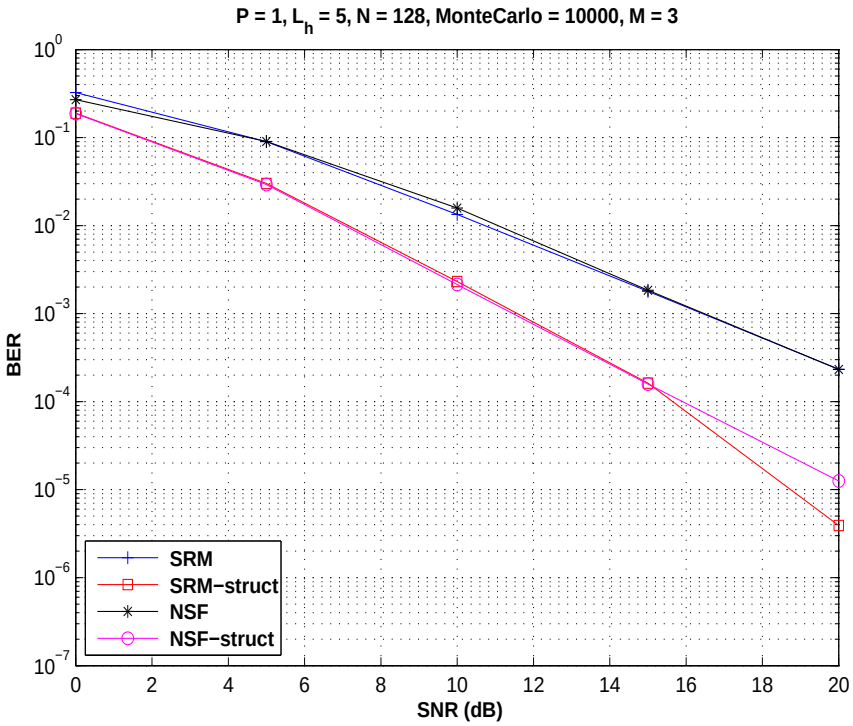


Figure II.2: The BER versus SNR for structured and non-structured estimators.

to moderate SNR while it still has the capabiltiy to provide a noticeable gain at high SNR.

# Bibliography


[1] D.T.M. Slock,; "Blind FIR Channel Estimation in Multichannel Cyclic Prefix Systems" in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, July 2004.

[2] J. A. Cadzow, "Signal Enhancement – A Composite Property Mapping Algorithm," *IEEE Trans. Acoust., Speech and Signal Processing*, Vol. 36, No. 1, Jan. 1988.

[3] D.T.M. Slock "From Sinusoids in Noise to Blind Deconvolution in Communications". In *Communications, Computation, Control and Signal Processing: a Tribute to Thomas Kailath*, A. Paulraj, V. Roychowdhury and C.D. Schaper, editors, Kluwer Academic Publishers, Boston, 1997.

[4] S.M. Omar, D.T.M. Slock "Singular Block Toeplitz Matrix Approximation and Application to Multi-Microphone Speech Dereverberation". *10th IEEE International Workshop on MultiMedia Signal Processing*, October 8-10, 2008, Cairns, Queensland, Australia.

[5] M.Gurelli and C.L.Nikias,"A New Eigenvector-Based Algorithm for Multichannel Blind Deconvolution of Input Colored Signals," *in Proc. ICASSP, 1993, pp.448-451.*

[6] L.A.Baccala and S.Roy, "A New Time-Domain Blind Identification Method," *IEEE Signal Processing Letters*,Vol. 1,no.6, pp.89-91, June 1994.

[7] G. Xu, H. Liu, L.Tong and T. Kailath "A Least Squares Approach to Blind Channel Identification," *IEEE Transactions on Signal Processing*,Vol. 43,no.12, pp.2982-2993, Dec 1995.

[8] Y. Hua "Fast Maximum Likelihood for Blind Identification of Multiple FIR Channels" *IEEE Transactions on Signal Processing*,Vol. 44,no.3, pp.661-672,March 1996.

[9] J.Ayadi, D.T.M. Slock,; "On Linear Channel-based Noise Subspace Parameterizations for Blind Multichannel Identification " in *Proc. IEEE Signal Processing Advances in Wireless Communications*,March 2001.

[10] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue,; "Subspace methods for the blind identification of multichannel FIR filters" in *IEEE Trans. Signal Processing,*,vol. 43, pp. 516525, Feb. 1995.