UNIVERSITAT POLITÈCNICA
DE CATALUNYA

Ph.D. Dissertation

# Generalized consistent estimation in arbitrarily high dimensional signal processing

Author: Francisco Rubio

Advisor: Dr. Xavier Mestre

Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)
Parc Mediterrani de la Tecnologia
Av. Canal Olímpic s/n
08860 Castelldefels (Barcelona), Spain

April 28, 2008

# Abstract

The theory of statistical signal processing finds a wide variety of applications in the fields of data communications, such as in channel estimation, equalization and symbol detection, and sensor array processing, as in beamforming, and radar systems. Indeed, a large number of these applications can be interpreted in terms of a parametric estimation problem, typically approached by a linear filtering operation acting upon a set of multidimensional observations. Moreover, in many cases, the underlying structure of the observable signals is linear in the parameter to be inferred. This dissertation is devoted to the design and evaluation of statistical signal processing methods under realistic implementation conditions encountered in practice.

Traditional statistical signal processing techniques intrinsically provide a good performance under the availability of a particularly high number of observations of fixed dimension. Indeed, the original optimality conditions cannot be theoretically guaranteed unless the number of samples increases asymptotically to infinity. Under this assumption, a statistical characterization can be often afforded by using the large-sample theory of sample covariance matrices. In practice, though, the application of these methods to the implementation of, for instance, training schemes in communication systems and adaptive procedures for radar detection problems, must rely on an observation window of finite length. Moreover, the dimension of the received signal samples (e.g. number of array sensors in multi-antenna systems) and the observations window size are most often comparable in magnitude. Under these situations, approaches based on the classical multivariate statistical analysis significantly lose efficiency or cannot even be applied. As a consequence, the performance of practical solutions in some real situations might turn out to be unacceptable.

In this dissertation, a theoretical framework for characterizing the efficiency loss incurred by classical multivariate statistical approaches in conventional signal processing applications under the practical conditions mentioned above is provided. Based on the theory of the spectral analysis of large-dimensional random matrices, or random matrix theory (RMT), a family of new statistical inference methods overcoming the limitations of traditional inferential schemes under comparably large sample-size and observation dimension is derived. Specifically, the new class of consistent estimators generalize conventional implementations by proving to be consistent even

for arbitrarily high-dimensional observations (i.e., for a limited number of samples per filtering degree-of-freedom).

In particular, the proposed theoretical framework is shown to properly characterize the performance of multi-antenna systems with training preambles in the more meaningful asymptotic regime defined by both sample size and dimension increasing without bound at the same rate. Moreover, the problem of optimum reduced-rank linear filtering is reviewed and extended to satisfy the previous generalized consistency definition. On the other hand, a double-limit asymptotic characterization of a set of vector-valued quadratic forms involving the negative powers of the observation covariance is provided that generalizes existing results on the limiting eigenvalue moments of the inverse Wishart distribution. Using these results, a new generalized consistent eigenspectrum estimator based on the inverse-shifted power method is derived that uniquely relies on the SCM and does not require matrix eigendecomposition. The effectiveness of the previous spectral estimator is demonstrated upon its application to the construction of an improved source power estimator that is robust to inaccuracies in the knowledge of both noise level and true covariance matrix.

In order to alleviate the computation complexity issue associated with practical implementations involving matrix inversions, a solution to the two previous problems is afforded in terms of the positive powers of the SCM. To that effect, a class of generalized consistent estimators of the covariance eigenspectrum and the power level are obtained on the Krylov subspace defined by the true covariance matrix and the signature vector associated with the intended parameter. In practice, filtering solutions are very often required to robustly operate not only under sample-size constraints but also under the availability of an imprecise knowledge of the signature vector. Finally, a signal-mismatch robust filtering architecture is proposed that is consistent in the doubly-asymptotic regime.

# Resum

La teoria del processat estadístic del senyal troba un ampli ventall d'aplicacions en els camps de les comunicacions de dades, com per exemple en els problemes d'estimació i equalització del canal, y en la detecció de símbols, així com també en el processat amb agrupacions de sensors, com per exemple en el problema de conformació de feix i en sistemes de radar. Certament, un gran nombre d'aquestes aplicacions poden ser interpretades com un problema d'estimació paramètrica, típicament resolt mitjançant una operació de filtrat lineal actuant sobre un conjunt d'observacions multidimensionals. A més, en molts casos, l'estructura subjacent dels senyals observables és lineal en el paràmetre a inferir. Aquesta dissertació està dedicada al disseny i avaluació de mètodes de processat estadístic del senyal en condicions d'implementació realistes trobades a la pràctica.

Les tècniques tradicionals de processat estadístic del senyal proporcionen un rendiment satisfactori donada la disponibilitat d'un nombre particularment elevat d'observacions de dimensió finita. En efecte, les condicions d'optimalitat originals no poden garantir-se en teoria a menys que el nombre de mostres disponibles augmenti de forma asimptòtica. En base a aquesta suposició, en ocasions es pot obtenir una caracterització estadística fent ús de la teoria de grans mostres per a matrius de covariança mostral. A la pràctica, no obstant, l'aplicació d'aquests mètodes a la implementació de, per exemple, esquemes d'entrenament en sistemes de comunicacions i procediments adaptatius per a problemes de detecció radar, ha de basar-se necessariament en una finestra d'observació de longitud finita. A més, la dimensió de les mostres rebudes (per exemple, nombre de sensors de l'agrupació en sistemes multi-antena), i el tamany de la finestra d'observació són sovint comparables en magnitud. En aquestes situacions, els plantejaments basats en l'anàlisi estadístic multivariant clàssic perden eficiència de forma significativa, o ni tan sols poden ésser aplicats. Com a conseqüència, la fiabilitat d'implementacions pràctiques en certes situacions reals pot resultar inacceptable.

En aquesta tesi es proporciona un marc teòric per a la caracterització de la pèrdua d'eficiència que els enfocs estadístics clàssics experimenten en aplicacions típiques del processat del senyal en les condicions pràctiques mencionades amb anterioritat. En base a la teoria de l'anàlisi espectral de matrius aleatòries de grans dimensions, o teoria de matrius aleatòries (RMT), es con-

strueix una família de mètodes d'inferència estadística que superen les limitacions dels esquemes d'estimació tradicionals per a un tamany de mostra i dimensió de la observació comparativament grans. Específicament, els estimadors de la nova classe obtinguda generalitzen les implementacions d'ús comú essent consistents fins i tot per a observacions amb dimensió arbitràriament gran (per exemple, per un nombre limitat de mostres per grau de llibertat de filtrat).

En particular, el marc teòric proposat és emprat per a caracteritzar de forma adequada el rendiment de sistemes multi-antena amb preàmbuls d'entrenament en un règim asimptòtic més coherent definit per un tamany i dimensió de les mostres que creixen sense límit amb raó constant. A més, el problema de filtrat òptim de rang reduït és revisat i extès de forma que es satisfaci la definició anterior de consistència generalitzada. Per altra banda, es proporciona una caracterització asimptòtica en el doble límit d'un conjunt de formes quadràtiques de les potències negatives de la covariància de l'observació que generalitza els resultats existents referents als moments negatius de la distribució de Wishart. Per mitjà d'aquests resultats, es deriva un estimador consistent generalitzat de l'espectre d'autovalors basat en el mètode de la potència inversa amb desplaçament que fa ús únicament de la matriu de covariància mostral i que no requereix descomposició de la matriu en valors singulars. L'efectivitat de l'estimador espectral anterior es demostra mitjançant la seva aplicació a la construcció d'un estimador de potència de font millorat que és robust a imprecisions en el coneixement del nivell de soroll i de la matriu de covariància real.

Amb el propòsit de reduir la complexitat computacional associada a implementacions pràctiques basades en la inversió de matrius, s'aborda una solució als problemes anteriors en termes de les potències positives de la matriu de covariança mostral. A tal efecte, s'obtenen una classe d'estimadors consistents generalitzats de l'espectre de la matriu de covariança i del nivell de potència en el subespai de Krylov definit per la covariànça real i el vector de signatura associat al paràmetre d'interès. A la pràctica, amb freqüència es requereixen solucions de filtrat robustes en front a no només restriccions en el tamany de la mostra, sino també a la disponibilitat d'un coneixement imprecís del vector de signatura. Com a contribució final, es proposa una arquitectura de filtrat robust a constriccions de la signatura que és consistent en el règim doblement asimptòtic de referència al llarg de la tesi.

# Resumen

La teoría del procesado estadístico de la señal halla un amplio abanico de aplicaciones en los campos de las comunicaciones de datos, como por ejemplo en los problemas de estimación y ecualización del canal, y en la detección de símbolos, así como también en el procesado con arreglos de sensores, como por ejemplo en el problema de conformación de haz y en sistemas de radar. Ciertamente, un gran número de estas aplicaciones pueden ser interpretadas como un problema de estimación paramétrica, típicamente resuelto mediante una operación de filtrado lineal actuando sobre un conjunto de observaciones multidimensionales. Además, en muchos casos, la estructura subyacente de las señales observables es lineal en el parámetro a inferir. Esta disertación está dedicada al diseño y evaluación de métodos de procesado estadístico de la señal en condiciones de implementación realistas encontradas en la práctica.

Las técnicas tradicionales de procesado estadístico de la señal proporcionan un rendimiento satisfactorio dada la disponibilidad de un número particularmente elevado de observaciones de dimensión finita. En efecto, las condiciones de optimalidad originales no pueden garantizarse en teoría a menos que el número de muestras disponibles aumente de forma asintótica. En base a esta suposición, en ocasiones se puede obtener una caracterización estadística haciendo uso de la teoría de grandes muestras para matrices de covarianza muestral. En la práctica, no obstante, la aplicación de estos métodos a la implementación de, por ejemplo, esquemas de entrenamiento en sistemas de comunicaciones y procedimientos adaptativos para problemas de detección radar, debe necesariamente basarse en una ventana de observación de longitud finita. Además, la dimensión de las muestras recibidas (por ejemplo, número de sensores del arreglo en sistemas multi-antena), y el tamaño de la ventana de observación son a menudo comparables en magnitud. En estas situaciones, los planteamientos basados en el análisis estadístico multivariante clásico pierden eficiencia de forma significativa, o ni siquiera pueden ser aplicados. Como consecuencia, la fiabilidad de implementaciones prácticas en ciertas situaciones reales puede resultar inaceptable.

En esta tesis se proporciona un marco teórico para la caracterización de la pérdida de eficiencia que los enfoques estadísticos clásicos experimentan en aplicaciones típicas del procesado de la señal en las condiciones prácticas mencionadas con anterioridad. En base a la teoría del

análisis espectral de matrices aleatorias de grandes dimensiones, o teoría de matrices aleatorias (RMT), se construye una familia de métodos de inferencia estadística que superan las limitaciones de los esquemas de estimación tradicionales para un tamaño de muestra y dimensión de la observación comparativamente grandes. Específicamente, los estimadores de la nueva clase obtenida generalizan las implementaciones al uso siendo consistentes incluso para observaciones con dimensión arbitrariamente grande (por ejemplo, para un número limitado de muestras por grado de libertad de filtrado).

En particular, el marco teórico propuesto es empleado para caracterizar de forma adecuada el rendimiento de sistemas multi-antena con preámbulos de entrenamiento en un régimen asintótico más acorde definido por un tamaño y dimensión de las muestras que crecen sin límite con razón constante. Además, el problema de filtrado óptimo de rango reducido es revisado y extendido de forma que se satisfaga la definición anterior de consistencia generalizada. Por otro parte, se proporciona una caracterización asintótica en el doble límite de un conjunto de formas cuadráticas de las potencias negativas de la covarianza de la observación que generaliza los resultados existentes referentes a los momentos negativos de la distribución de Wishart. Por medio de estos resultados, se deriva un estimador consistente generalizado del espectro de autovalores basado en el método de la potencia inversa con desplazamiento que hace uso únicamente de la matriz de covarianza muestral y que no requiere descomposición de la matriz en valores singulares. La efectividad del estimador espectral anterior se demuestra mediante su aplicación a la construcción de un estimador de potencia de fuente mejorado que es robusto a imprecisiones en el conocimiento del nivel de ruido y de la matriz de covarianza real.

Con el propósito de reducir la complejidad computacional asociada a implementaciones prácticas basadas en la inversión de matrices, se aborda una solución a los problemas anteriores en términos de las potencias positivas de la matriz de covarianza muestral. A tal efecto, se obtienen una clase de estimadores consistentes generalizados del espectro de la matriz de covarianza y del nivel de potencia en el subespacio de Krylov definido por la covarianza real y el vector de firma asociado al parámetro de interés. En la práctica, con frequencia se requieren soluciones de filtrado robustas frente no sólo a constricciones en el tamaño de la muestra, sino también a la disponibilidad de un conocimiento impreciso del vector de firma. Como contribución final, se propone una arquitectura de filtrado robusto a constricciones de la firma que es consistente en el régimen doblemente asintótico de referencia a lo largo de la tesis.

*For my dear parents and my near family,*

# Acknowledgements

During the last four and half years, many things happened for which I am sincerely thankful. Since it is not always easy to say thanks, I would like to use this page to express my gratitude.

I am deeply indebted to my Ph.D. advisor, Dr. Xavier Mestre, for he has allowed me to dedicate these years to work on a dissertation in exactly the subject, the type of problems and the methodology that I wished to. Definitely, without his assistance and support, this thesis would not appear by far in its present form.

I would also thank Prof. Michael Honig and Prof. Dongning Guo to kindly host me during my stay at Northwestern University, and for their remarkable accessibility and readiness to assistance at anytime. I left impressed by their close work with the students in the research group.

I wish to thank Carles Antón and Carles Fernández for agreeing to be the external reviewers of my thesis and for completing the burocratic paperwork in a record time.

There were many people in all this time that have directly or indirectly contributed to the final state of this dissertation by making my way to its completion easier and enjoyable. In order to keep this section from resembling a phone directory as well as avoid forgetting any name, I would prefer to thank them all together. In particular, I would like to thank people at the CTTC, for building the social club that it is now and was always thanks to the existing great atmosphere (at times too great for work!). I would like to stress the fact that the gratitude extends to those who still are in the center, those who are not anymore (all of them are somewhere else!), and those who were here for just a short period of time. Of course, special thanks should be conferred to some of these people, but I am pretty sure they already know who I mean. Moreover, I would also like to thank the group of guys I met at Northwestern University, whom I am indebted for making my stay in Evanston much more pleasant. Altogether, the names add up to many, so that I guess my decision is now more clear.

Antes de terminar, me gustaría expresar mi mayor gratitud a mis padres, Jesús y Loli, a mis hermanos, Jesús Jr., David y Ángel, y al resto de la "near family", Patricia, Álvaro y Daniel, por su apoyo incondicional y verdadero, desde lo lejos, durante todos estos años.

---

This dissertation was supported by the Predoctoral Fellowship Program of the Centre Tecnològic de Telecomunicacions de Catalunya and the Predoctoral FI Program of the Generalitat de Catalunya.

*There is nothing more practical than a good theory*
*(James Clerk Maxwell, 1831-1879),*

# Contents

# Notation

In general, uppercase boldface letters ($\mathbf{A}$) denote matrices, lowercase boldface letters ($\mathbf{a}$) denote (column) vectors and italics ($a$) denote scalars and generic non-commutative random variables. Calligraphic letters will denote sets, or subspaces with dimension given by a subscripts.

| | |
|---|---|
| $\mathbf{A}^T, \mathbf{A}^*, \mathbf{A}^H$ | Transpose, complex conjugate and Hermitian (i.e., complex conjugate transpose) of a matrix $\mathbf{A}$, respectively. |
| $\mathbf{A}^{-1}, \mathbf{A}^{\#}$ | Inverse and Moore-Penrose pseudoinverse of $\mathbf{A}$, respectively. |
| $\mathbf{A}^{1/2}$ | Positive definite Hermitian square-root of $\mathbf{A}$, i.e. $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$. |
| $\mathrm{Tr}\left[\mathbf{A}\right]$ | Trace of a matrix $\mathbf{A}$. |
| $\mathrm{Vec}\left[\mathbf{A}\right]$ | Column vector formed stacking the columns of $\mathbf{A}$ on top of one another. |
| $\rho\left(\mathbf{A}\right)$ | Spectral radius of a square $M \times M$ matrix $\mathbf{A}$ with eigenvalues $\lambda_m, m = 1, \ldots, M$, i.e. $\rho\left(\mathbf{A}\right) = \max_{1\leq m\leq M}\left(\lvert\lambda_m\rvert\right)$. |
| $\lVert\mathbf{A}\rVert, \lVert\mathbf{A}\rVert_{\mathrm{tr}}$ $\lVert\mathbf{A}\rVert_F, \lVert\mathbf{A}\rVert_W$ | Induced (spectral or strong), trace, Frobenius and weak norm, respectively, of a square $M \times M$ matrix $\mathbf{A}$, i.e. $\lVert\mathbf{A}\rVert = \left(\rho\left(\mathbf{A}^H\mathbf{A}\right)\right)^{1/2}$, $\lVert\mathbf{A}\rVert_{\mathrm{tr}} = \mathrm{Tr}\left[\left(\mathbf{A}^H\mathbf{A}\right)^{1/2}\right]$, $\lVert\mathbf{A}\rVert_F = \left(\mathrm{Tr}\left[\mathbf{A}^H\mathbf{A}\right]\right)^{1/2}$ and $\lVert\cdot\rVert_W = \frac{1}{M^{1/2}}\lVert\mathbf{A}\rVert_F$. |
| $\lVert\mathbf{a}\rVert$ | Euclidean norm of a vector $\mathbf{a}$, i.e. $\lVert\mathbf{a}\rVert = \left(\mathbf{a}^H\mathbf{a}\right)^{1/2}$. |
| $\lVert\mathbf{a}\rVert_{\mathsf{W}}$ | Weighted norm of $\mathbf{a}$, i.e. $\lVert\mathbf{a}\rVert_{\mathsf{W}} = \left(\mathbf{a}^H\mathbf{W}\mathbf{a}\right)^{1/2}$ (with Hermitian positive definite $\mathbf{W}$). |
| $\left[\mathbf{a}\right]_i$ | The $k$th entry of a vector $\mathbf{a}$ (corresponding to the $i$th row and the $j$th column). |
| $\left[\mathbf{A}\right]_{i,j}$ | The entry of a matrix in row $i$ of column $j$ of a matrix $\mathbf{A}$ (row and column indices begin at 1). |
| $\mathbf{A} \geq \mathbf{B}$, $\mathbf{A} > \mathbf{B}$ | The matrix $\mathbf{A} - \mathbf{B}$ is positive semidefinite and positive definite respectively. |
| $\mathbf{I}_M$ | The $M \times M$ identity matrix. |

| | |
|---|---|
| $\mathbf{0}_{M\times N}$ | An $M\times N$ matrix with all-zero entries. |
| $\mathsf{C},\mathsf{N},\mathsf{R},\mathsf{Z}$ | The set of all natural, integer, real and complex numbers, respectively. |
| $\mathsf{C}^M,\mathsf{R}^M$ | The set of $M$-dimensional vectors with entries in $\mathsf{C}$ and $\mathsf{R}$, respectively. |
| $\mathsf{C}^+$ | The set $\{z\in\mathsf{C}:\mathrm{Im}\{z\}>0\}$. |
| $\mathsf{R}^+$ | The set of all strictly positive real numbers. |
| $\mathsf{C}^{M\times N},\mathsf{R}^{M\times N}$ | The set of $M\times N$ matrices with entries in $\mathsf{C}$ and $\mathsf{R}$, respectively. |
| $\mathsf{C}[x],\mathsf{R}[x]$ | The set of polinomials in $x$ with coefficients in $\mathsf{C}$ and $\mathsf{R}$, respectively. |
| $\mathsf{C}[[x]],\mathsf{R}[[x]]$ | The set of formal power series in $x$ with coefficients in $\mathsf{C}$ and $\mathsf{R}$, respectively. |
| $\left[z^{-l}\right]\{f(z)\}$ | The operator extracting the coefficient of $z^{-l}$ in the series expansion of a $f(z)$. |
| $T(m,k)$ | The set of all $k$-tuples of natural numbers $(m_1,\ldots,m_k)$ satisfying $1m_1+2m_2+\ldots+km_k=k$ such that $m=m_1+\ldots+m_m$, with $m,k\in\mathsf{N}$. |
| $\mathrm{j}$ | Imaginary unit ($\mathrm{j}=\sqrt{-1}$). |
| $\mathrm{Re}\{\cdot\},\mathrm{Im}\{\cdot\}$ | Real and imaginary part, respectively. |
| $\lvert a\rvert,\mathrm{sign}(a)$ | Absolute value and sign of a real valued $a$. |
| $[a]^+$ | Maximum between $a$ (real valued) and zero. |
| $\Pr[A]$ | Probability of a certain event $A$. |
| $\mathsf{E}[\cdot]$ | Mathematical expectation. |
| $\mathsf{E}_{\mathsf{B}}[\cdot]$ | Mathematical expectation with respect to the statistics in $\mathbf{B}$. |
| $\mathsf{E}[\cdot\,\vert\mathcal{F}]$ | Conditional expectation given the $\sigma$-field $\mathcal{F}$. |
| $\#\{\cdot\}$ | Cardinality of a set. |
| $\log(\cdot)$ | Natural logarithm. |
| $\delta_0(\cdot)$ | Dirac delta. |
| $\delta_{m,n}$ | Kronecker delta. |
| $\mathcal{I}_\Omega$ | Indicator function over the set $\Omega$. |

| | |
|---|---|
| $\int_X g(x)\,\mu(dx)$ | Lebesgue integral of a function $g$ with respect to a measure $\mu$. |
| $\int_X g(x)dF(x)$ | Lebesgue-Stieltjes integral of a function $g$ with respect to a function $F$ of bounded variation. |
| $\stackrel{a.s.}{\to}$ | Convergence with probability one (or a.s. convergence). |
| $a \asymp b$ | Both quantities $a, b$ are asymptotic equivalents, i.e., $\lvert a-b\rvert \stackrel{a.s.}{\to} 0$. |
| sup | Supremum (lowest upper bound). If the set is finite, it coincides with the maximum (max). |
| $\mathrm{supp}\,(f)$ | Support of a probability density (nonnegative Lebesgue integrable) function $f$. |

# Acronyms

**a.s.** almost surely.
**AV** Auxiliary Vector.
**AWGN** Additive White Gaussian Noise.
**BER** Bit Error Rate.
**BLUE** Best Linear Unbiased Estimator.
**CDF** Cumulative Distribution Function.
**CDMA** Code Division Multiple Access.
**CG** Conjugate Gradient.
**CGM** Conjugate Gradient Method.
**CH** Cayley-Hamilton.
**CLT** Central Limit Theorem.
**CS** Cross-Spectral.
**CSI** Channel State Information.
**DCRCB** Doubly-Consistent Robust Capon Beamformer.
**DCRCB** Doubly-Constrained Robust Capon Beamformer.
**DL** Diagonal Loading.
**DoA** Direction of Arrival.
**DoF** Direction of Arrival.
**DS-CDMA** Direct Sequence Code Division Multiple Access.
**EVD** Eigenvalue Decomposition.
**GMRES** Generalized Minimal Residual Method.
**GSA** General Statistical Analysis.
**GSC** Generalized Sidelobe Canceller.
**GML** Gaussian Maximum Likelihood.
**i.i.d.** independent and identically distributed.
**LCMV** Linearly Constrained Minimum Variance.
**LHS** Left Hand Side.
**LMMSE** Linear Minimum Mean Square Error.
**MAI** Multiple Access Interference.
**MIMO** Multiple Input Multiple Output.

| | |
|---|---|
| **ML** | Maximum Likelihood. |
| **MOE** | Minimum Output Energy. |
| **MSE** | Mean Square-Error. |
| **MMSE** | Minimum Mean Squared Error. |
| **MSINR** | Maximum Signal-to-Interference Ratio. |
| **MSWF** | Multi-Stage Wiener Filter. |
| **MV** | Minimum Variance. |
| **MVUE** | Minimum Variance Unbiased Estimator. |
| **NCRCB** | Norm-Constrained Robust Capon Beamformer. |
| **PCA** | Principal Component Analysis. |
| **PCF** | Partial Cancellation Factor. |
| **PDF** | Probability Density Function. |
| **PIC** | Parallel Interference Cancellation. |
| **POR** | Power of $\mathbf{R}$. |
| **PPIC** | Partial Parallel Interference Cancellation. |
| **RCB** | Robust Capon Beamformer. |
| **RHS** | Right Hand Side. |
| **RMT** | Random Matrix Theory. |
| **RX** | Receiver. |
| **SCM** | Sample Covariance Matrix. |
| **SDM** | Steepest Descent Algorithm. |
| **SINR** | Signal to Interference plus Noise Ratio. |
| **SLLN** | Strong Law of the Large Numbers. |
| **SMI** | Sample Matrix Inversion. |
| **SNR** | Signal-to-Noise Ratio. |
| **SOI** | Signal of Interest. |
| **SOI** | Source of Interest. |
| **SOS** | Second-Order Statistics. |
| **SVD** | Singular Value Decomposition. |
| **TX** | Transmitter. |
| **ULA** | Uniform Linear Array. |
| **WBE** | Welch Bound Equality. |
| **w.r.t.** | with respect to. |

# FOREWORD and Research Motivation

Despite its long history and a solid evolution towards a mathematically well founded science, signal theory still face nowadays extremely important challenges, both in the scientific as well as in its application aspects. The theory of statistical signal processing finds a wide variety of applications in the fields of data communications, such as in channel estimation, equalization and symbol detection, and sensor array processing, as in beamforming as well as in radar and sonar systems. Alternatively, although not emphasized throughout this work, inferential methods from mathematical statistics are also extensively applied to signal processing problems in other disciplines such as astrophysics, biomedicine, seismology, and many other fields of interest for the scientific and engineering community. Indeed, a large number of these applications can be interpreted in terms of a parametric estimation problem, typically approached by a linear filtering operation acting upon a set of multidimensional observed samples. Moreover, in many cases, the underlying structure of the observable signals is linear in the parameter to be inferred. This dissertation is concerned with the general and certainly fundamental problem of discrete-time linear filtering of noisy signals aimed at the estimation of a linearly described unknown random parameter.

Most commonly applied estimation methods rely on the second-order statistics of the observed random vector process, particularly often via an inverse operation. In practice, the lack of true covariance information leads to implementations based on the empirical statistics of the received data samples. This fact immediately reveals the functional relevance of the sample covariance matrix (SCM) to practical estimation problems in signal processing. Unfortunately, two particular problems related to SCM-based implementations can be readily identified: the sample-support requirements and the computational complexity. In addition, these problems are quickly aggravated as the observation dimension becomes higher. Clearly, in cases requiring the covariance matrix to be inverted, an increasingly larger computational complexity is to be expected due to the inversion operation. Moreover, a particularly limited number of samples of relatively large dimension may especially contribute to a severe degradation of the estimation performance. In order to mitigate these limitations, a number of different schemes have been proposed in the engineering literature that can be essentially categorized into two broad families: diagonal loading regularization techniques and linear reduced-rank filtering methods. While the actual complexity problem has been increasingly relaxed over the past three decades due to the advent of modern computational methods and devices, despite its unquestionable interest in practice, little analytical insight can still be drawn from the broad literature about estimation problems characterized by the availability of a finite number of samples of arbitrarily high dimension.

In order to extend conventional approaches and review classical techniques under more general conditions, a significant effort has been recently placed by the mathematical statistics community on the study of high-dimensional data analysis methods as well as the optimal statistical inference under the more meaningful limiting regime defined by both the size of the data set and the data dimension going to infinity at a fixed rate (see [Bai05, Sri07, Joh07, Rao07] and references therein). Classical procedures are then obtained as special cases in the new framework (see also [Hal05] and cited work therein for high dimension, low sample-size data analysis and its application to statistical classification in genetics and medical imaging). Due to the relevance of the eigenvalue spectrum of the observation covariance matrix, of special interest for optimum signal processing are high-dimensional statistical data analysis and inferencial methods based on the theory of the spectral analysis of large-dimensional random matrices, or random matrix theory (RMT). In particular, this rather sophisticated branch of the mathematical theory of multivariate statistical analysis provides a characterization of the asymptotic behaviour of the eigenvalue spectrum of certain random matrix models.

In this dissertation, a RMT-based theoretical framework for characterizing the efficiency loss incurred by classical multivariate statistical approaches in conventional signal processing applications under the practical conditions mentioned above is provided. Of special interest are limiting results concerning SCM-type random matrices that are derived under the aforedefined double-asymptotic regime. While this class of results does allow for a limiting description in terms of only the eigenvalues, optimal signal processing solutions in practice usually depend as well upon the set of matrix eigenvectors. Therefore, an extension of some recently published results from random matrix theory is afforded in this thesis that is key for the theoretical characterization and further development of classical signal waveform and power estimation methods discussed throughout the dissertation. The limitations of current architectures has been analyzed and new robust and more convenient alternatives have been proposed, solving the structural limitations of traditional solutions. Specifically, a new class of generalized consistent estimators is introduced that allows for a considerably improved performance under a limited number of observations per filtering degree of freedom. In particular, the proposed estimators are derived such as to consistently represent an arbitrarily accurate approximation of the actual parameter as not only the number of samples but also the observation dimension increases without bound at the same rate. Consequently, the new constructions generalize traditional implementations that prove to be consistent only for an increasing sample size of strictly fixed dimension.

### Optimum Linear Filtering

A large number of signal processing applications can be interpreted in terms of a filtering problem on a set of multidimensional observations in order to extract a certain parameter of interest. In many of these applications, the underlying structure of the observations is linear in the parameter to be estimated. Under the assumption of a linear signal model, a number of estimators derived from different criteria and relying on the second-order statistics of the observed samples are identified as equivalent. For the sake of clarity of presentation, let us consider a collection of $N$ observations $\left\{\mathbf{y}(n) \in \mathsf{C}^M\right\}$ that are to be processed by a linear filter $\mathbf{w}$. The output of the filter can be expressed as $\mathbf{w}^H \mathbf{y}(n)$. In the following, we review the classical optimum solutions for the filter $\mathbf{w}$. In order to review optimum choices of $\mathbf{w}$, let us consider the following linear data model that properly defines the structure of a vast number of estimation problems in statistical signal processing, namely,

$$\mathbf{y}(n) = x(n)\,\mathbf{s} + \mathbf{n}(n)\,,$$

where $x(n)$ is the unknown parameter to be estimated, which is observed in noise with a given (unknown) interference $\mathbf{n}(n) \in \mathsf{C}^M$ after being operated upon by $\mathbf{s} \in \mathsf{C}^M$, a known *signature* vector associated with the parameter. A natural criterion to design a linear estimator of the desired parameter consists in trying to eliminate the undesired contribution from the observed data while keeping the desired component undistorted. This criterion can be formulated as an optimization problem in terms of a linear transformation acting on the covariance matrix of the observations, namely,

$$\arg \min_{\mathsf{w} \in \mathsf{C}^M} \mathbf{w}^H \mathbf{R} \mathbf{w} \ \text{ subject to } \mathbf{w}^H \mathbf{s} = \gamma,$$

where $\gamma \in \mathsf{C}$ is a fixed constant value. The optimal filter is found as the solution of a standard linearly constrained quadratic optimization problem, conventionally obtained by the method of Lagrange multipliers as

$$\mathbf{w}_{\mathsf{opt}} = \gamma^* \frac{\mathbf{R}^{-1}\mathbf{s}}{\mathbf{s}^H \mathbf{R}^{-1} \mathbf{s}},$$

where $(\cdot)^*$ denotes complex conjugation. The previous filtering structure is known in the engineering literature as the linearly constrained minimum variance (LCMV) filter. For the special case of a real constant $\gamma = 1$, it is also known as the minimum variance distortionless response (MVDR) filter. Formally, this intuitive criterion may be formulated from a statistical estimation perspective as the problem of constructing a linear estimator minimizing the mean square-error (MSE), i.e.,

$$\mathsf{MSE}(\hat{x}) = \mathsf{E}\left\{|x - \hat{x}|^2\right\} = \mathsf{var}(\hat{x}) + [\mathsf{bias}(\hat{x})]^2\,.$$

The problem of obtaining a linear transformation minimizing $\mathsf{MSE}(\hat{x})$ under the unbiasedness constraint is equivalent to the optimization problem above, and the result is usually referred to as the best linear unbiased estimator (BLUE). Interestingly enough, when the observations are Gaussian or have a linear model structure, the minimum variance unbiased estimator (MVUE) of the desired parameter turns out to be linear [Kay93]. Thus, in these cases both BLUE and MVUE are equivalent.

Next, we consider an alternative metric defined in terms of the ratio between the power of the desired and undesired components, respectively, at the output of the filter, and that is usually motivated from the application point of view. In the engineering literature, the signal of interest (SOI) and the noise are conventionally assumed to be independent and jointly distributed wide-sense stationary random processes, with mean zero and SOI power and noise covariance given, respectively, by $\mathsf{E}\left[x^{*}\left(n\right)x\left(n\right)\right]=\sigma_{x}^{2}\delta_{m,n}$ and $\mathsf{E}\left[\mathbf{n}\left(m\right)\mathbf{n}^{H}\left(n\right)\right]=\mathbf{R}_{N}\delta_{m,n}$. Then, from the above assumptions, the covariance matrix of the observations takes the form

$$\mathbf{R}=\mathsf{E}\left\{\left(\mathbf{y}-\mathsf{E}\left\{\mathbf{y}\right\}\right)\left(\mathbf{y}-\mathsf{E}\left\{\mathbf{y}\right\}\right)^{H}\right\}=\sigma_{x}^{2}\mathbf{s}\mathbf{s}^{H}+\mathbf{R}_{N}.$$

Then, we define the so-called signal-to-interference-plus-noise ratio (SINR) as

$$\mathsf{SINR}(\mathbf{w})=\frac{\sigma_{x}^{2}\left|\mathbf{w}^{H}\mathbf{s}\right|^{2}}{\mathbf{w}^{H}\mathbf{R}_{N}\mathbf{w}}=\left(\frac{\mathbf{w}^{H}\mathbf{R}\mathbf{w}}{\sigma_{x}^{2}\left|\mathbf{w}^{H}\mathbf{s}\right|^{2}}-1\right)^{-1},$$

which allows us to formulate the optimum estimator problem as

$$\mathbf{w}_{opt}=\arg\max_{\mathbf{w}}\mathsf{SINR}(\mathbf{w}).$$

Equivalently, all previous estimation solutions can be shown to maximize the SINR at the output of the discrete-time linear filter. It is worth noting that the linear transformation obtained from the Bayesian minimum mean squared error (MMSE) criterion, namely $\mathbf{w}_{\mathsf{LMMSE}}=\mathbf{R}^{-1}\mathbf{s}$, is also equivalent in terms of the SINR to the previous filtering solutions.

Before concluding our review, we consider the two fundamental applications of the previous signal model motivating the research work in this thesis: the problems of *signal waveform estimation* and *signal power estimation* in the contexts of symbol detection in wireless communications and spatial filtering in sensor array signal processing.

Classical examples of estimation problems in the engineering literature based on the second-order statistics of the received observations are those of multiuser symbol detection in code-division multiple access (CDMA) systems and the coherent reception of sources using an antenna array. In these two cases, the signature vector includes (possibly distorted by the channel versions of) the spreading sequence of the desired user in a CDMA application and the spatial steering vector of the intended source in array processing. Classical filtering solutions for linear multiuser detection [Ver98] are based on the knowledge of the spreading sequences of all users as well as information about their channels and the background noise level. In some scenarios, like for example in the downlink, it is unrealistic to assume that a particular user will have access to the information of the rest of the system. In these situations, the detection task is better approached from a point of view of multiple access interference (MAI) suppression [Hon95, Mad94]. The linear MMSE interference-suppresion receiver [Poo98] is equivalent to $\mathbf{w}_{\mathsf{LMMSE}} = \mathbf{R}^{-1}\mathbf{s}$, where $\mathbf{R}$ is the covariance matrix of the received observations. On the other hand, the MVDR beamformer is proportional to the column vector $\mathbf{w}_{\mathsf{MVDR}} = \mathbf{R}^{-1}\mathbf{s}$, regardless of whether the signal of interest is present or not in the received signal, where $\mathbf{R}$ is here the covariance matrix of the array observation.

Under general operation conditions, the implementation of the previous waveform estimators requires the knoledge of the received signal amplitude, or, equivalently, the SOI power. Alternatively, the analogous problem of estimating the amplitude of a number of received users or sources is also of special interest in the fields of wireless communications and array processing, as, for instance, in the computation of the SINR required for power control algorithms. If the knowledge of the signature vector associated to all received signals is available, maximum likelihood (ML) methods usually deliver rather accurate power estimate (see e.g. [Ott93]). If, as in many practical situations, only the signature vector associated with the intended user or signal is known, from the minimum variance filtering formulation above, the SOI power can be approximated by $\mathsf{E}\left[|\hat{x}(n)|^2\right] = \mathbf{w}^H\mathbf{R}\mathbf{w}$. In particular, in the array processing literature, the Capon SOI power estimate is defined as

$$\sigma^2_{\mathsf{CAPON}} = \frac{1}{\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}}.$$

Finally, a related problem in sensor array signal processing is the estimation of the density of the power angular spectrum, usually with the purpose of estimating the direction of arrival (DoA) of the different sources [Sto05]. Consider a (finite) record of $N$ spatial samples obtained with an array of $M$ sensors or antennas $\{y_m(n), n = 1, \ldots, N, m = 1, \ldots, M\}$ . A number of $K$ different sources are supposed to impinge in the antenna array from different directions. We are now concerned with the problem of Under the assumption of narrowband signals and linear array elements, the array observation $\mathbf{y}(n) = \left[y_1(n) \cdots y_M(n)\right]^T$ is additively decomposed as

$$\mathbf{y}(n) = \sum_{k=1}^{K} s_k(n)\,\mathbf{a}(\theta_k) + \mathbf{n}(n),$$

where $s_k(n)$ is the symbol transmitted by the $k$th source at the discrete-time instant $n$, $\mathbf{a}(\theta_k)$ is its spatial signature vector (also steering vector or array transfer vector) and $\mathbf{n}(n)$ is additive noise. Here, $\theta_k$ denotes the $k$th source's direction of arrival, which is the parameter of interest in this problem. In order to identify a possible arrival from a certain direction $\theta$, Capon's method chooses a spatial filter $\mathbf{h}$ such that the power at its output is minimized subject to the constraint that the frequency $\omega_k$ is passed undistorted, i.e.,

$$\arg\min_{\mathsf{h}\in\mathsf{C}^M} \mathbf{h}^H\mathbf{R}\mathbf{h} \quad \text{subject to}\, \mathbf{h}^H\mathbf{a}(\theta) = 1,$$

where $\mathbf{R}$ is the covariance matrix of the array observation $\mathbf{y}(n)$. The solution to the filter design problem is

$$\mathbf{h}_{\mathsf{opt}} = \frac{\mathbf{R}^{-1}\mathbf{a}\left(\theta\right)}{\mathbf{a}^{H}\left(\theta\right)\mathbf{R}^{-1}\mathbf{a}\left(\theta\right)}.$$

Then, the power at the filter output, i.e. the power of the input data centered at the spatial frequency parametrized by $\theta$, is $1/\mathbf{a}^{H}\left(\theta\right)\mathbf{R}^{-1}\mathbf{a}\left(\theta\right)$. Furthermore, assuming constant unit gain over the filter bandpass and zero outside, and denoting by $\beta$ the bandwidth of the filter $\mathbf{h}_{\mathsf{opt}}$, the value of the (spatial) power spectral density can be finally approximated by $1/\beta\mathbf{a}^{H}\left(\theta\right)\mathbf{R}^{-1}\mathbf{a}\left(\theta\right)$. Different criteria may be used to determine the value of the bandwidth. A choice of $\beta$ is proposed in [Lag86] that results in a normalization of the output power by the squared norm of the filter solution, i.e.,

$$\beta = \left\|\mathbf{h}\right\|^{2} = \frac{\mathbf{a}^{H}\left(\theta\right)\mathbf{R}^{-2}\mathbf{a}\left(\theta\right)}{\left[\mathbf{a}^{H}\left(\theta\right)\mathbf{R}^{-1}\mathbf{a}\left(\theta\right)\right]^{2}}.$$

Applying this normalization, the Capon DoA estimates are obtained as the locations of the $K$ largest peaks of the following spatial spectrum estimate

$$\mathsf{E}\left[\left|\mathbf{h}^{H}\mathbf{y}\left(n\right)\right|^{2}\right] = \frac{\mathbf{a}^{H}\left(\theta\right)\mathbf{R}^{-1}\mathbf{a}\left(\theta\right)}{\mathbf{a}^{H}\left(\theta\right)\mathbf{R}^{-2}\mathbf{a}\left(\theta\right)}.$$

After the previous review on the statistical signal theory of relevance for the dissertation, in the following we describe, chapter by chapter, the contributions of this thesis.

### Thesis Outline and Contribution

In **Chapter 1**, the technical background on the mathematical tools underlying the theoretical framework developed throughout the thesis is introduced. In particular, the Stieltjes transform method is revised, and a new result concerning the asymptotic convergence of the eigenvalues and eigenvectors of a class of general random matrix ensembles is introduced. As an example of the importance and practical interest of the study of the asymptotics of spectral functions of not only random matrix eingevalues but also the associated eigenvectors in an appropriate double-limit regime underlying the entire dissertation, an analytical characterization of the transient regime of a training-based multiple-input multiple-output (MIMO) system exploiting the full diversity order of an arbitrarily correlated MIMO fading channel via optimal beamforming and combining is presented. The focus is on practical scenarios where no channel state information is available at either transmitter or receiver and the length of the training phase is comparable in magnitude to the system size.

A related technical contribution has been published in:

- F. Rubio, D. Guo, M. Honig, X. Mestre, *On Optimal Training and Beamforming in Uncorrelated MIMO Systems with Feedback*, Conference on Information Sciences and Systems (CISS 2008), Princeton, NJ, USA, March 19-21, 2008.

Other contributions on the study of the performance analysis of wireless communications systems via RMT tools can be found in

- F. Rubio, X. Mestre, *Asymptotic performance of code-reference spatial filters for multicode DS/CDMA*, IEEE International Conference on Acoustics, Speech, and Signal Processing. Philadelphia (USA), March 18-23, 2005, and

- F. Rubio, X. Mestre, *A comparative study of different self-reference beamforming architectures for multicode DS/CDMA*, XI National Symposium of Radio Science. Poznan (Poland) April 7-8, 2005,

where the negative moments of the Marcenko-Pastur distribution are used to analytically characterize the output signal-to-interference-plus-noise ratio and the asymptotic covariance of three different code-reference spatial filters for multicode DS/CDMA systems, and

- F. Rubio, X. Mestre, *Semi-blind ML channel estimation for MC-CDMA systems with code-multiplexed pilots*, VI IEEE Workshop on Signal Processing Advances in Wireless Communications, New York (USA) June 5-8, 2005,

where the performance of an iterative ML channel estimator for MC-CDMA systems is addressed relying on results about the asymptotic distribution of the spectrum of large-dimensional Fourier matrices.

After an introductory theoretical analysis confirming the practical interest of the theoretical framework, an improved estimator of the optimal reduced-rank linear minimum mean-square error (MMSE) filter is derived in **Chapter 2** that is consistent for arbitrarily high-dimensional observations. The proposed implementation results from an enhanced approximation of the MMSE filter coefficients on the reduced-dimensional subspace. The new filter construction generalizes traditional filter realizations based on directly replacing the theoretical covariance matrix by its sample estimate, and being consistent when all dimensions in the model but the number of samples remain bounded.

Part of the results concerning the application of the proposed method can be found in:

- F. Rubio, X. Mestre, *On the design of practical reduced-rank DS-CDMA receivers*, 49th IEEE Globecom Conference, San Francisco, California (USA), 27 November-1 December, 2006.

- F. Rubio, X. Mestre, *Design of reduced-rank MVDR beamformers under finite sample-support*, 4th IEEE Workshop on Sensor Array and Multi-channel Processing (SAM 2006). Waltham, Massachusetts (USA), July 12-14, 2006.

- F. Rubio, X. Mestre, *Analysis of multi-stage receivers under finite sample-support*, 31st IEEE International Conference on Acoustics, Speech, and Signal Processing. Toulouse (France), May 14-19, 2006.

- F. Rubio, X. Mestre, *Consistent reduced-rank LMMSE estimation with a limited number of samples per observation dimension.* Submitted to IEEE Transactions on Signal Processing.

Traditional estimators based on the eigendecomposition of the sample covariance matrix are known to be particularly sensitive to finite sample-size situations. In order to illustrate the application of the proposed framework to common spectral estimation problems, an estimator of the eigenspectrum of the array observation covariance matrix is presented in **Chapter 3** that builds upon the well-known power method and allows for an improved performance in the low sample-size, relatively large observation dimension regime. A family of scalar estimation problems involving the negative powers of arbitrarily correlated Wishart matrices is then characterized and further exemplified through the fundamental problem of source power estimation in sensor array signal processing. In particular, an extension of the Capon method delivering remarkably accurate approximations provided that there is available a precise knowledge of the noise variance, is consistently approximated using the SCM in an asymptotic regime that allows to take into consideration a limited number of samples per array sensor.

This technical contribution has been submitted to:

- F. Rubio, X. Mestre, *On the estimation of the covariance eigenspectrum of array sample observations*, 5th IEEE Workshop on Sensor Array and Multi-channel Processing (SAM 2008).
- F. Rubio, X. Mestre, *On the Eigenspectrum Inference of Array Covariance Matrices and the Problem of Source Power Estimation.* In preparation.

In order to relax the computational complexity related to the previous class of estimators, the focus is turned in **Chapter 4** into more practically affordable solutions based on the positive powers of the unknown second-order statistics of the observed samples. In particular, the application of Krylov subspace methods to two fundamental problems in sensor array signal processing is considered, namely the problem of estimating the power of an intended source and the estimation of the principal eigenspace and dominant eigenmodes of a structured observation covariance matrix. Specifically, a generalized consistent estimation of a certain class of power functions is proposed that allows for an implementation of Krylov subspace methods achieving both a moderate computational complexity and an improved performance under limited sample-size and relatively large observation dimension.

Part of the contribution can be found in:

- F. Rubio, X. Mestre, *Improved consistent estimation in Krylov subspaces*, in Proc. Asilomar Conference on Signals, Systems and Computers 2007, Pacific Grove, CA, USA, Nov. 4-7, 2007.
- F. Rubio, X. Mestre, *Generalized consistent estimation on low-rank Krylov subspaces of arbitrarily high dimension.* Submitted to IEEE Transactions on Signal Processing.

Finally, in **Chapter 5**, the consistency of sample robust Capon beamforming (RCB) solutions that are constructed under signature-mismatch constraints from a set of received array observations is revised. Particular emphasis is placed on the class of robust filters heuristically modeling the adverse effects of practical finite sample-size conditions as due to an imperfect knowledge of the effective spatial signature. In contrast, and as in practice, a small sample-size relative to the array dimension is identified as the actual source of filter estimation errors under unknown second-order statistics. Accordingly, a new alternative approach to RCB design is proposed in this dissertation that explicitly addresses both the signature-mismatch problem and the limitations due to a finite sample-size.

This technical contribution has been submitted to:

- F. Rubio, X. Mestre, *Generalized consistent robust Capon beamforming for arbitrarily large arrays*. Submitted to 2008 European Signal Processing Conference (EUSIPCO-2008).
- F. Rubio, X. Mestre, *A Class of Doubly-Consistent Robust Capon Beamformers*. In preparation.

Some concluding remarks and topics for future work are provided in the **Afterword**, followed by two **Appendices** on some asymptotic convergence results of particular significance for the developments throughout the thesis and a precursor on the combinatorics of set partitions.

# Chapter 1

# Technical Background and Mathematical Tools

In this chapter, a brief exposition of the main mathematical techniques used in this thesis is given. After introducing the elements of the theory of the *spectral analysis of large-dimensional random matrices*, a review of the Stieltjes transform approach to the study of the asymptotic spectrum of a class of random matrix models is provided. Finally, Girko's general statistical analysis of large observations is shortly outlined. In particular, the theory of G-estimation allows us to derive estimators of certain functions of the eigenvalue spectrum of the theoretical covariance matrix that are given in terms of the spectrum of its sample estimate and are consistent in the double-limit regime considered throughout the dissertation.

## 1.1 Preliminaries on Random Matrix Theory

In order to motivate our fundamental exposition about the spectral properties of large dimensional random matrices, consider the $M \times M$ complex Hermitian matrix $\mathbf{M}$, with eigenvalues $\lambda_1(\mathbf{M}) \leq \lambda_2(\mathbf{M}) \leq \ldots \lambda_M(\mathbf{M})$ and associated eigenvectors $\mathbf{e}_m$, $m = 1, \ldots, M$. Because the entries of $\mathbf{M}$ are random variables, so are its eigenvalues and eigenvectors. A key element in the study of the spectrum of random matrices of increasing dimensions is the *empirical distribution function* of the eigenvalues, which is defined for the spectrum of $\mathbf{M}$ as

$$F_{\mathsf{M}}^{M}(\lambda) = \frac{1}{M} \sum_{m=1}^{M} \mathcal{I}_{(\lambda_m(\mathsf{M}) \leq \lambda)}. \tag{1.1}$$

Clearly, as defined in (1.1), $F_{\mathsf{M}}^{M}(\lambda)$ is a random (right-continuous nondecreasing) probability distribution function, possibly atomic, i.e., with discontinuities at discrete points. The major driving idea behind random matrix theory relies on the fact that, for a certain class of random matrix ensembles, as the dimension of the matrix increases without bound with a fixed aspect ratio, the empirical distribution function $F_{\mathsf{M}}^{M}(\lambda)$ converges almost surely towards a well-defined limiting probability distribution function (possibly defective)[1] with a compactly supported density.

In order to establish the *limiting spectral distribution* of a given random matrix ensemble, two approaches have been essentially reported in the RMT literature (see [Bai99] for a thorough review and also the monograph on the spectral analysis of large-dimensional random matrices [Bai06]). On the one hand, the *moment method* has been succesfully applied to the Wigner matrix ensemble, the sample covariance and $F$ matrices, and, more recently, to Toeplitz, Hankel and Markov matrices (see [Bai99, Section 2] for arguments in connection to Wigner, SCM and $F$ matrices, as well as [Bry06] for the results concerning Toeplitz, Hankel and Markov matrices). In the following, we discuss the more relevant and practically useful approach based on the *Stieltjes transform*.

**Definition 1** *(Stieltjes transform of probability measures) Let $G$ be a probability distribution function on $\mathsf{R}$. Then, the Stieltjtes transform of $G$ is defined as*

$$m_G(z) = \int_{\mathsf{R}} \frac{dG(\lambda)}{\lambda - z}, \qquad z \in \mathsf{C}^+. \tag{1.2}$$

Only compactly supported measures will be hereafter considered. References dealing with the Stieltjes transform of probability measures and its properties include [Akh61, Chapter 6][Akh65, Sections 3.1-2][Kre77, Appendix][Lax02, Chapter 32][Hia00, Chapter 3][Ger03] (see also [Hac07, Proposition 2.2]). In particular, the following properties hold:

(**P1**) *$m_G(z)$ is an analytic function on $C^+$. Moreover, $m_G(\mathsf{C}^+) \subset \mathsf{C}^+$, $\overline{m_G(z)} = m_G(\bar{z})$, and* $\lim_{y\to\infty} -\mathrm{j}\, y m_G(\mathrm{j}\, y) = 1$.

See e.g. [Rud87, Theorem 10.7] for a proof.

(**P2**) *$m_G(z)$ satifies*

$$|m_G(z)| \leq \frac{1}{\mathrm{Im}\{z\}}, \tag{1.3}$$

$$\mathrm{Im}\{m_G(z)\} > 0. \tag{1.4}$$

Let us write $z = x + \mathrm{j}\, y$. Indeed, regarding (1.3), note that

$$\left|\frac{1}{\lambda - z}\right| \leq \frac{1}{\sqrt{(\lambda - x)^2 + y}} \leq \left|\frac{1}{y}\right|,$$

[1] I.e., with total variation less than one. Convergence of such sub-probability measures will be regarded as *vague* convergence. Alternatively, we will talk about *weak* convergence of proper probability measures and convergence in *distribution* when regarded to a random variable.

and, accordingly,

$$|m_G(z)| \le \frac{1}{y}\int_{\mathsf{R}} dG(\lambda) \le \frac{1}{y}.$$

Moreover, on the other hand, observe that

$$\begin{aligned} y &= \int_{\mathsf{R}} \operatorname{Im}\left\{\frac{1}{\lambda - z}\right\} dG(\lambda) \\ &= \int_{\mathsf{R}} \operatorname{Im}\left\{\frac{\lambda - z^*}{|\lambda - z|^2}\right\} dG(\lambda) \\ &= \int_{\mathsf{R}} \frac{y}{(\lambda - x)^2 + y} dG(\lambda), \end{aligned}$$

which is strictly positive, as (1.4) states.

(**P3**) *(Probability measures over* $\mathsf{R}^+$*) If* $G(0) = 0$*, then* $m_G(z)$ *is analytic over* $\mathsf{C} - \mathsf{R}^+$*. Moreover, if* $z \in \mathsf{C}^+$ *then* $z m_G(z) \in \mathsf{C}^+$*, and the following holds [Hac07, Proposition 2.2]:*

$$|m_G(z)| \le \begin{cases} \frac{1}{|\operatorname{Im}\{z\}|} & if\ z \in \mathsf{C} - \mathsf{R} \\ \frac{1}{|z|} & if\ z \in (-\infty, 0) \\ \frac{1}{\operatorname{dist}(z, \mathsf{R}^+)} & if\ z \in \mathsf{C} - \mathsf{R}^+, \end{cases}$$

*where* dist *stands for Euclidean distance.*

(**P4**) *Let* $m(z)$ *be a function satisfying (P1). Then, there is a probability distribution* $H$ *function such that* $m(z)$ *is its Stieltjes transform. Moreover, if, as in (P3),* $z m(z) \in \mathsf{C}^+$ *for* $z \in \mathsf{C}^+$*, then* $H(0) = 0$*, and* $m(z)$ *has an analytic continuation on* $\mathsf{C} - \mathsf{R}^+$*.*

See [Kre77, Appendix].

(**P5**) *(Stieltjes inversion formula) For any continuity points* $a < b$ *of* $G$*, we have*

$$G(b) - G(a) = \lim_{y \to 0^+} \frac{1}{\pi} \int_a^b \operatorname{Im}\{m_G(x + \mathrm{j}\, y)\}\ dx. \tag{1.5}$$

To see (1.5), note first that

$$\begin{aligned} \int_a^b \frac{1}{\pi} \operatorname{Im}\{m_G(x + \mathrm{j}\, y)\}\ dx &= \frac{1}{\pi}\int_a^b \int_{\mathsf{R}} \frac{y}{(u - x)^2 + y^2} dG(u)\, dx & (1.6) \\ &= \frac{1}{\pi}\int_{\mathsf{R}} \int_a^b \frac{y}{(u - x)^2 + y^2} dx dG(u) & (1.7) \\ &= \frac{1}{\pi}\int_{\mathsf{R}} \left(\arctan\left(\frac{b - x}{y}\right) - \arctan\left(\frac{a - x}{y}\right)\right) dG(u), & (1.8) \end{aligned}$$

where (1.6) is showed in (**P2**), the equality in (1.7) follows from Fubini's theorem and the integral in (1.8) results from the fact $\int \frac{dx}{1+x^2} = \arctan(x)$. Then, since the integrand in (1.8) is bounded, from the dominated convergence theorem we have

$$\lim_{y \to 0^+} \left(\arctan\left(\frac{b - x}{y}\right) - \arctan\left(\frac{a - x}{y}\right)\right) = \pi \mathcal{I}_{[a,b]}, \tag{1.9}$$

for any two continuity points $a < b$ of $G$. Therefore, using (1.9), we indeed have finally

$$\lim_{y \to 0^+} \int_a^b \frac{1}{\pi} \operatorname{Im} \{m_G (x + \mathrm{j}\, y)\} \; dx = \int_a^b dG(u) = G(b) - G(a).$$

The convenience of using the Stieltjes transform in order to study the limiting behavior of a matrix eigenvalue spectrum is evident from the following property, namely establishing a method to prove convergence of the empirical spectral distribution from the convergence of the associated Stieltjes transform. More specifically,

(**P6**) *If $\{G_n\}$ is a sequence of probability distribution functions and the sequence of associated Stieltjes transforms $\{m_{G_n}(z)\}$ converges pointwise to $m(z)$ for $z \in \mathsf{C}^+$, then there exists a probability distribution function $G$ with Stieltjes transform $m_G = m$, such that $\{G_n\}$ converges weakly to $G$. More generally,*

$$\{G_n\} \to G \Leftrightarrow \{m_{G_n}(z)\} \to m_G.$$

Indeed, the sufficiency part follows from the Helly-Bray theorem (see e.g. [Rao73, Section 2c.4]), whereas the necessary condition can be argued using Helly's selection theorem (see e.g. [Chu01, Theorem 4.3.3]), by extracting a subsequence of $\{G_n\}$ converging to $G$ with an associated Stieltjes transform that is, by (**P4**) and (**P5**), in one-to-one correspondance with $m_G(z)$ (see for instance [Ger03, Theorem 1]).

Thus, as weak convergence of probability measures can be established from their Fourier transform via Lévy's convergence theorem, the limiting behavior of the empirical spectral distribution of a matrix with increasing dimensions can be equivalently determined by its Stieltjes transform representation. An important advantage of using the latter is that the eigenvalue density function can be easily recovered from the Stieltjes transform of the distribution function. In particular,

(**P7**) [Sil95c, Theorem 1.1] *If, for $x_o \in \mathsf{R}$, the limit $\lim_{z \in \mathsf{C}^+ : z \to \infty} \operatorname{Im}\{m_G(z)\}$ exists and is denoted as $\operatorname{Im}\{m_G(x_o)\}$, then $G$ is differentiable at $x_o$ and its derivative is equal to $\frac{1}{\pi}\operatorname{Im}\{m_G(x_o)\}$, i.e.,*

$$\frac{dG(x)}{dx} = \lim_{y \to 0^+} \frac{1}{\pi} \operatorname{Im}\{m_G(x + \mathrm{j}\, y)\}. \tag{1.10}$$

The fundamental connection between the Stieltjes transform and the spectrum of random matrices is now apparent if we consider as the probability distribution function the empirical distribution function of the eigenvalues of a certain random matrix ensamble. Furthermore, the convenience of using the Stieltjes transform in order to study the limiting behavior of a matrix eigenvalue spectrum essentially relies on its representation in terms of the matrix resolvent, namely, for the matrix $\mathbf{M}$, $\mathbf{Q}(z) = (\mathbf{M} - z\mathbf{I}_M)^{-1}$. In particular, from the spectral theorem for

Hermitian matrices, we can write

$$\frac{1}{M}\operatorname{Tr}\left[\mathbf{Q}(z)\right] = \frac{1}{M}\operatorname{Tr}\left[\left(\mathbf{M} - z\mathbf{I}_M\right)^{-1}\right] = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{\lambda_m(\mathbf{M}) - z} = \int_{\mathsf{R}}\frac{dF_{\mathsf{M}}^{M}(\lambda)}{\lambda - z},$$

which is the Stieltjes transform of $F_{\mathsf{M}}^{M}$, i.e., $m_{F_{\mathsf{M}}^{M}}(z)$.

Consider further the eigenvalue moments of the spectral distribution of $\mathbf{M}$, namely,

$$M_k^{\mathsf{M}} = \int_{\mathsf{R}}\lambda^k dF_{\mathsf{M}}^{M}(\lambda) = \frac{1}{M}\operatorname{Tr}\left[\mathbf{M}^k\right], \quad k = 1, 2, \ldots. \tag{1.11}$$

The Stieltjes transform representation of a spectral distribution function can also be used to obtain the eigenvalue moments of the matrix ensamble without the need for integration. Indeed, observe that

$$M_k^{\mathsf{M}} = -\frac{1}{k!}\lim_{z\to 0}\frac{\partial^k}{\partial z^k}\left\{z^{-1}m_{F_{\mathsf{M}}^{M}}\left(z^{-1}\right)\right\}.$$

Hence, as for empirical spectral distributions, the asymptotic convergence of the eigenvalue moments can be established from the limit of the sequence of Stieltjes transforms for an ever increasing $M$. In particular, in Chapter 4, the Stieltjes transform is applied as a moment generating function in order to obtain the asymptotic eigenvalue moments of SCM-type matrices with outer correlations (cf. Appendix A). Next, a review of the Stieltjes transform representation as a power series expansion is provided that can help in answering a fundamental question in the context of the study of the asymptotic behavior of matrix spectral distributions: Is it possible to find the limiting empirical distribution function of sums and products of matrix ensembles in terms of their individual asymptotic spectrum?

### 1.1.1 Power series representation of Stieltjes transforms

Consider a spectral distribution function $G$, with Stieltjes transform $m_G(z)$. Since

$$\frac{1}{z-\lambda} = \frac{1}{z}\sum_{k=0}^{\infty}\left(\frac{\lambda}{z}\right)^k,$$

by the linearity of the integral, we can write the Stieltjes transform of $G$ as

$$m_G(z) = -\sum_{k=0}^{\infty}\frac{1}{z^{k+1}}\int_{\mathsf{R}}\lambda^k dG(\lambda),$$

which is the Laurent series expansion of the integrand (or Taylor series about $\infty$). Note that due to the compactness assumption on the eigenvalue density, the power series expansion of the Stieltjes transform is guaranteed to be holomorphic at infinity. Let $M_k^G$ be, with some abuse of notation, the $k$th eigenvalue moment of $G$. Then,

$$m_G(z) = -\frac{1}{z}\sum_{k=0}^{\infty}\frac{M_k^G}{z^k}. \tag{1.12}$$

The Laurent series $m_G(z)$ has an inverse for composition, say $k_G(z)$, with an expansion [Voi92, Remark 3.3.3]

$$k_G(z) = -\frac{1}{z} - \sum_{k=0}^{\infty} R_{k+1}^G \, z^k. \tag{1.13}$$

Analogously to the classical moment-cumulant problem, the series expansion coefficients $R_k^G$ are usually regarded as the cumulants associated with the spectral distribution function $G$. From their correspondance through a compositional inverse operation, the moments and cumulants of a spectral distribution can be found in terms of each other via the Lagrange inversion formula (cf. Appendix B). In Chapter 5, the relation between eigenvalue moments and cumulants is exploited in order to characterize the limiting behavior of a vector-valued defined empirical distribution of the eigenvalues and associated eigensubspaces of SCM-type matrices (see discussion in Appendix B on the appropriate combinatorial framework).

Consider now two independent random matrix ensembles, namly obtained from $M$-dimensional complex Hermitian matrices $\mathbf{X}$ and $\mathbf{Y}$, with empirical eigenvalue distribution functions $F_{\mathsf{X}}^M$ and $F_{\mathsf{Y}}^M$, respectively. We further assume that, as $M \to \infty$, with probability one, $F_{\mathsf{X}}^M \to F_{\mathsf{X}}$ and $F_{\mathsf{Y}}^M \to F_{\mathsf{Y}}$, where $F_{\mathsf{X}}$ and $F_{\mathsf{Y}}$ are proper probability distribution functions with compactly supported density. Suppose we are interested in the asymptotic spectrum of the ensembles $\mathbf{X}+\mathbf{Y}$ and $\mathbf{XY}$. In the case of classical (commutative) random variables, a closed-form expression of the probability density function of the sum and the product of two ensembles is known to exist, and is namely given for two absolutely continuous random variables $X$ and $Y$ by [Roh76]

$$f_{X+Y}(s) = \int_{-\infty}^{+\infty} \int_{-\infty}^{s-x} f_{X,Y}(x,y)\, dy dx, \tag{1.14}$$

and

$$f_{XY}(s) = \int_{-\infty}^{+\infty} f_{X,Y}\left(x, \frac{s}{x}\right) \frac{1}{|x|} dx, \tag{1.15}$$

respectively, where $f_X$, $f_Y$, $f_{X+Y}$ and $f_{XY}$ are the probability density function of $X$, $Y$, $X+Y$ and $XY$, and $f_{X,Y}$ is the joint probability density function of the random variables. If $X$ and $Y$ are independent, the expresion in (1.14) reduces to the convolution of the corresponding densities. Moreover, under the condition of independence, the result in (1.15) can be implemented via the Mellin transform, since the transform of the product of densities is the product of density transforms [Spr79].

The previous results are not applicable in the context of spectral distributions, since the algebra of Hermitian matrices gives rise to a non-commutative probability space. However, consider the power series expansion

$$r_G(z) = \sum_{k=0}^{\infty} R_{k+1}^G \, z^k, \tag{1.16}$$

where we have used the same notation as above. The series expansion in (1.16) is regarded as the *R-transform* in the literature of free probability theory [Voi92], and is related to the Stieltjes transform as

$$r_G(z) = m_G^{\langle -1 \rangle}(-z) - z^{-1}. \tag{1.17}$$

Interestingly enough, it turns out that, under certain conditions on the matrix ensembles (such as i.i.d. Gaussian matrix elements and unitarily invariance) the limiting spectral distribution of the sum of two matrices can be found from the sum of the R-transforms of the individual ensembles. In particular, if $F_{\mathsf{X}}$ and $F_{\mathsf{Y}}$ satisfy the kind of conditions mentioned above[2], we have

$$r_{F_{\mathsf{X}+\mathsf{Y}}}(z) = r_{F_{\mathsf{X}}}(z) + r_{F_{\mathsf{Y}}}(z),$$

where $F_{\mathsf{X}+\mathsf{Y}}$ is the limiting distribution function of the eigenvalues of $\mathbf{X}+\mathbf{Y}$. Consequently, from (1.17), the asymptotic spectral distribution can be obtained by first recovering its Stieltjes transform via

$$m_{F_{\mathsf{X}+\mathsf{Y}}}\left(r_G(-z) - z^{-1}\right) = z, \tag{1.18}$$

and then applying the Stieltjes inversion formula in (**P5**).

Consider again the distribution function $G$ and the following formal power series, namely,

$$\chi_G(z) = \sum_{k=1}^{\infty} M_k^G \, z^k. \tag{1.19}$$

In particular, observe that (1.19) can be found in terms of the Stieltjes transform of $G$ as

$$\chi_G(z) = -z^{-1} m_G\left(-z^{-1}\right) - 1. \tag{1.20}$$

Additionally, consider further the series expansion

$$s_G(z) = \frac{1+z}{z}\chi_G^{\langle -1 \rangle}(z), \tag{1.21}$$

which is regarded as the *S-transform* in the context of free probability [Voi87]. Analogously to the R-transform and the addition of matrix spectra, the limiting spectral distribution of the product of two random matrix ensembles can be directly obtained, under the conditions mentioned above, from the product of the S-transforms of the limiting individual spectral distribution. More precisely, under the previous conditions,

$$s_{F_{\mathsf{X}\mathsf{Y}}}(z) = s_{F_{\mathsf{X}}}(z)\, s_{F_{\mathsf{Y}}}(z),$$

[2] Namely, if the limiting empirical spectral distributions $F_{\mathsf{X}}$ and $F_{\mathsf{Y}}$ define two mutually *free* (non-commutative) random variables, where the notion of *freeness* can be thought of as a definition of statistical independence in non-commutative probability spaces. Equivalently, if $F_{\mathsf{X}}^M$ and $F_{\mathsf{Y}}^M$ define two random variables, $\mathsf{X}$ and $\mathsf{Y}$, that are *asymptotically free*. See e.g. [Tul04] for examples of free random matrix ensembles appearing in engineering applications.

where $F_{\mathsf{XY}}$ is the limiting distribution function of the eigenvalues of $\mathbf{XY}$. Thus, the asymptotic spectral distribution can be obtained, as before, by first recovering its Stieltjes transform from (1.21) and (1.20), and finally using the Stieltjes inversion formula.

In principle, the limiting spectral distribution of random ensembles resulting from linear operations on matrix models with arbitrary known limiting spectra can be obtained using the R-transform and the S-transform as outlined above. However, operative problems may quickly arise when calculating the inverse functions in (1.17), (1.18) and (1.21), as a closed-form expression may be in many situations not possibly obtained. Moreover, the statistical assumptions on the matrix ensemble definition that must be satisfied in order for the conditions on the individual limiting distributions to be fulfilled may represents a practical limitation. Finally, still more restrictive is the fact that an expression of the limiting Stieltjes transform is assumed to be available for finding the R- and S-transforms.

In the application of RMT to engineering problems in signal processing and wireless communications, the analytical description of the asymptotic spectrum of a certain random matrix models of much practical interest can be usually afforded by means of solely *Stieltjes transform methods*. Accordingly, on the one hand, the Gaussian distributional assumption as well as the existence and availability of an explicit closed-form expression of the Stieltjes transform of the matrix ensemble can be relaxed. On the other hand, the asymptotic convergence analysis of the ensemble spectrum can be extended to a more general class of spectral functions including also the associated eigensubspaces. Clearly, by considering as well the convergence and asymptotic behavior of the matrix eigenvectors, the range of application of the obtained results is expanded to cover a broader family of statistical signal processing problems.

In the following, we provide a brief compilation of known existing results obtained via the Stieltjes transform based approach about the asymptotic eigenvalue spectrum of some random matrix models.

### 1.1.2 Examples of random matrix ensembles

Throughout the next results, the sequences of random matrices are assumed to be defined on a common probability space. Moreover, the matrix $\Xi$ will denote an $M \times N$ complex random matrix, such that the real and imaginary parts of the entries are i.i.d. random variables with mean zero, variance 1/2 and bounded moments, and $\mathbf{X} = N^{-1/2}\Xi$. Furthermore, define $c$ , $\lim_{N\to\infty} M/N$ and $\beta = c^{-1}$.

**Theorem 1.1** *[Yin86][Sil95a, Theorem 1.1] (Sample covariance matrix with outer correlations) Let* $\mathbf{R}$ *be a* $M \times M$ *Hermitian non-negative definite matrix, whose eigenvalues are uniformly bounded for all* $M$ *and have an empirical distribution function that converges almost surely,* *as $M \to \infty$, to a nonrandom distribution function $H$. Define $\mathbf{B} = \mathbf{R}^{1/2}\mathbf{X}\mathbf{X}^H\mathbf{R}^{1/2}$ with $\mathbf{R}^{1/2}$ denoting any Hermitian square-root of the matrix $\mathbf{R}$. Then, as $M, N \to \infty$, with $c < +\infty$, almost surely, the empirical distribution function of $\mathbf{B}$, say $F_{\mathsf{B}}^M$, converges weakly to the distribution function $F_{\mathsf{B}}$ with associated Stieltjes transform $m_F(z)$ such that, for each $z \in \mathsf{C}^+$, $m = m_F(z)$ is the unique solution in the set $\{m \in \mathsf{C} : -(1-c)/z + cm \in \mathsf{C}^+\}$ to the following equation, namely,*

$$m = \int \frac{dH(\lambda)}{\lambda(1 - c - czm) - z}. \tag{1.22}$$

In particular, observe that the random matrix $\mathbf{B}$ resembles the sample estimate of the covariance matrix of a collection of multidimensional observations. Indeed, let $\{\mathbf{y}(n)\}$, $n = 1, \ldots, N$, be a sequence of i.i.d. $M$-dimensional vector samples with mean zero and covariance $\mathsf{E}\left[\mathbf{y}(m)\mathbf{y}^H(n)\right] = \mathbf{R}\delta_{m,n}$. The minimum variance unbiased estimator of the theoretical covariance matrix $\mathbf{R}$ is the sample covariance matrix, namely [And03, Mui82]

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{y}(n)\mathbf{y}^H(n) \equiv \frac{1}{N}\mathbf{Y}\mathbf{Y}^H, \tag{1.23}$$

where we have defined $\mathbf{Y} = \left[\mathbf{y}(1)\ldots\mathbf{y}(N)\right] \in \mathsf{C}^{M\times N}$. Furthermore, note that we can write the $n$th sample observation statistically equivalently as $\mathbf{y}(n) = \mathbf{R}^{1/2}\mathbf{\Xi}$, so that $\hat{\mathbf{R}} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H = \frac{1}{N}\mathbf{R}^{1/2}\mathbf{\Xi}\mathbf{\Xi}^H\mathbf{R}^{1/2} = \mathbf{R}^{1/2}\mathbf{X}\mathbf{X}^H\mathbf{R}^{1/2} = \mathbf{B}$. If the entries of $\mathbf{\Xi}$ are normally distributed (in which case $\hat{\mathbf{R}}$ is also the ML estimator of $\mathbf{R}$), then we have $N^{-1}\hat{\mathbf{R}} \sim \mathcal{W}_M(N, \mathbf{R})$, i.e., the (normalized) SCM $N^{-1}\hat{\mathbf{R}}$ follows a central Wishart distribution with N degrees of freedom and mean $N\mathbf{R}$.

**Example 1** *(Sample covariance matrix with outer correlations) $\mathbf{R}$ has dimension $M = 40$ eigenvalues 1, 2.5, 4 and 4.5 with same multiplicity. The number of samples to construct $\hat{\mathbf{R}}$ is $N = 400$. In Figure 1.1, the empirical histogram and the theoretical limiting spectral density are compared.*

As a special case of the result in Theorem (1.1), consider the standard (uncorrelated) central Wishart matrix, i.e., $\mathbf{R} = \mathbf{I}_M$. Then, the integral in (1.22) yields

$$m = \frac{1}{\lambda(1 - c - czm) - z},$$

so that the Stieltjes transform can be obtained as the solution of the following canonical equation, namely,

$$czm^2 - (1 - c - z)m + 1 = 0,$$

which is a second-degree polynomial in $m$, with coefficients being polynomials in $z$ and whose solution in $\mathsf{C}^+$ is given as a function of $z$ by

$$m = \frac{1 - c - z + \sqrt{(z - 1 - c)^2 - 4c}}{2cz}. \tag{1.24}$$

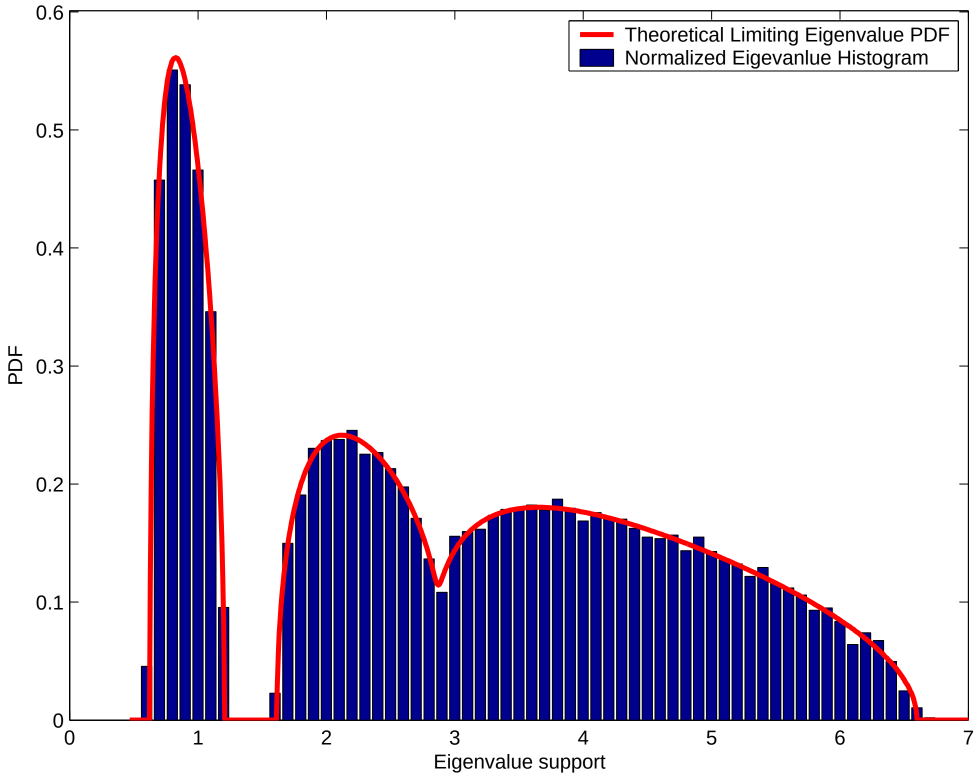


Figure 1.1: Limiting spectral density and normalized histogram of $\hat{\mathbf{R}}$ in Example 1

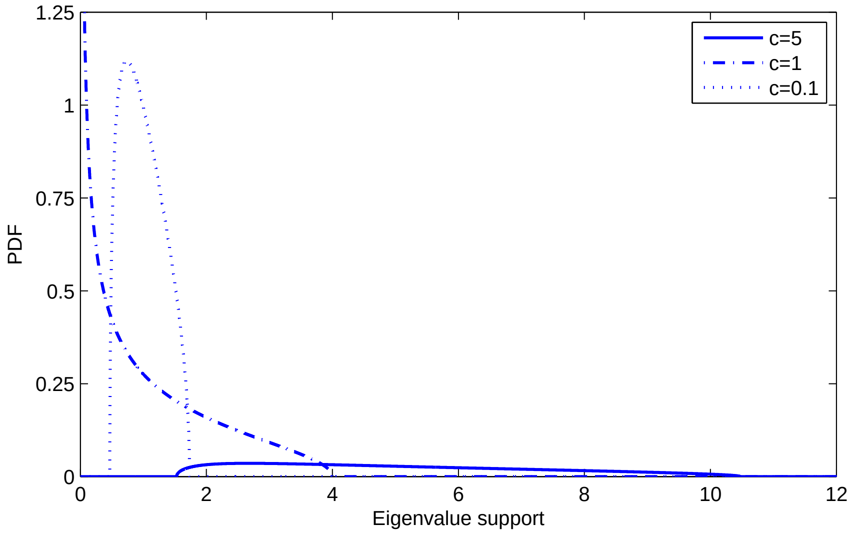


Figure 1.2: Marčenko-Pastur density for the three different aspect ratios of the matrix $\mathbf{X}$ in Example 2

Moreover, from (**P5**) and (**P6**) the eigenvalue density function can be found as

$$p_{\mathsf{B}}(x) = \begin{cases} \frac{1}{2\pi c x}\sqrt{(b-x)(x-a)} & if\ a \le x \le b, \\ 0 & \text{otherwise} \end{cases} \tag{1.25}$$

and has a point mass $1 - c^{-1}$ at the origin if $c > 1$, where $a = (1-\sqrt{c})^2$ and $b = (1+\sqrt{c})^2$. Indeed, note that the imaginary part of $m = m_F(z)$ goes to zero as $z$ approaches the real line and lies outside the interval $[a, b]$, and, therefore, (**P7**) applies. The function in (1.25) is the density function of the Marčenko-Pastur law, whose moments are given in Corollary 4.4. In particular, for $c = 1$, the $k$th moment is equal to the Catalan number $C(k)$ (cf. Appendix B).

**Example 2** *(Marčenko-Pastur law)* $\mathbf{X}$ *has dimensions* $M$ *and* $N$ *such that their ratio is* $c$ *is equal to* 0.5, 1 *and* 5. *Figure 1.2 illustrates the density of the Marčenko-Pastur law for the previous three cases.*

Alternatively, let again $\mathbf{R} = \mathbf{I}_M$ and consider the matrix ensemble $\mathbf{B}^{-1} = \left(\mathbf{X}\mathbf{X}^H\right)^{-1}$. In order to find the limiting spectral distribution of $\mathbf{B}^{-1}$, observe that, from (1.19) and (1.20), we have

$$-\frac{1}{z} m\left(-\frac{1}{z}\right) = 1 + \sum_{k=1}^{\infty} M_k\, z^k,$$

for general Stieltjes transform $m(z)$ and $M_k$ the $k$th moment of the associated distribution. Then, noting that the $k$th moment of $\mathbf{B}$ is the $-k$th moment of $\mathbf{B}^{-1}$, it is easy to verify that

$$-\frac{1}{z}m_F\left(-\frac{1}{z}\right) = -zm_G(z) - 1,$$

where $m_F(z)$ is the Stieltjes transform of the limiting spectral distribution of $\mathbf{B}$ and $m_G(z)$ is that of the limiting spectrum of $\mathbf{B}^{-1}$. Consequently, without further calculation we have that

$$m_G(z) = \frac{1}{z^2}m_F\left(-\frac{1}{z}\right) - \frac{1}{z}.$$

Obviously, the density function of the associated distribution can also be obtained from $p_{\mathsf{B}}(x)$ in (1.25) without the need of integration by using the fact that [Pap91, pag. 94]

$$p_{\mathsf{B}^{-1}}(x) = \frac{1}{x^2}p_{\mathsf{B}}\left(\frac{1}{x}\right),$$

and considering $c \in (0,1)$, to guarantee the existence of the inverse, as

$$p_{\mathsf{B}}(x) = \begin{cases} \frac{1}{2\pi c x^2}\sqrt{(xb-1)(1-ax)}\, if\ \frac{1}{b} \leq x \leq \frac{1}{a}, \\ 0 \qquad\qquad \text{otherwise} \end{cases}$$

Furthermore, observe that the spectrum of $\mathbf{R}^{1/2}\mathbf{X}\mathbf{X}^H\mathbf{R}^{1/2}$ and that of $\mathbf{R}\mathbf{X}\mathbf{X}^H$ coincides. Then, let $\mathbf{W}$ be a matrix defined equivalently as $\mathbf{X}$ and independent of the latter (with possibly different aspect ratio). The special case defined by $\mathbf{R} = \left(\mathbf{W}\mathbf{W}^H\right)^{-1}$ is known as the general multivariate $F$ matrix, and its limiting eigenvalue distribution was obtained in [Sil85] via the method of moments.

Finally, consider the matrix $\left(\mathbf{X}\mathbf{X}^H\right)^{1/2}$. As before, direct transformation of the density function of the Marčenko-Pastur law by using [Pap91, pag. 96]

$$p_{\mathsf{B}^{1/2}}(x) = 2xp_{\mathsf{B}}\left(x^2\right)\mathcal{I}_{[0,\infty)},$$

yields

$$p_{\mathsf{B}}(x) = \begin{cases} \frac{1}{\pi c}\sqrt{\left(b-x^2\right)\left(x^2-a\right)}\, if\ \sqrt{a} \leq x \leq \sqrt{b}, \\ 0 \qquad\qquad \text{otherwise,} \end{cases} \tag{1.26}$$

which, for $c = 1$, is regarded in the RMT literature as *quarter circle law* [Wig67].

**Example 3** *(Quarter Circle Law)* $\mathbf{X}$ *has dimensions* $M = N = 100$*, so that* $c = 1$*. In Figure 1.3, the empirical histogram and the theoretical limiting spectral density are compared.*

**Theorem 1.2** *[Mar67][Sil95b, Theorem 1.1] Consider a* $N \times N$ *Hermitian matrix* $\mathbf{A}$ *with empirical spectral distribution converging, almost surely, vaguely to a nonrandom (possibly defective) probability distribution function* $A$*. Moreover, suppose* $\mathbf{T}$ *is a* $N \times N$ *real diagonal matrix, whose*

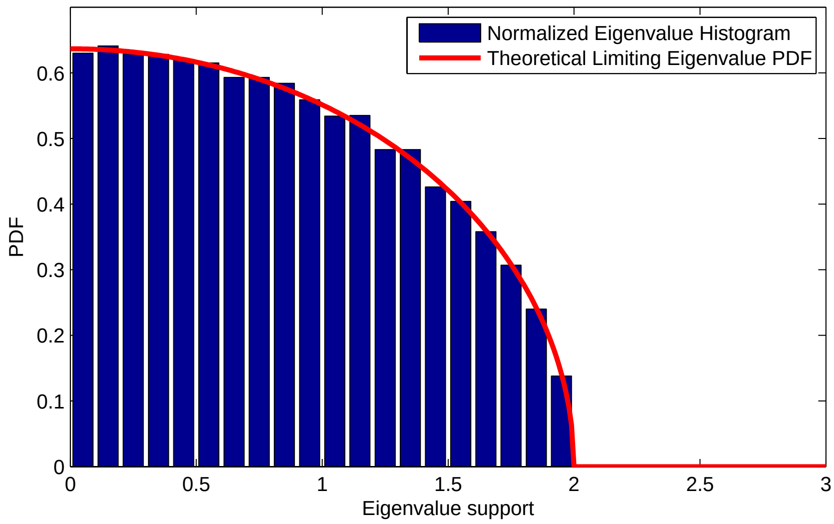


Figure 1.3: Limiting spectral density and normalized histogram of matrix ensemble in Example 3

*empirical spectral distribution converges almost surely to a deterministic probability distribution function* $T$ *as* $N \to \infty$*. Furthermore, define* $\mathbf{B} = \mathbf{A} + \mathbf{X}^H \mathbf{T} \mathbf{X}$*. Then, as* $M, N \to \infty$*, with* $c < +\infty$*, almost surely, the empirical spectral distribution of* $\mathbf{B}$ *converges vaguely to a non-random distribution function* $F_{\mathsf{B}}$ *with associated Stieltjes transform* $m_F(z)$ *such that, for each* $z \in \mathsf{C}^+$*,* $m = m_F(z)$ *is the unique solution in* $\mathsf{C}^+$ *to the following equation, namely,*

$$m = m_A\left(z - c\int \frac{\tau}{1+\tau m} dT(\tau)\right). \tag{1.27}$$

From (1.27), some straightforward connections with the spectral results obtained above can be readily established. For instance, consider the case in which $\mathbf{A} = \mathbf{0}_{N\times N}$. Then, $m_A(z) = -\frac{1}{z}$ and we simply have

$$m_F(z) = \frac{1}{c\int \frac{\tau\, dT(\tau)}{1+\tau m} - z}. \tag{1.28}$$

From (1.28), a functional inverse for the Stieltjes transform $m_F(z)$ can be easily found as

$$z = -\frac{1}{m} + c\int \frac{\tau\, dT(\tau)}{1+\tau m}. \tag{1.29}$$

All of the analytical behavior of the limiting distribution can be extracted from (1.29) [Sil95c]. Moreover, consider the Marčenko-Pastur distribution (i.e., $\mathbf{T} = \mathbf{I}_N$). Then, the Stieltjes transform $m = m_F(z)$ solves the following simplification of equation (1.29), namely,

$$z = -\frac{1}{m} + \frac{c}{1+m},$$

resulting in the quadratic equation

$$zm^2 - (z + 1 - c)\, m + 1 = 0,$$

with solution in $\mathsf{C}^+$ is given by

$$m = \frac{\beta - 1 - z + \sqrt{(z-a)\,(z-b)}}{2z}. \tag{1.30}$$

Again, since the imaginary part of $m = m_F(z)$ vanishes as $z$ approaches $\mathsf{R}$ and lies outside the interval $[a, b]$, from (**P7**) the limiting distribution has a density given, for $x \neq 0$, by

$$p_{\mathsf{B}}(x) = \begin{cases} \frac{1}{2\pi x}\sqrt{(b-x)\,(x-a)}\, if \;\; a \leq x \leq b, \\ 0 \qquad\qquad\qquad\qquad \text{otherwise,} \end{cases} \tag{1.31}$$

with a mass at $1 - \beta$ at zero when $\beta < 1$.

Alternatively, the Stieltjes transform and distribution function in (1.30) and (1.31), respectively, can also be directly obtained from the corresponding expressions in (1.24) and (1.25). Indeed, note that, for any matrix $\mathbf{C} \in \mathsf{C}^{M\times N}$, with $M/N \to c$, the spectra of $\mathbf{C}^H\mathbf{C}$ and $\mathbf{C}\mathbf{C}^H$ differ by $|M - N|$ zero eigenvalues, and so can their associated empirical eigenvalue distribution functions be related as

$$F_{\mathsf{C}^H\mathsf{C}}(x) = (1-c)\,\mathcal{I}_{[0,\infty)} + cF_{\mathsf{C}\mathsf{C}^H}(x), \tag{1.32}$$

where $F_{\mathsf{C}^H\mathsf{C}}$ and $F_{\mathsf{C}\mathsf{C}^H}$ are the spectral distribution functions of the matrices $\mathbf{C}^H\mathbf{C}$ and $\mathbf{C}\mathbf{C}^H$, respectively. Consequently, we further have that

$$m_{\mathsf{C}^H\mathsf{C}}(z) = -\frac{1-c}{z} + cm_{\mathsf{C}\mathsf{C}^H}(z), \tag{1.33}$$

where $m_{\mathsf{C}^H\mathsf{C}}(z)$ and $m_{\mathsf{C}\mathsf{C}^H}(z)$ are correspondingly defined. Thus, using the relations in (1.32) and (1.33) along with the Stieltjes transform and density function of the Marčenko-Pastur distribution, equations (1.30) and (1.31) can be equivalently obtained. More generally, using the identity in (1.33), it can be seen that (1.22) and (1.27) are equivalent. Hence, it is apparent that the restriction on $\mathbf{T}$ being a diagonal matrix can be dropped in order to broaden the class of matrix ensembles for which Theorem 1.2 holds.

**Theorem 1.3** *[Pau07, Theorem 1] Consider the class of matrices of the form* $\mathbf{B} = \mathbf{R}^{1/2}\mathbf{X}\mathbf{T}\mathbf{X}^H\mathbf{R}^{1/2}$ *with* $\mathbf{R}^{1/2}$ *and* $\mathbf{T}$ *defined as in Theorem 1.1 and Theorem 1.2, respectively. Then, with probability one, as* $M, N \to \infty$, *with* $c < +\infty$, *the empirical spectral distribution of* $\mathbf{B}$ *converges weakly to a probability distribution function* $F_{\mathsf{B}}$ *with associated Stieltjes transform* $m_F(z)$ *such that, for each* $z \in \mathsf{C}^+$, $m(z) = m_F(z)$ *is given by*

$$m(z) = \int \frac{1}{a\displaystyle\int \frac{b}{1+cbe}\, dT(b) - z}\, dH(a), \tag{1.34}$$

*where $e = e(z)$ is the unique solution in $\mathbb{C}^+$ to the following equation, namely,*

$$e = \int \frac{a}{a \int \frac{b}{1+cbe} dT(b) - z} dH(a), \tag{1.35}$$

In particular, in the case of the entries of $\mathbf{X}$ being Gaussian distributed, $\mathbf{B}$ can be thought of as modelling the sample covariance matrix of a collection of multidimensional observations with *spatio-temporal correlations* defined by $\mathbf{R}$ and $\mathbf{T}$, respectively, namely, $\mathbf{B} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H$, where $\mathbf{Y} = \mathbf{R}^{1/2}\mathbf{\Xi}\mathbf{T}^{1/2} \in \mathbb{C}^{M\times N}$ is matrix-variate normal distributed, i.e., $\mathbf{Y} \sim \mathcal{CMN}_{M\times N}(\mathbf{0}_{M\times N}, \mathbf{R}, \mathbf{T})$, or, equivalently, $\mathrm{Vec}(\mathbf{Y}) \sim \mathcal{CN}_{MN}(\mathbf{0}_{MN}, \mathbf{R}\otimes\mathbf{T})$ and $\mathbf{E}[\mathbf{B}] = \mathrm{Tr}[\mathbf{T}]\mathbf{R}$ (see e.g. [Kol05, Chapter 2]).

In general, observe that, if we let $\mathbf{R} = \mathbf{I}_M$, then we have $e(z) = m(z)$ and equation (1.34) yields the Stieltjes transform in (1.28).

Alternatively, if we instead have $\mathbf{T} = \mathbf{I}_N$, then we find that $m = m_F(z)$ is the unique solution to

$$m(z) = \int \frac{1}{a\frac{1}{1+ce(z)} - z} dH(a). \tag{1.36}$$

Indeed, from Lemma A.7, if we define

$$w(z) = 1 - c - czm(z), \tag{1.37}$$

it can be easily checked that

$$w(z) = 1 - c - z\frac{c}{M}\mathrm{Tr}\left[(w(z)\mathbf{R} - z\mathbf{I}_M)^{-1}\right] = \frac{1}{1 + \frac{c}{M}\mathrm{Tr}\left[\mathbf{R}(w(z)\mathbf{R} - z\mathbf{I}_M)^{-1}\right]},$$

so that we have

$$w(z) \equiv \frac{1}{1 + ce(z)},$$

establishing the fact that equations (1.22) and (1.36) coincide. In particular, note that, if additionally $\mathbf{R} = \mathbf{I}_M$, then we have $e(z) = m(z)$ and, accordingly,

$$1 - c - czm(z) \equiv \frac{1}{1 + cm(z)}.$$

### 1.1.3 Description of asymptotic spectrum of SCM-type matrices

In this dissertation, random matrix ensembles with a SCM-type structure will be of most relevance, as we shall be interested in parameter estimation methods relying upon the second-order statistics of the set of received observations. For the purposes of statistical inference based on the sample estimate of the true covariance matrix, an analytical characterization of the eigenvalue support of the SCM is often of required. In the following, we briefly summarize some recently existing results on the asymptotic behavior of the support of the spectral distribution of the random matrix ensemble introduced in Theorem 1.1 (see [Mes06a] for further details). In particular, some notational definitions of convenience are introduced that will be repetedly used in our developments in the sequel.

From the one-to-one correspondance between the Stieltjes transform and the spectral distribution of a random matrix ensemble, in order to study the behavior of the support of the limiting eigenvalue density of $\mathbf{B} = \hat{\mathbf{R}}$ in Theorem 1.1, one may proceed by characterizing the solutions to $m = m_F(z)$. To that effect, we note that the limiting Stieltjes transform in (1.22) can be written as

$$m = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{\lambda_m(\mathbf{R})(1-c-czm)-z} \tag{1.38}$$

$$= \frac{(1-c)f-z}{czf}, \tag{1.39}$$

where we have defined the new variable $f$ in terms of $m$ as

$$f = \frac{z}{1-c-czm}. \tag{1.40}$$

From (1.39) and (1.40), it can be easily checked that $f$ and $m$ are the inverse of each other for $z \neq 0$, and establish a bijection between the set $\{m \in \mathbb{C} : m \neq (1-c)/cz\}$ and $\{f \in \mathbb{C} : f \neq 0\}$. Then, inserting (1.39) into (1.38), we see that in order to find the solutions of the limiting Stieltjes transform equation $m$, we can equivalently look at the solutions in $f$ to

$$f\left(1-\frac{c}{M}\sum_{m=1}^{M}\frac{\lambda_m(\mathbf{R})}{\lambda_m(\mathbf{R})-f}\right) = z. \tag{1.41}$$

Indeed, using the definition of $w(z)$ in (1.37), observe that we can rewrite $w(z)$ as

$$\begin{aligned} w(z) &= 1-c-z\frac{c}{M}\operatorname{Tr}\left[(w(z)\mathbf{R}-z\mathbf{I}_M)^{-1}\right] \\ &= 1-c\left[\frac{1}{M}\sum_{m=1}^{M}1+\frac{z}{w(z)\lambda_m(\mathbf{R})-z}\right] \\ &= 1-\frac{c}{M}\sum_{m=1}^{M}\frac{w(z)\lambda_m(\mathbf{R})}{w(z)\lambda_m(\mathbf{R})-z}. \end{aligned}$$

Consequently, from (1.40), we have

$$w(z) = 1-\frac{c}{M}\sum_{m=1}^{M}\frac{\lambda_m(\mathbf{R})}{\lambda_m(\mathbf{R})-f(z)}, \tag{1.42}$$

so that we get $fw = z$ as in (1.41).

Using the previous definitions, a direct characterization of the eigenvalue support can be afforded as follows. Let us define the following function of $f$, namely,

$$\Phi(f) = f\left(1-\frac{c}{M}\sum_{m=1}^{M}\frac{\lambda_m(\mathbf{R})}{\lambda_m(\mathbf{R})-f}\right). \tag{1.43}$$

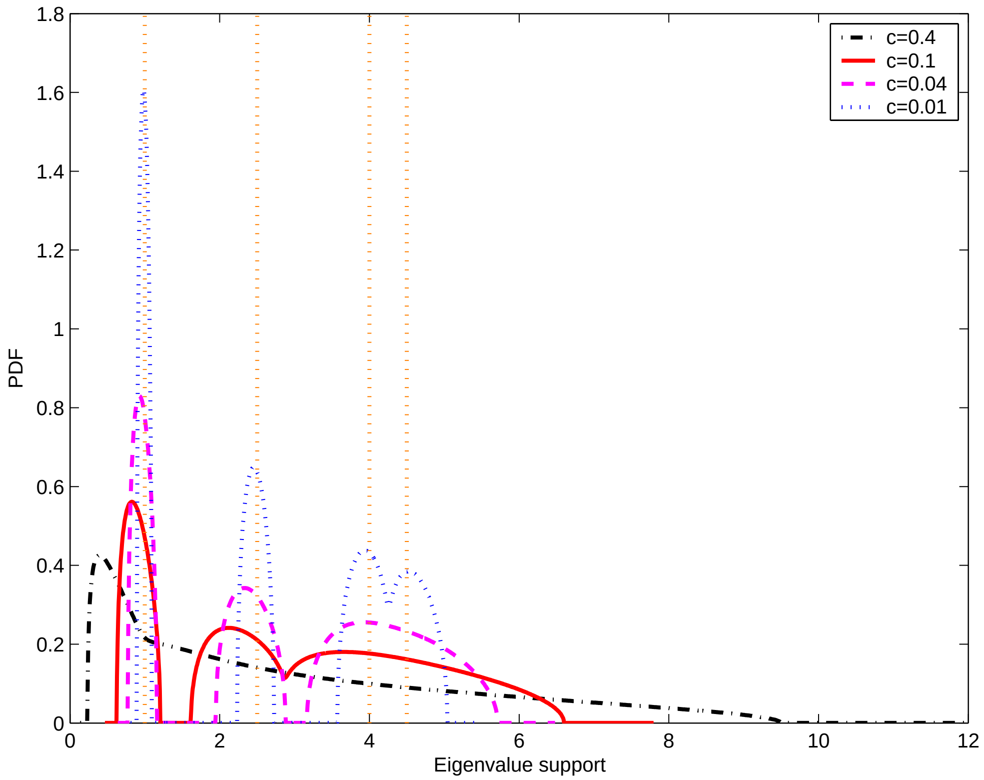


Figure 1.4: Limiting eigenvalue density for the random SCM-type ensamble in Example 1. The splitting phenomena of eigenvalue clusters is illustrated as a function of the sample support $N$.

Then, interestingly enough, it turns out that the support of the eigenvalue density can be described by using $\Phi(f)$ in (1.43) as the union of $Q$ intervals, namely $\left[x_1^-, x_1^+\right] \cup \ldots \cup \left[x_Q^-, x_Q^+\right]$, where

$$x_q^- = \Phi\left(f_q^-\right), \; x_q^+ = \Phi\left(f_q^+\right), \tag{1.44}$$

and $f_1^- < f_1^+ \leq f_2^- < f_2^+ \leq \ldots \leq f_Q^- < f_Q^+$ are the unique real-valued solutions in $f$ to $\Phi'(f) = 0$, or, equivalently,

$$\frac{1}{M}\sum_{m=1}^{M}\left(\frac{\lambda_m(\mathbf{R})}{\lambda_m(\mathbf{R}) - f}\right)^2 = \frac{1}{c}. \tag{1.45}$$

The previous procedure for the analytical characterization of the theoretical asymptotic density of the eigenvalue spectrum can be illustrated by using Figure 1.4, where the density function associated with the random SCM-type ensamble in Example 1 is depicted for different values of the ratio $M/N$. The eigevalue cluster splitting characterized by (1.44) is exemplified by letting the number of sample observations grow for a fixed observation dimension (i.e., by decreasing the ratio $c$). Indeed, given two consecutive eigenvalues, there exists a minimum number of samples per observation dimension that guarantees the corresponding eigenvalue

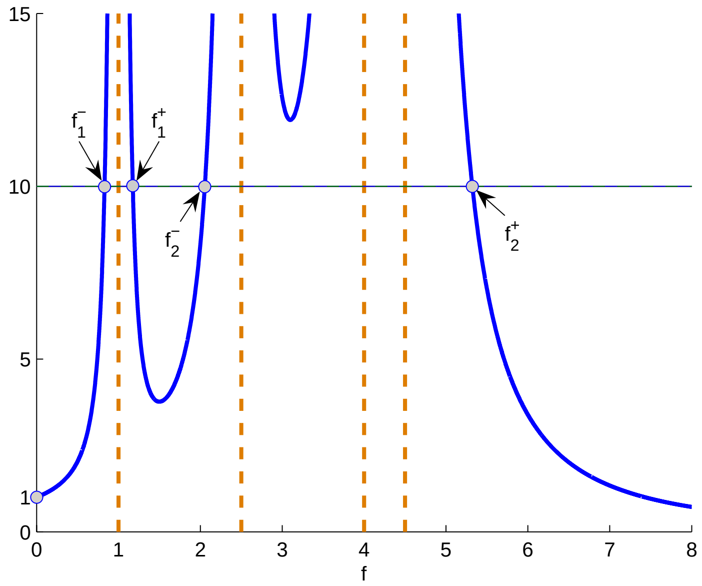


Figure 1.5: LHS of (1.45) for Example 1. The solutions of $\Phi'(f) = 0$ characterizing the clusters constituting the eigenvalue density are shown. Furthermore, the splitting effect of the eigenvalue clusters can be predicted from the minima of the function (namely, the solutions to $\Phi''(f) = 0$) and the location of the horizontal bar given by the ratio $N/M$.

cluster to split. Specifically, consider the $k$th eigenvalue of $\mathbf{R}$. It turns out that the cluster of the asymptotic eigenvalue distribution corresponding to this eigenvalue can be ensured to be separated from the clusters associated with adjacent eigenvalues if and only if

$$\frac{N}{M} > \frac{1}{M} \sum_{m=1}^{M} \left( \frac{\lambda_m(\mathbf{R})}{\lambda_m(\mathbf{R}) - \xi} \right)^2, \tag{1.46}$$

where is the $k$th real-valed solution to $\Phi''(f) = 0$, i.e.,

$$\frac{1}{M} \sum_{m=1}^{M} \frac{\lambda_m^2(\mathbf{R})}{(\lambda_m(\mathbf{R}) - f)^3} = 0.$$

In Figure 1.5, the LHS of (1.45), obtained from $\Phi'(f) = 0$, is depicted for the SCM ensemble in Example 1. The cluster splitting behavior described by (1.46) for a particular sample support $(N)$ can be predicted from the location of the minima of the displayed function.

Finally, Figure 1.6 shows the solutions to $\Phi(f) = 0$, namely the LHS of

$$\frac{1}{M} \sum_{m=1}^{M} \frac{\lambda_m(\mathbf{R})}{\lambda_m(\mathbf{R}) - f} = \frac{1}{c}. \tag{1.47}$$

In particular, the equation in (1.47) represents a canonical equation whose roots will be of

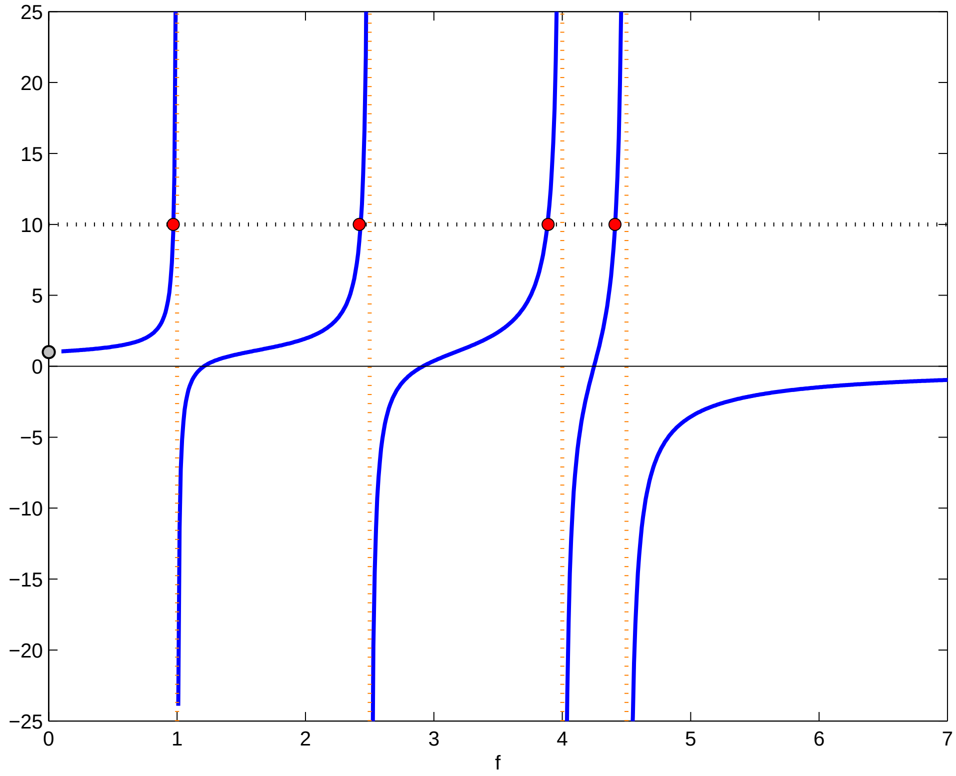


Figure 1.6: LHS of (1.47) for Example 1. The solutions to $\Phi(f) = 0$ are shown in red.

special interest in our estimation framework. Furthermore, the function $w(z)$ in (1.37) and its equivalent representation in (1.42), as well as $f(z) = z/w(z)$, will prove very useful for the derivation of the estimators proposed in this thesis (see motivational remarks in Section 1.3.3).

### 1.1.4 Vector-valued spectral distributions

The asymptotic convergence results outlined above are concerned with the behavior of matrix eigenvalue spectra. Very often, performance analyses and algorithm designs in statistical signal processing and wireless communications can be relied on the eigenvalues of a certain random matrix model, so that the previous results are of unquestionable practical interest, as it can be drawn from the vast engineering literature that is based on Theorems 1.1 to 1.3. However, in many other situations, the characterization of an objective function is required that depends upon not only the eigenvalues but also the associated eigensubspaces. For instance, consider the following quantity, namely, $\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}$, where $\mathbf{R}$ is the covariance matrix of, for example, a set of observed samples, and $\mathbf{s}$ is a given (nonrandom) signature vector. As it will be clarified in the subsequent section, the proposed estimation methods in this thesis will be based on the correction of the limit in the doubly-asymptotic regime of the corresponding traditional estimator based on the SCM.

In particular, consider the limit of the random quantity $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}$. Let the SCM be decomposed as $\hat{\mathbf{R}} = \hat{\mathbf{Q}}\hat{\mathbf{\Lambda}}\hat{\mathbf{Q}}^H$, where $\hat{\mathbf{\Lambda}}$ is a diagonal matrix with the eigenvalues of $\hat{\mathbf{R}}$ and $\hat{\mathbf{Q}}$ is the matrix with the associated eigenvectors. From the spectral theorem, it is clear that $\mathbf{s}^H\hat{\mathbf{Q}}\hat{\mathbf{\Lambda}}^{-1}\hat{\mathbf{Q}}^H\mathbf{s}$. Thus, it is obvious that, in order to establish the limit of $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}$, not only the asymptotic behavior of the SCM spectrum but also that of the sample eigenvectors is required. While there are fairly many results in the literature of RMT about the eigenvalues of random matrices of increasing dimensions, not much has been reported about the eigenvectors.

Building on the fact that the eigenvectors of a Wishart matrix are Haar distributed, that is, they follow a uniform distribution over the group of unitary matrices, the following result can be established.

**Theorem 1.4** *[Sil86, Theorem 2] Let* $\mathbf{W}$ *be a* $N \times N$ *standard complex central Wishart matrix with eigendecomposition* $\mathbf{W} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, *such that* $\mathbf{y} = \begin{bmatrix} y_1 \cdots y_N \end{bmatrix} = \mathbf{U}\mathbf{x}$ *has a uniform distribution over the unit sphere* $\left\{\mathbf{x} \in \mathsf{R}^N : \|\mathbf{x}\| = 1\right\}$ *(i.e., if* $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$, *then* $\mathbf{y} \sim \mathbf{z}/\|\mathbf{z}\|$*). If the entries of* $\mathbf{W}$ *have all moments finite, then for every* $t \in [0,1]$,

$$\frac{X_N(t)}{\sqrt{N}} \to 0,$$

*almost surely, as* $N \to \infty$, *where*

$$X_N(t) = \sqrt{\frac{N}{2}} \sum_{n=1}^{\lfloor tN \rfloor} y_n^2 - \frac{1}{N} \sim \sqrt{\frac{N}{2}} \frac{1}{\|\mathbf{z}\|^2} \sum_{n=1}^{\lfloor tN \rfloor} |\{\mathbf{z}\}_n|^2 - \frac{\|\mathbf{z}\|^2}{N},$$

*where $\lfloor \cdot \rfloor$ denotes the greatest integer function.*

Observe that, equivalently, we have

$$\sum_{n=1}^{\lfloor tN \rfloor} y_n^2 \to t.$$

almost surely, as $N \to \infty$. The previous result is like a law of large numbers, showing some degree of similarity in the large matrix limit to the uniformity over the unit sphere in the finite case.

In order to extend the asymptotic convergence results reviewed above to the study of the limiting behavior of random quantities involving the eigensubspaces of a certain random matrix ensemble, it will be convenient to define the following empirical distribution function based on the spectrum as well as the associated eigenspace of a random matrix $\mathbf{M} \in \mathsf{C}^{M \times M}$, namely,

$$H_{\mathsf{M}}^{M}(\lambda) = \sum_{m=1}^{M} |a_m|^2 \mathcal{I}_{(\lambda_m(\mathsf{M}) \leq \lambda)}, \tag{1.48}$$

where $a_m = \{\mathbf{a}\}_m$, $\mathbf{a} = \mathbf{U}\mathbf{x}$, with $\mathbf{U}$ being the matrix of eigenvectors of $\mathbf{M}$ and $\mathbf{x}$ being defined as in Theorem 1.4. Clearly, $H_{\mathsf{M}}^{M}(\lambda)$ in (1.48) is a random probability distribution function, with a Stieltjes transform given by

$$m_H(z) = \mathbf{x}^H (\mathbf{M} - z\mathbf{I}_M)^{-1} \mathbf{x}. \tag{1.49}$$

Note that (1.48) represents a weighted version of the empirical eigenvalue distribution function in (1.1), with the weights non-trivially characterizing the asymptotic convergence of the random quantity in (1.48), or, equivalently, the associated Stieltjes transform in (1.49). In particular, if $a_m = \frac{1}{M}$, $m = 1, \ldots, M$, then $H_{\mathsf{M}}^{M}(\lambda)$ in (1.48) and $F_{\mathsf{M}}^{M}(\lambda)$ in (1.1) clearly coincide.

The following result extends the spectral convergence theorems introduced previously in this chapter (cf. Section 1.1.2) to the convergence of the Stieltjes transform of the more general distribution function in (1.48).

**Theorem 1.5** *Let $\mathbf{B} = \mathbf{A} + \mathbf{R}^{1/2}\mathbf{X}\mathbf{T}\mathbf{X}^H\mathbf{R}^{1/2}$ with the matrices $\mathbf{A}$ and $\mathbf{R}^{1/2}\mathbf{X}\mathbf{T}\mathbf{X}^H\mathbf{R}^{1/2}$ being defined as in Theorems 1.2 and 1.3, respectively. Moreover, consider a nonrandom matrix $\mathbf{\Theta} \in \mathsf{C}^{M \times M}$ with uniformly bounded Frobenius norm for all $M$. Then, with probability one, as $M, N \to \infty$, with $c < +\infty$, for each $z \in \mathsf{C}^+$,*

$$\left| \mathrm{Tr}\left[ \mathbf{\Theta} \left( (\mathbf{B} - z\mathbf{I}_M)^{-1} - (\mathbf{A} + x(z)\mathbf{R} - z\mathbf{I}_M)^{-1} \right) \right] \right| \to 0, \tag{1.50}$$

*where*

$$x(z) = \frac{1}{M} \mathrm{Tr}\left[ \mathbf{T} (\mathbf{I}_M + ce\mathbf{T})^{-1} \right], \tag{1.51}$$

*and $e = e(z)$ is the unique solution in $\mathsf{C}^+$ to the following equation, namely,*

$$e = \frac{1}{M} \mathrm{Tr}\left[ \mathbf{R} (\mathbf{A} + x(z)\mathbf{R} - z\mathbf{I}_M)^{-1} \right]. \tag{1.52}$$

**Corollary 1.1** *(Asymptotic convergence of eigenvectors) Let* $\boldsymbol{\Theta} = \mathbf{a}_2\mathbf{a}_1^H$*, with* $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{C}^M$ *two deterministic vectors with uniformly bounded Euclidean norm for all* $M$*. Then, the result establishes the asymptotic convergence of the class of vector-valued Stieltjes transforms defined by (1.49), with* $\mathbf{M}$ *being equal to the family of random matrix ensambles defined by* $\mathbf{B}$ [3]*.*

**Corollary 1.2** *(Limiting Stieltjes transform of spectral distributions) Let* $\boldsymbol{\Theta} = \frac{1}{M}\mathbf{I}_M$*. Then, if* $\mathbf{A} = \mathbf{0}_{M\times M}$*, (1.50) is equivalent to the convergence result in Theorem 1.3; if* $\mathbf{R} = \mathbf{I}_M$*, (1.50) coincides with the result in Theorem 1.2; finally, if* $\mathbf{A} = \mathbf{0}_{M\times M}$ *and* $\mathbf{T} = \mathbf{I}_N$*, we get (1.22).*

**Proof.** In the following, we give a sketch of the proof of Theorem 1.5 that shall summarize the key elements and techniques involved in the derivations of the convergence results in Theorems 1.1 to 1.3.

Let us first write

$$\mathbf{R}^{1/2}\mathbf{X}\mathbf{T}\mathbf{X}^H\mathbf{R}^{1/2} = \frac{1}{N}\sum_{n=1}^{N} t_n \mathbf{R}^{1/2}\boldsymbol{\xi}_n\boldsymbol{\xi}_n^H\mathbf{R}^{1/2} \equiv \frac{1}{N}\sum_{n=1}^{N}\mathbf{y}_n\mathbf{y}_n^H,$$

where $t_i = \{\mathbf{T}\}_{i,i}$, and define

$$\mathbf{B}_n = \mathbf{B} - \frac{1}{N}\mathbf{y}_n\mathbf{y}_n^H.$$

For the sake of notational convenience, we will use the following definitions in our derivations,

$$\begin{aligned}
\mathbf{Q}(z) &= (\mathbf{B} - z\mathbf{I}_M)^{-1}, \\
\mathbf{Q}_n(z) &= (\mathbf{B}_n - z\mathbf{I}_M)^{-1}, \\
\mathbf{P}(z) &= (\mathbf{A} + x(z)\mathbf{R} - z\mathbf{I}_M)^{-1}
\end{aligned}$$

Furthermore, define

$$x(z) = \frac{1}{N}\sum_{n=1}^{N}\frac{t_n}{1 + t_n\frac{c}{M}\operatorname{Tr}[\mathbf{R}\mathbf{Q}(z)]}.$$

Now, consider the equality

$$\mathbf{B} - z\mathbf{I}_M = \mathbf{A} - (z\mathbf{I}_M - x(z)\mathbf{R}) + \frac{1}{N}\mathbf{R}^{1/2}\boldsymbol{\Xi}\mathbf{T}\boldsymbol{\Xi}^H\mathbf{R}^{1/2} - x(z)\mathbf{R}.$$

We will proceed by factoring the difference of inverses as

$$\mathbf{P}(z) - \mathbf{Q}(z) = \mathbf{P}(z)\left(\frac{1}{N}\mathbf{R}^{1/2}\boldsymbol{\Xi}\mathbf{T}\boldsymbol{\Xi}^H\mathbf{R}^{1/2} - x(z)\mathbf{R}\right)\mathbf{Q}(z),$$

[3] From the assumptions on $\mathsf{a}_1, \mathsf{a}_2$, it is clear that simple normalization of the vectors by their norm yields a random sub-probability distribution function $F_{\mathsf{M}}^{M}(\lambda)$ in (1.48), being a proper distribution if $\mathsf{a}_1 = \mathsf{a}_2$.

where we have used the resolvent identity, i.e., $\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{B}^{-1}(\mathbf{A} - \mathbf{B})\mathbf{A}^{-1}$. Furthermore, we expand the middle factor as

$$\begin{aligned}\left(\frac{1}{N}\mathbf{R}^{1/2}\boldsymbol{\Xi}\mathbf{T}\boldsymbol{\Xi}^H\mathbf{R}^{1/2} - x(z)\mathbf{R}\right)\mathbf{Q}(z) &= \frac{1}{N}\sum_{n=1}^{N} t_n\mathbf{R}^{1/2}\boldsymbol{\xi}_n\boldsymbol{\xi}_n^H\mathbf{R}^{1/2}\mathbf{Q}(z) - x(z)\mathbf{R}\mathbf{Q}(z)\\ &= \frac{1}{N}\sum_{n=1}^{N} t_n\mathbf{R}^{1/2}\boldsymbol{\xi}_n\boldsymbol{\xi}_n^H\mathbf{R}^{1/2}\mathbf{Q}(z) - \frac{t_n\mathbf{R}\mathbf{Q}(z)}{1 + t_n\frac{1}{N}\operatorname{Tr}[\mathbf{R}\mathbf{Q}(z)]}\\ &= \frac{1}{N}\sum_{n=1}^{N}\frac{t_n\mathbf{R}^{1/2}\boldsymbol{\xi}_n\boldsymbol{\xi}_n^H\mathbf{R}^{1/2}\mathbf{Q}_n(z)}{1 + t_n\frac{1}{N}\boldsymbol{\xi}_n^H\mathbf{R}^{1/2}\mathbf{Q}(z)\mathbf{R}^{1/2}\boldsymbol{\xi}_n} - \frac{t_n\mathbf{R}\mathbf{Q}(z)}{1 + t_n\frac{1}{N}\operatorname{Tr}[\mathbf{R}\mathbf{Q}(z)]},\end{aligned}$$

where, in the last equality, we have used the Sherman-Morrison formula as

$$\mathbf{Q}(z) = \mathbf{Q}_n(z) - \frac{1}{N}\frac{t_n\mathbf{Q}_n(z)\mathbf{R}^{1/2}\boldsymbol{\xi}_n\boldsymbol{\xi}_n^H\mathbf{R}^{1/2}\mathbf{Q}_n(z)}{1 + t_n\frac{1}{N}\boldsymbol{\xi}_n^H\mathbf{R}^{1/2}\mathbf{Q}_n(z)\mathbf{R}^{1/2}\boldsymbol{\xi}_n}. \tag{1.53}$$

Now, observe that we can write

$$\begin{aligned}&\left(\boldsymbol{\Xi}\boldsymbol{\Xi}^H - x(z)\mathbf{R}\right)\mathbf{Q}(z) =\\ &= \frac{t_n}{1 + t_n\frac{1}{N}\operatorname{Tr}[\mathbf{R}\mathbf{Q}(z)]}\frac{1}{N}\sum_{n=1}^{N}\frac{t_n\left(\frac{1}{N}\operatorname{Tr}[\mathbf{R}\mathbf{Q}(z)] - \frac{1}{N}\boldsymbol{\xi}_n^H\mathbf{R}^{1/2}\mathbf{Q}_n(z)\mathbf{R}^{1/2}\boldsymbol{\xi}_n\right)}{1 + t_n\frac{1}{N}\boldsymbol{\xi}_n^H\mathbf{R}^{1/2}\mathbf{Q}_n(z)\mathbf{R}^{1/2}\boldsymbol{\xi}_n}\mathbf{R}^{1/2}\boldsymbol{\xi}_n\boldsymbol{\xi}_n^H\mathbf{R}^{1/2}\mathbf{Q}_n(z)\\ &\quad + \frac{t_n}{1 + t_n\frac{1}{N}\operatorname{Tr}[\mathbf{R}\mathbf{Q}(z)]}\frac{1}{N}\sum_{n=1}^{N}\mathbf{R}^{1/2}\boldsymbol{\xi}_n\boldsymbol{\xi}_n^H\mathbf{R}^{1/2}\mathbf{Q}_n(z) - \mathbf{R}\mathbf{Q}_n(z)\\ &\quad + \frac{t_n}{1 + t_n\frac{1}{N}\operatorname{Tr}[\mathbf{R}\mathbf{Q}(z)]}\frac{1}{N}\sum_{n=1}^{N}\mathbf{R}\mathbf{Q}_n(z) - \mathbf{R}\mathbf{Q}(z).\end{aligned}$$

Consequently, noting from Lemma A.8 that

$$\left|\frac{t_n}{1 + t_n\frac{1}{N}\operatorname{Tr}[\mathbf{R}\mathbf{Q}(z)]}\right| < +\infty,$$

we just need to show that, as $M, N \to \infty$ with $M/N \to c < +\infty$, almost surely,

$$\left|\frac{1}{N}\sum_{n=1}^{N}\operatorname{Tr}\left[\boldsymbol{\Theta}_1\left(\mathbf{R}^{1/2}\boldsymbol{\xi}_n\boldsymbol{\xi}_n^H\mathbf{R}^{1/2}\mathbf{Q}_n(z) - \mathbf{R}\mathbf{Q}_n(z)\right)\right]\right| \to 0 \tag{1.54}$$

$$\left|\frac{1}{N}\sum_{n=1}^{N}\mathbf{a}^H\left(\mathbf{R}\mathbf{Q}_n(z) - \mathbf{R}\mathbf{Q}(z)\right)\mathbf{b}\right| \to 0 \tag{1.55}$$

$$\left|\frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{N}\operatorname{Tr}[\mathbf{R}\mathbf{Q}(z)] - \frac{1}{N}\boldsymbol{\xi}_n^H\mathbf{R}^{1/2}\mathbf{Q}_n(z)\mathbf{R}^{1/2}\boldsymbol{\xi}_n\right)\frac{t_n\mathbf{a}^H\mathbf{R}^{1/2}\boldsymbol{\xi}_n\boldsymbol{\xi}_n^H\mathbf{R}^{1/2}\mathbf{Q}_n(z)\mathbf{b}}{1 + t_n\frac{1}{N}\boldsymbol{\xi}_n^H\mathbf{R}^{1/2}\mathbf{Q}_n(z)\mathbf{R}^{1/2}\boldsymbol{\xi}_n}\right| \to 0 \tag{1.56}$$

where we have defined $\tilde{\boldsymbol{\Theta}} = \mathbf{P}(z)\boldsymbol{\Theta}$.

In particular, the result in (1.54) follows directly from Lemma A.5, using $\mathbf{C} = \tilde{\boldsymbol{\Theta}}\mathbf{R}^{1/2}$ and $\mathbf{U}_n(z) = \mathbf{R}^{1/2}\mathbf{Q}_n$.

Let us now consider (1.55). As before, applying (1.53) we just need to show

$$\left| \frac{1}{N} \sum_{n=1}^{N} \frac{t_n}{1 + t_n \frac{1}{N} \boldsymbol{\xi}_n^H \mathbf{R}^{1/2} \mathbf{Q}_n(z) \mathbf{R}^{1/2} \boldsymbol{\xi}_n} \frac{1}{N} \boldsymbol{\xi}_n^H \mathbf{R}^{1/2} \mathbf{Q}_n(z) \tilde{\boldsymbol{\Theta}} \mathbf{Q}_n(z) \mathbf{R}^{1/2} \boldsymbol{\xi}_n \right| \to 0$$

According to Lemma A.2, it is enough to prove that

$$\max_{1 \le m,n \le N} \mathsf{E} \left[ \left| \frac{1}{N} \frac{t_n \boldsymbol{\xi}_n^H \mathbf{R}^{1/2} \mathbf{Q}_n(z) \tilde{\boldsymbol{\Theta}} \mathbf{Q}_n(z) \mathbf{R}^{1/2} \boldsymbol{\xi}_n}{1 + t_n \frac{1}{N} \boldsymbol{\xi}_n^H \mathbf{R}^{1/2} \mathbf{Q}_n(z) \mathbf{R}^{1/2} \boldsymbol{\xi}_n} \right|^p \right] \le \frac{C}{N^{1+\delta}}, \tag{1.57}$$

for some constants $C$, $\delta > 0$ and $p > 1$ not depending on $N$. Using the Cauchy-Schwarz inequality, we can write the expectation in (2.78) as

$$\begin{aligned} &\mathsf{E} \left[ \left| \frac{1}{N} \frac{t_n \boldsymbol{\xi}_n^H \mathbf{R}^{1/2} \mathbf{Q}_n(z) \tilde{\boldsymbol{\Theta}} \mathbf{Q}_n(z) \mathbf{R}^{1/2} \boldsymbol{\xi}_n}{1 + t_n \frac{1}{N} \boldsymbol{\xi}_n^H \mathbf{R}^{1/2} \mathbf{Q}_n(z) \mathbf{R}^{1/2} \boldsymbol{\xi}_n} \right|^p \right] \\ &\le \mathsf{E}^{1/2} \left[ \left| \frac{t_n}{1 + t_n \frac{1}{N} \boldsymbol{\xi}_n^H \mathbf{R}^{1/2} \mathbf{Q}_n(z) \mathbf{R}^{1/2} \boldsymbol{\xi}_n} \right|^{2p} \right] \mathsf{E}^{1/2} \left[ \left| \frac{1}{N} \boldsymbol{\xi}_n^H \mathbf{R}^{1/2} \mathbf{Q}_n(z) \tilde{\boldsymbol{\Theta}} \mathbf{Q}_n(z) \mathbf{R}^{1/2} \boldsymbol{\xi}_n \right|^{2p} \right]. \end{aligned}$$

Again, we note that, from Lemma A.8,

$$\left| \frac{t_n}{1 + t_n \frac{1}{N} \boldsymbol{\xi}_n^H \mathbf{R}^{1/2} \mathbf{Q}_n(z) \mathbf{R}^{1/2} \boldsymbol{\xi}_n} \right| < +\infty,$$

and so is also its expectation bounded. Therefore, we just need to show

$$\max_{1 \le m,n \le N} \frac{1}{N^p} \mathsf{E}^{1/2} \left[ \left| \boldsymbol{\xi}_n^H \mathbf{R}^{1/2} \mathbf{Q}_n(z) \tilde{\boldsymbol{\Theta}} \mathbf{Q}_n(z) \mathbf{R}^{1/2} \boldsymbol{\xi}_n \right|^{2p} \right] \le \frac{C}{N^{1+\delta}}, \tag{1.58}$$

but this holds for $p \ge 2$ , since the expectation in (1.58) is bounded for $p \ge 1$ from Lemma A.4.

Finally, the convergence in (1.56) can be similarly proved following the same line of reasoning as for (1.55). ■

Corollaries 1.1 and 1.2 establish the asymptotic convergence of the eigenvalues and associated eigensubspaces of a broad class of random matrix ensembles through the characterization of the limiting behavior of appropriately defined Stieltjes transforms. The double-limit asymptotics of the spectrum of certain random matrix models has been extensively used in the literature for the performance characterization and design of communication systems. In this dissertation, we focus on the performance and consistent estimation of traditional statistical signal processing methods in the dobly-asymptotic regime. To that effect, an explicit characterization of the limiting convergence of the eigenvectors of sample covariance matrices that is less common in the literature will be required.

Before proceeding with a short introduction to our generalized consistent estimation framework, in the following section we provide and example of application of the limiting characterization of the eigenvectors of random matrices to a typical problem in wireless communications.

In particular, the limiting eigensubspace characterization of the random matrix ensamble underlying the type of information-plus-noise covariance matrices, and complementing the model in Theorem 1.5 is studied.

## 1.2 Application example of the limiting convergence of random eigenspaces

In this section, an analitycal characterization of the transient regime of a training-based multiple-input multiple-output (MIMO) system exploiting the full diversity order of an arbitrarily correlated MIMO fading channel via optimal beamforming and combining is presented. No channel state information is assumed to be available at either the transmitter or the receiver side, so that the design of the optimal transmit beamformer and receive combiner is necessarily based on a finite collection of samples observed during a training phase. The focus is on practical scenarios where the length of the training sequence is comparable in magnitude to the system size. In these situations, the performance of the MIMO system can be expected to suffer from a considerable degradation. In order to characterize the actual performance under the previous realistic conditions, a large-system performance analysis is proposed that builds upon the theory of large-dimensional random matrices. In particular, the asymptotic spectral characterization of information-plus-noise covariance matrices in [Doz07b, Theorem 1.1] is extended to the case involving also the eigensubspaces. The proposed method is numerically validated in the context of a typical MIMO application.

### 1.2.1 Diversity Analysis of MIMO Systems with Limited Training

The performance of MIMO channels can be significantly enhanced if the channel state is known to the transmitter, the receiver, or both. Alternative blind techniques applied in order to avoid channel training may often incur in a nonnegligible loss of performance and a fairly increased computational complexity. In practice, the coefficients of a MIMO channel often vary over time and need to be estimated. If the channel state varies slowly, one may carry out some measurements in order to learn the channel statistics and estimate (or predict) its instantaneous realization. Typically, the channel coefficients are measured at the receiver by having the transmitter send known training vectors. Knowledge of the channel at the receiver can be sent to the transmitter via feedback channels [Lov04].

The impact of the realistic availability of an imprecise channel state information (CSI) in the capacity due to multiplexing gains predicted for MIMO systems is summarized in [Gol03]. On the other hand, the tradeoff between the time and the power allocated to training operation and data transmission was evaluated in [Has03]. In particular, the authors provide the optimum number of pilots and training power allocation of a training-based MIMO system in the sense of maximizing a lower-bound on the Shannon capacity over the class of ergodic block-fading (memoryless and uncorrelated) channels, as a function of the number of transmit and receive antennas, the received signal-to-noise ratio (SNR) and the length of the fading coherence time. Earlier related contributions include [Mar99b], as well as [Mar99c, Zhe02], where the number of channel uses available for training and the optimal input distribution achieving capacity at high SNR over unknown block-fading uncorrelated MIMO channels with a finite coherence time interval is investigated. Furthermore, the effects of pilot-assisted channel estimation on achievable data rates over frequency-flat time-varying channels is analyzed in [Sam03]. Along with the time-division multiplexing training scheme considered in the previous works, a tight lower-bound on the maximum mutual information of a MIMO system using superimposed pilots is derived in [Col07].

Much less effort has been placed in understanding the consequences of the lack of CSI on the achieved diversity gain of an unknown MIMO channel that is learned by means of a training sequence of finite length. Indeed, perfect knowledge of the channel realization can be used in general to modulate each transmitted symbol onto a beamforming vector matched to the channel in order to improve the received SNR. In particular, if the MIMO channel is completely known to the transmitter, the evident choice of the beamforming vector is the right eigenvector of the channel matrix corresponding to the maximum singular value in amplitude, which maximizes the received SNR. In [Rub08], the problem of optimal transmit beamforming maximizing the received SNR over unknown MIMO channels with given Gaussian statistics is addressed.

In the following application example, we will focus on the problem of achieving full diversity gain over an unknown, *arbitrary* block-fading MIMO channel by optimal transmit beamforming and receive combining. We assume a certain given amount of channel uses is allocated for training purposes at the beginning of each coherence interval, such that both sides can learn the channel from a sequence of known training beams. Instead of following the generally suboptimal approach consisting of obtaining an intermediate estimate of the channel matrix to be used for further processing, we pursue the direct estimation of both optimal (channel-adapted) beamformer vector and receive combiner using the sequence of pilots during the so-called training phase. In particular, we are interested in the actual empirical performance obtained from a limited number of training samples per degree-of-freedom. To that effect, we provide a large-system analysis of such a training-based MIMO scheme that allows us to consider, as in practice, the number of transmit and receive antennas, as well as the length of training sequence to be comparable in magnitude.

The application example is structured as follows. In Section 1.2.2, the problem of pilot-aided transmiter and receiver estimation is addressed. Section 1.2.3 provides a large system performance analysis of the transient estimation regime, which is numerically validated in Section

1.2.4. After the final discussion in Section 1.2.5, the derivation of the asymptotic convergence results are given in the appendices.

### 1.2.2 Channel model and transceiver estimation

Consider the linear vector channel model corresponding to a MIMO transmission system with $M$ receive antennas and $K$ transmit antennas, namely, the received signal is expressed as

$$\mathbf{y}(n) = \mathbf{H}\mathbf{x}(n) + \mathbf{n}(n), \quad n = 1, 2, \ldots \tag{1.59}$$

where $\mathbf{x}(n) \in \mathsf{C}^K$ represents the transmitted signal, $\mathbf{n}(n) \in \mathsf{C}^M$ is the background noise, and $\mathbf{H} \in \mathsf{C}^{M \times K}$ models an arbitrary MIMO channel matrix. The noise process is assumed to be wide-sense stationary, with independent and identically distributed (i.i.d.) standarized complex Gaussian[4] vector entries such that $\mathsf{E}\left[\mathbf{n}(l)\,\mathbf{n}(m)^H\right] = \sigma_n^2 \delta_{l,m} \mathbf{I}_M$, where $\delta_{l,m}$ is the Kronecker delta function. Specifically, one wishes to modulate a sequence of transmitted symbols $x(n)$ onto a (unit-norm) beamforming vector $\mathbf{v} \in \mathsf{C}^K$ ($\mathbf{x}(n) = \mathbf{v}x(n)$), so that the received signal becomes

$$\mathbf{y}(n) = \mathbf{H}\mathbf{v}x(n) + \mathbf{n}(n), \quad n = 1, 2, \ldots \tag{1.60}$$

As mentioned above, the purpose of using multiple antennas here is to enhance through beamforming the SNR at the receiver side and after matched filtering, namely,

$$\mathsf{SNR} = \frac{\left|\mathbf{u}^H \mathbf{H} \mathbf{v}\right|^2}{\sigma_n^2}, \tag{1.61}$$

where $\mathbf{u} \in \mathsf{C}^M$ represents the receiver matched to the MIMO channel. In particular, the receiver and transmiter vectors maximizing the SNR are resp. the right and left top singular vectors of $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, henceforth denoted by $\mathbf{u}_1$ and $\mathbf{v}_1$. Accordingly, the maximum achievable $\mathsf{SNR}$ is

$$\max_k \left\{ \mathsf{SNR}_k = \frac{\sigma_k^2}{\sigma_n^2} \right\} = \mathsf{SNR}_1 = \frac{\sigma_1^2}{\sigma_n^2}, \tag{1.62}$$

where $\sigma_k^2 = [\mathbf{\Sigma}]_{k,k}^2$ is the power over the $k$th channel eigenmode and $\mathsf{SNR}_k$ is its associated signal-to-noise ratio.

We assume that no CSI is available at either the transmitter or the receiver side, and that a sequence of $N$ fixed pilot beams $\mathbf{b}(n) \in \mathsf{C}^K$ consuming a certain given amount of training energy are available for transceiver estimation purposes. Accordingly, the received signal becomes ($\mathbf{x}(n) = \mathbf{b}(n)$)

$$\mathbf{y}(n) = \mathbf{H}\mathbf{b}(n) + \mathbf{n}(n), \quad n = 1, 2, \ldots$$

[4] A random variable is standarized complex Gaussian if its real and imaginary parts are i.i.d. with mean zero and variance $1/2$.

By collecting the column vector observations at different instants of time in a matrix $\mathbf{Y} \in \mathsf{C}^{M \times N}$, we can rewrite

$$\begin{aligned}\mathbf{Y} &= [\mathbf{y}(1), \ldots, \mathbf{y}(N)] \\ &= \mathbf{HB} + \mathbf{N},\end{aligned} \tag{1.63}$$

where we have defined

$$\begin{aligned}\mathbf{B} &= [\mathbf{b}(1), \ldots, \mathbf{b}(N)], \\ \mathbf{N} &= [\mathbf{n}(1), \ldots, \mathbf{n}(N)].\end{aligned}$$

In the following, we consider the problem of empirical estimation of the optimal transceiver given a fixed training energy budget (i.e., power allocation strategy across pilot beams and length of training phase), namely,

$$\{\mathbf{u}, \mathbf{v}\} = \arg \max_{\mathsf{u},\mathsf{v}: \|\mathsf{B}\|_F^2 \leq E} \mathsf{E}\left[\mathsf{SNR} | \mathbf{y}(1), \ldots, \mathbf{y}(N)\right],$$

where $E$ determines the constraint on the total energy consumed by training. In particular, note that the total energy constraint will be related to the power allocated to the beamvector pilots sended during the training phase, as well as the lenght of this training window (i.e., number of training beams). Furthermore, for estimation purposes, observe that $\mathbf{u}_1$ is the top eigenvector of $\mathbf{HH}^H$, whereas $\mathbf{v}_1$ is the top eigenvector of $\mathbf{H}^H\mathbf{H}$. The achieved system performance based on pilot-assisted transceiver estimation clearly depends on the selection of training beams. In this work, we will focus on the more relevant case in practice of orthogonal training. In particular, we assume that the training phase is defined by a set of orthogonal (unitary) beams satisfying the training budget constraint, such that $\mathbf{BB}^H = E/K\mathbf{I}_K$. In other words, the training sequences (colum vectors of $\mathbf{B}$) satisfy the Welch-bound equality (WBE) [Wel74, Ver98]. In the multiuser detection literature, WBE signature sequences are known to maximize the sum capacity achieved by overloaded symbol-synchronous code-division multiple-access channels with equal average-input-energy constraints [Mas93, Rup94]. The optimality of WBE sequences for transmit beamforming schemes maximizing the received SNR is discussed in [Rub08]. For the purposes of statistically analyzing the effect of limited training in the performance of pilot-assisted MIMO systems, it will be in order to assume in the sequel the following model for the training matrix, namely, $\mathbf{B} = E/K\mathbf{U}^H$, where the columns of $\mathbf{U} \in \mathsf{C}^{N \times K}$ are orthogonal, such that $\mathbf{U}^H\mathbf{U} = \mathbf{I}_K$.

**Receiver estimation.** Since the top eigenvector of $\mathbf{HH}^H$ is equal to the principal eigenvector of the covariance matrix of the received observations, namely,

$$\mathbf{R} = \mathsf{E}\left[\mathbf{y}(n)\mathbf{y}^H(n)\right] = E/K\mathbf{HH}^H + \sigma_n^2\mathbf{I}_M,$$

the problem of estimating $\mathbf{u}_1$ can be directly approached by equivalently finding an estimator of the top eigenvector of $\mathbf{R}$. To that effect, we may use the sample estimate of the latter, namely the sample covariance matrix (SCM), i.e.,

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{y}(n)\,\mathbf{y}^H(n) = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H. \tag{1.64}$$

From the strong law of large numbers, the SCM is a consistent estimator of the theoretical covariance matrix. Note that, as the length of the training phase $N$ increases, the SCM converges to its average $\mathsf{E}\left[\hat{\mathbf{R}}\right] = \mathbf{R}$. In fact, the SCM is the minimum variance unbiased estimator of $\mathbf{R}$ [And03]. Moreover, for Gaussian observations, the maximum-likelihood (ML) estimator of the principal eigenvector of $\mathbf{R}$ is the corresponding eigenvector of $\hat{\mathbf{R}}$ [Mui82].

**Transmitter estimation.** In order to find an estimator of the optimum transmitter, consider the following construction based on the (known) training vectors, namely,

$$\hat{\mathbf{C}} = \frac{1}{N}\left(\mathbf{B}^{\#}\right)^H \mathbf{Y}^H\mathbf{Y}\mathbf{B}^{\#}, \tag{1.65}$$

where $(\cdot)^{\#}$ denotes the Moore-Penrose pseudoinverse, i.e., $\mathbf{B}^{\#} = \left(\mathbf{B}\mathbf{B}^H\right)^{-1}\mathbf{B}$. Indeed, note that, as $N$ goes to infinity, almost surely, $\hat{\mathbf{C}} \to \mathbf{C}$, where

$$\begin{aligned}\mathbf{C} &= \mathsf{E}\left[\hat{\mathbf{C}}\right]\\ &= \frac{1}{N}\mathbf{H}^H\mathbf{H} + \sigma_n^2\frac{MK}{EN}\mathbf{I}_K.\end{aligned}$$

In the following section, we provide an analytical characterization of the performance of a training-based MIMO system under the realistic assumption of a training phase length comparable in magnitude with the system dimension.

### 1.2.3 Large system performance analysis

In this section, we are interested in assessing the performance of a training-based MIMO system under a limited training budget. In particular, we will concentrate on the effect of a bounded ratio between training sample-size and number of degrees of freedom. In this work, in order to study the effect of the energy budget limitation as essentially due to a finite training sequence length, we assume a fixed power allocation across training beams given by $\|\mathbf{b}(n)\|^2 = 1$, $n = 1, \ldots, N$.

Using the principal eigenvectors of $\hat{\mathbf{R}}$ and $\hat{\mathbf{C}}$, denoted in the sequel by $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{v}}_1$, respectively, as the estimators of resp. the optimum receiver and transmitter maximizing the SNR, we are interested in evaluating the performance loss incurred in practice by the use of the estimated solutions, namely,

$$\widehat{\mathsf{SNR}} = \left| \sum_{k=1}^{K \wedge M} \sqrt{\mathsf{SNR}_k}\ \hat{\mathbf{u}}_1^H \mathbf{u}_k \mathbf{v}_k^H \hat{\mathbf{v}}_1 \right|^2 . \tag{1.66}$$

Observe that the lack of an accurate estimate will contribute to the spread of power over the different orthogonal subchannels (similar to a linear programming suboptimal solution to the power allocation problem). In order to analitically characterize the performance measure in (1.66), it is enough to characterize the projection of the transceiver estimate obtained from a finite training sample-support onto the eigensubspaces spanned by the different right and left singular vectors. Indeed, for an unlimited training energy budget, as $N \to \infty$ (infinite training phase length), we clearly have $\hat{\mathbf{u}}_1^H \mathbf{u}_k \mathbf{v}_k^H \hat{\mathbf{v}}_1 \to 1\delta_{1,k}$, and, accordingly,

$$\widehat{\mathsf{SNR}} \to \mathsf{SNR}_1 .$$

The (finite-dimensional) statistical analysis of the quantity in (1.66) for finite system-size and limited training energy is rather intricate, and only an asymptotic characterization in the large-sample regime might be affordable [And03, Mui82]. Therefore, we focus on a large-system analysis of (1.66) and let not only the number of training samples ($N$), but also both the number transmit ($K$) and receive ($M$) antennas (i.e., the system dimension) go to infinity at a constant rate, defined by $\alpha = M/N$ and $\beta = K/N$. Since the previous asymptotic framework better matches realistic deployment conditions in practice, we may expect our results to more approppriately model the system performance in a practical setting characterized by a limited amount of training beams per degree-of-freedom.

Regarding the projections in the summation in (1.66) involving the estimates $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{v}}_1$, we may rely on the following procedure based on the power method for finding the eigenvalues and associated eigenvectors of an arbitrary Hermitian matrix. In particular, let us concentrate for instance on the top eigenvector of $\hat{\mathbf{R}}$ as the estimate of the optimal receiver. Then, consider the following quantity, namely,

$$\frac{\boldsymbol{v}^H \left( \hat{\mathbf{R}} - \xi \mathbf{I}_M \right)^{-1} \mathbf{u}_k}{\left( \boldsymbol{v}^H \left( \hat{\mathbf{R}} - \xi \mathbf{I}_M \right)^{-2} \boldsymbol{v} \right)^{1/2}}, \tag{1.67}$$

where $\boldsymbol{v} \in \mathsf{C}^K$ is any vector with a non-zero component in the direction of $\hat{\mathbf{u}}_1$ and $\xi = \lambda_1 \left( \hat{\mathbf{R}} \right) + \epsilon$, with $\lambda_1 \left( \hat{\mathbf{R}} \right)$ being the maximum eigenvalue of $\hat{\mathbf{R}}$ and $\epsilon$ being a small strictly positive constant. Indeed, $\hat{\mathbf{u}}_1^H \mathbf{u}_k$ can be arbitrarily well approximated by the expression in (1.67) for an arbitrarily small $\epsilon > 0$. For the purpose of analysis, we can use $\boldsymbol{v} = \mathbf{u}_k$ (as $M, N$ go to infinity, with probability one, $\hat{\mathbf{u}}_1$ has a non-zero component in the direction of $\mathbf{u}_k$, for each $k$). Then, we finally have

$$\frac{\mathbf{u}_k^H \left(\hat{\mathbf{R}} - \xi \mathbf{I}_M\right)^{-1} \mathbf{u}_k}{\left(\mathbf{u}_k^H \left(\hat{\mathbf{R}} - \xi \mathbf{I}_M\right)^{-2} \mathbf{u}_k\right)^{1/2}}. \tag{1.68}$$

Note that an equivalent procedure follows for the optimal combiner at the receiver side by replacing the sample covariance matrix $\hat{\mathbf{R}}$ with the matrix $\hat{\mathbf{C}}$ in (1.65), and $\mathbf{u}_k$ with $\mathbf{v}_k$. In particular, using the previous procedure, an arbitrarily well approximated SNR estimate in (1.66) can be obtained as

$$\widehat{\mathsf{SNR}}\left(\xi_1, \xi_2\right) = \left| \sum_{k=1}^{K \wedge M} \sqrt{\mathsf{SNR}_k} \frac{\mathbf{u}_k^H \left(\hat{\mathbf{R}} - \xi_1 \mathbf{I}_M\right)^{-1} \mathbf{u}_k}{\left(\mathbf{u}_k^H \left(\hat{\mathbf{R}} - \xi_1 \mathbf{I}_M\right)^{-2} \mathbf{u}_k\right)^{1/2}} \frac{\mathbf{v}_k^H \left(\hat{\mathbf{C}} - \xi_2 \mathbf{I}_K\right)^{-1} \mathbf{v}_k}{\left(\mathbf{v}_k^H \left(\hat{\mathbf{C}} - \xi_2 \mathbf{I}_K\right)^{-2} \mathbf{v}_k\right)^{1/2}} \right|^2, \tag{1.69}$$

where $\xi_1 = \lambda_1\left(\hat{\mathbf{R}}\right) + \epsilon_1$ and $\xi_2 = \lambda_1\left(\hat{\mathbf{C}}\right) + \epsilon_2$, with $\epsilon_1$ and $\epsilon_2$ being two two arbitrarily small strictly positive constants.

For the purposes of validating the proposed analytical characterization, we consider a Rayleigh MIMO channel matrix with particularly low-rank, such that the highest eigenmode alone essentially characterizes the full diversity gain that can be achieved over the channel. Note that, apart from simplifying the numerical validation, such a scenario renders specially relevant the accurate analysis and estimation of the diversity gain achieved by a MIMO system. Thus, as an approximation of $\widehat{\mathsf{SNR}}\left(\xi_1, \xi_2\right)$, we consider

$$\widetilde{\mathsf{SNR}}\left(\xi_1, \xi_2\right) = \left| \sqrt{\mathsf{SNR}_1} \frac{\mathbf{u}_1^H \left(\hat{\mathbf{R}} - \xi_1 \mathbf{I}_M\right)^{-1} \mathbf{u}_1}{\left(\mathbf{u}_1^H \left(\hat{\mathbf{R}} - \xi_1 \mathbf{I}_M\right)^{-2} \mathbf{u}_1\right)^{1/2}} \frac{\mathbf{v}_1^H \left(\hat{\mathbf{C}} - \xi_2 \mathbf{I}_K\right)^{-1} \mathbf{v}_1}{\left(\mathbf{v}_1^H \left(\hat{\mathbf{C}} - \xi_2 \mathbf{I}_K\right)^{-2} \mathbf{v}_1\right)^{1/2}} \right|^2, \tag{1.70}$$

In particular, we build upon the fact that the expression in (1.68) is given in terms of the resolvent of $\hat{\mathbf{R}}$. Interestingly enough, the so-called Stieltjes transform of the empirical eigenvalue distribution function of this matrix is also defined in terms of its resolvent. Hence, we may resort to the theory of the spectral analysis of large-dimensional random matrices in order to analytically characterize resolvent-type expressions of $\hat{\mathbf{R}}$ as a function of its limiting spectral distribution. In particular, the following theorem provides an analytical characterization of the asymptotic behavior of the spectrum of $\hat{\mathbf{R}}$.

**Theorem 1.6** *[Doz07b, Theorem 1.1] Let* $\hat{\mathbf{R}} = \frac{1}{N}\left(\mathbf{S}+\mathbf{N}\right)\left(\mathbf{S}+\mathbf{N}\right)^H$*, where* $\mathbf{N}$ *is a* $M \times N$ *complex random matrix, whose entries have independent and identically distributed real and imaginary parts, with mean zero, variance* $\sigma^2/2$ *and bounded moments, and* $\mathbf{S}$ *a* $M \times N$ *complex matrix, such that the empirical distribution function of the eigenvalues of* $N^{-1}\mathbf{S}\mathbf{S}^H$ *converges almost surely to the probability distribution function of the eigenvalues of the nonrandom matrix*

*$\mathbf{\Psi} \in \mathsf{C}^{M\times M}$, as $M, N \to \infty$ with $M/N \to c < +\infty$. Moreover, consider two $M$-dimensional deterministic complex vectors $\mathbf{a}, \mathbf{b}$ with uniformly bounded Euclidean norm for all $M$. Then, with probability one, in the previous asymptotic regime, the Stieltjes transform of the empirical eigenvalues distribution function of $\hat{\mathbf{R}}$ converges as*

$$\left|\frac{1}{M}\operatorname{Tr}\left[\left(\hat{\mathbf{R}} - z\mathbf{I}_K\right)^{-1}\right] - \frac{1}{M}\operatorname{Tr}\left[\left(\frac{1}{1+\sigma^2 cm}\mathbf{\Psi} + \sigma^2 w\left(z\right)\mathbf{I}_M - z\mathbf{I}_M\right)^{-1}\right]\right| \to 0, \tag{1.71}$$

*where*

$$w\left(z\right) = 1 - c - czm, \tag{1.72}$$

*and $m = m\left(z\right)$ is the solution for any $z \in \mathsf{C}^{+}$ of the following functional equation in $m$, namely,*

$$m = \frac{1}{M}\operatorname{Tr}\left[\left(\frac{1}{1+\sigma^2 cm}\mathbf{\Psi} + \sigma^2 w\left(z\right)\mathbf{I}_M - z\mathbf{I}_M\right)^{-1}\right]. \tag{1.73}$$

Furthermore, since, for any $M \times N$ random matrix $\mathbf{B}$, the Stieljtes transforms of $\mathbf{B}\mathbf{B}^H$ and $\mathbf{B}^H\mathbf{B}$ are related as

$$m_{\mathsf{B}\mathsf{B}^H}\left(z\right) = -\frac{1-c}{z} + cm_{\mathsf{B}^H\mathsf{B}}\left(z\right), \tag{1.74}$$

the result in Proposition 1.6 can be used to characterize the asymptotic spectrum of $\hat{\mathbf{C}}$.

The result in Theorem 1.6 can be easily extended to the case in which $\mathbf{N}$ is matrix-variate normal distributed with arbitrary spatio-temporal correlations. The latter is a special case corresponding to a Kronecker correlation structure of the more general result in [Hac07], where a random matrix $\mathbf{N}$ with an arbitrary variance profile is considered.

Note that the asymptotic convergence established by (1.71) concerns only the spectrum of the matrix $\hat{\mathbf{R}}$. However, both numerator and denominator in (1.68) are given as functions of also the eigensubspaces of the matrix. Thus, an extension of Theorem 1.6 characterizing the asymptotic behavior of vector-valued quadratic forms of the resolvent of $\hat{\mathbf{R}}$ is required. The following proposition provides an asymptotic description of the spectrum and the eigenvectors of the matrix $\hat{\mathbf{R}}$.

**Proposition 1.1** *Let $\hat{\mathbf{R}}$ be defined as in Theorem 1.6. Furthermore, consider two $M$-dimensional deterministic complex vectors $\mathbf{a}_1, \mathbf{a}_2$ with uniformly bounded Euclidean norm for all $M$. Then, as $M, N \to \infty$, $M/N \to c < +\infty$ , almost surely for any $z \in \mathsf{C}^{+}$,*

$$\left|\mathbf{a}_1^H\left(\hat{\mathbf{R}} - z\mathbf{I}_K\right)^{-1}\mathbf{a}_2 - \mathbf{a}_1^H\left(\frac{1}{1+\sigma^2 cm}\mathbf{\Psi} + \sigma^2\left(1-c-czm\right)\mathbf{I}_M - z\mathbf{I}_M\right)^{-1}\mathbf{a}_2\right| \to 0, \tag{1.75}$$

*where $m$ is given in Theorem 1.6.*

**Proof.** See Appendix A. ■

Hence, from the properties of the Stieltjes transform of probability measures outlined in Section, an asymptotic equivalent of (1.68) can be obtained as

**Corollary 1.3** *Let* $\mathbf{a}_1 = \mathbf{a}_2 = \mathbf{u}_1$ *in Proposition 1.1 and consider* $\sigma^2 = 1$*, and* $z = \xi$*, with* $\xi = \lambda_1\left(\hat{\mathbf{R}}\right) + \epsilon$*,* $\epsilon > 0$*. Then, as* $M, N \to \infty$*,* $M/N \to c < +\infty$ *,*

$$\frac{\mathbf{u}_1\left(\hat{\mathbf{R}} - \xi\mathbf{I}_M\right)^{-1}\mathbf{u}_1}{\left(\mathbf{u}_1^H\left(\hat{\mathbf{R}} - \xi\mathbf{I}_M\right)^{-2}\mathbf{u}_1\right)^{1/2}} \asymp \frac{U_1(\xi)}{\left(U_2(\xi)\right)^{1/2}}, \tag{1.76}$$

*where*

$$U_1(\xi) = \mathbf{u}_1^H\mathbf{T}^{-1}\mathbf{u}_1, \tag{1.77}$$

$$U_2(\xi) = \zeta(\xi)\,\mathbf{u}_1^H\mathbf{T}^{-2}\mathbf{u}_1, \tag{1.78}$$

*with*

$$\mathbf{T} = \frac{1}{1+cm}\mathbf{\Psi} + w(\xi)\,\mathbf{I}_M - \xi\mathbf{I}_M,$$

*and*

$$\zeta(\xi) = \frac{1+cm}{1+cm - \frac{1}{N}\operatorname{Tr}\left[\mathbf{T}^{-1}\right] + \left(w(\xi) - \xi(1+cm)\right)\frac{1}{N}\operatorname{Tr}\left[\mathbf{T}^{-2}\right]}.$$

**Proof.** See Appendix B. ■

Finally, based on the results in Proposition 1.1 and Corollary 1.3, we have the following asymptotic limit for the proposed approximation in (1.66) of the SNR in (1.70), namely,

**Proposition 1.2** *Under the previous statistical assumptions, for* $\mathsf{SNR}_k$ *uniformly bounded for all* $k$*, as* $M, N \to \infty$*,* $M/N \to c < +\infty$ *,*

$$\widetilde{\mathsf{SNR}}(\xi_1, \xi_2) \asymp \left|\sqrt{\mathsf{SNR}_1}\,\frac{U_1(\xi_1)}{\left(U_2(\xi_1)\right)^{1/2}}\frac{V_1(\xi_2)}{\left(V_2(\xi_2)\right)^{1/2}}\right|^2, \tag{1.79}$$

*where* $V_1$ *and* $V_2$ *are defined equivalently to, respectively,* $U_1$ *and* $U_2$ *for the covariance matrix* $\hat{\mathbf{C}}$.

**Proof.** See Appendix C. ■

In the following section, we numerically evaluate the accuracy of the approximant asymptotic equivalent (1.79) in describing the transient regime of pilot-aided MIMO transceivers with limited training.

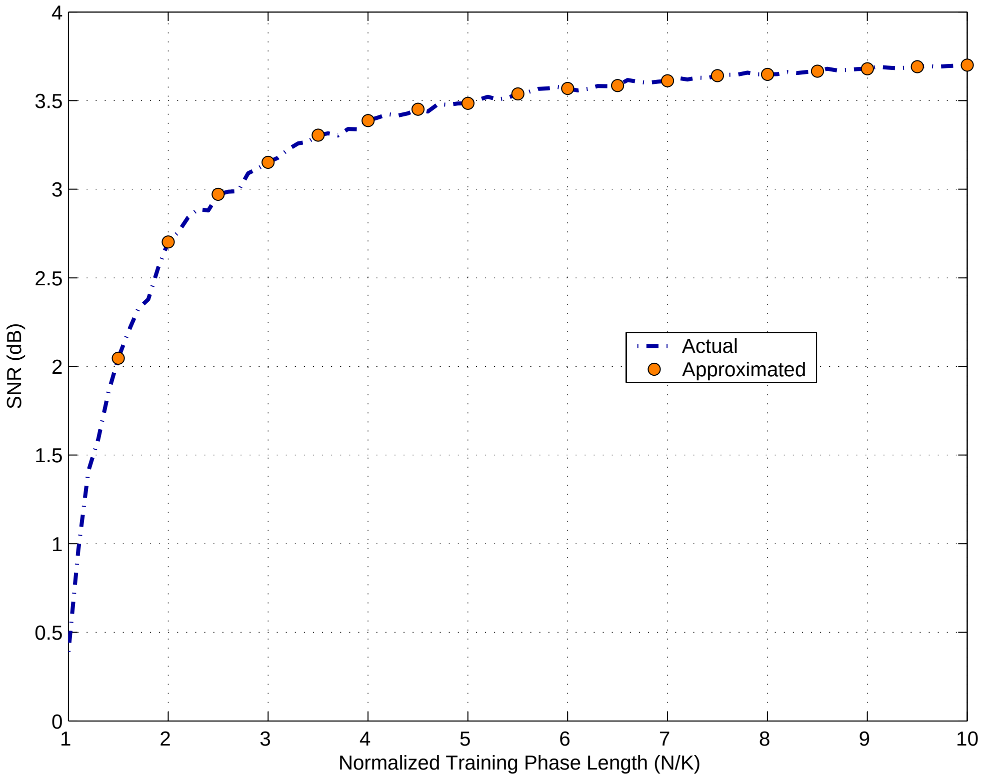


Figure 1.7: Simulated and theoretically predicted SNR performance of MIMO system with empirically estimated optimal transceiver. $K = 10, M = 8$.

### 1.2.4 Numerical results

In this section, we numerically validate the analytical characterization of the transient SNR performance of a training-based MIMO system. Specifically, we assume that both optimal transmit beamformer and receive combiner are empirically estimated from a finite collection of channel observations during a training phase of given length as described in Section 1.2.2. In particular, the empirical performance in terms of averaged received SNR conditioned to the available training samples, i.e., $\mathsf{E}\left[\mathsf{SNR}\left(N+1\right)|\,\mathbf{y}\left(1\right),\ldots,\mathbf{y}\left(N\right)\right]$, is compared with the large-system performance approximation afforded in Section 1.2.3. Figure 1.7 shows both numerically simulated and theoretically approximated SNR performance for a MIMO system with $K = 10$ transmit antennas and $M = 8$ receive antennas, versus the length of the training phase (normalized by $K$). The noise variance is assumed to be one.

### 1.2.5 Concluding remarks

In this section, we have presented an analitycal characterization of the transient regime of a training-based MIMO system exploiting the full diversity order of an arbitrarily correlated MIMO fading channel via optimal beamforming and combining. Since no channel state information is in practice available at either the transmitter or the receiver side, the design of the optimal transmit beamformer and receive combiner is most often based on a finite collection of samples observed during a training phase. If the length of the training sequence is comparable in magnitude to the system size, the performance of the MIMO system can be expected to suffer from a considerable degradation. While the finite-size statistical analysis of the problem is rather involved, the characterization based on the limiting behavior in the large-sample asymptotic regime does not provide any insight into the transient performance. In order to shed some light on the actual performance under the previous practical conditions, we have proposed a large-system SNR performance analysis that build upon the theory of large-dimensional random matrices. In particular, our results are based on an extension of the asymptotic spectral characterization of information-plus-noise covariance matrices to the case involving also the eigen-subspaces via certain vector-valued quadratic forms of the resolvent. A power-iteration-based approach allows for a limiting description of the projection of the sample principal eigenvector onto the true principal eigenspaces. The proposed method is numerically validated in the context of a typical application of training-based MIMO systems.

## 1.3 General Statistical Analysis

In this section, we deal with the problem of statistical inference for general, arbitrarily large-dimensional observations, which is namely the ultimate motivation for the present dissertation. In the following, we shortly summarize the basic idea underlying the proposed estimation approach based on RMT, and then outline the fundamentals of Girko's general statistical analysis of observations of large dimensions. In particular, GSA has served as a motivating approach to the inferential methods proposed in this thesis.

### 1.3.1 Statistical inference based on the SCM

In this dissertation, we shall be interested in parametric statistical methods based on the second-order statistics of a collection of observed samples. More specifically, a class of theoretically optimal solutions given in terms of the true covariance matrix (and, therefore, from the spectral factorization theorem, as a function of its eigenvalues and eigenvectors) will be approximated or estimated using the SCM. In particular, we propose a familly of estimators of the previous spectral functions of the covariance matrix that are consistent for an arbitrarily large observation dimension. Note that by imposing such a condition we are likely to obtain a better estimation performance in realistic scenarios, namely characterized in practice by the number of samples available and the observation dimension being comparable in magnitude. A possible approach consists of analyzing the behavior of the traditional estimator, consistent in the conventional limiting regime where the size of the sample increases without bound whereas its dimension remains fixed, in a more meaningful asymptotic regime that allows for both sample size and dimension to go to infinity at the same rate. Then, based on the limiting behavior established, a correction may be introduced in order for the newly found estimator to converge to the original quantity depending on the true covariance matrix.

### 1.3.2 G-estimation

We present next the basic rationale behind Girko's general statistical analysis of observations of large dimensions [Gir95, Gir98].

Let us begin with the key elements. In GSA, an instance of the Stieltjes transform is defined on the field of the real numbers and associated with a weighted version of the empirical eigenvalue distribution function of the SCM. By SCM, denoted as $\hat{\mathbf{R}}$, we will hereafter mean the random matrix ensemble in Theorem 1.22. Moreover, $\mathbf{R}$ will denote the theoretical covariance matrix.

We introduce the so-called real-valued Stieltjes transform (also known as $\eta$-transform in the engineering literature [Tul04, Section 2.2.2]) of a distribution function, say $G$, defined as

$$r_G(x) = \int \frac{dG(\lambda)}{1+\lambda x}, \qquad x \in \mathsf{R},\ x \geq 0. \tag{1.80}$$

Note that, as for the characterization of the Stieltjes transform in Section 1.1.1 as a MGF, observe that we can write

$$r_G(x) = \sum_{k=0}^{\infty} (-x)^k M_k^G,$$

where is the $k$th moment of $G$.

Interestingly enough, the real-valued Stieltjes transform in (1.80) turns out to be of special interest for the characterization of the asymptotic behavior of certain spectral function of $\mathbf{R}$ in terms of the eigenspectrum of $\hat{\mathbf{R}}$. In particular, let $F_{\mathsf{M}}^{M}(\lambda)$ in (1.1) and $H_{\mathsf{M}}^{M}(\lambda)$ in (1.48) be defined in terms of the eigenvalues and eigenvectors of $\mathbf{R}$, and let $r_F(x)$ and $r_H(x)$, denote their respective real-valued Stieltjes transforms, i.e.,

$$r_F(x) = \frac{1}{M} \operatorname{Tr}\left[(\mathbf{I}_M + x\mathbf{R})^{-1}\right], \tag{1.81}$$

and

$$r_H(x) = \mathbf{a}_1^H (\mathbf{I}_M + x\mathbf{R})^{-1} \mathbf{a}_2, \tag{1.82}$$

where $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{C}^M$ are two nonrandom vectors with uniformly bounded Euclidean norm for all $M$. Furthermore, observe that, when $z$ is restricted to the real negative axis, the following relations hold, namely,

$$r_F(x) = \frac{1}{x} m_F\left(-\frac{1}{x}\right), \qquad r_H(x) = \frac{1}{x} m_H\left(-\frac{1}{x}\right).$$

Assume that a given quantity depending on $\mathbf{R}$ through different combinations of the Stieltjes transforms of $r_F(x)$ and $r_H(x)$ is to be estimated. If the true covariance matrix $\mathbf{R}$ is directly replaced by its sample estimate $\hat{\mathbf{R}}$, only consistency in the classical sense, i.e. as the number of samples go to infinity, can be guaranteed. Since both the number of samples and the dimension of the observations are most often in practice comparable in magnitude, an (asymptotic) approximation of the given quantity in a double-limit regime considering a fixed constant ratio between both $M$ and $N$ will turn out to better resemble realistic situations. In the sequel, consistency related to this doubly-asymptotic regime will be referred to as $M, N$-consistency, as a generalization of classical $N$-consistency. In particular, note that $M, N$-consistent estimators are clearly consistent for arbitrarily high-dimensional observations.

According to General Statistical Analysis, $M, N$-consistent estimators of quantities defined in terms of Stieltjes transforms of the type shown above can be found by simply identifying a uniformly consistent estimator of the generic (real-valued) Stieltjes transform in (1.80) under the same asymptotic conditions. From this basic estimator of the real Stieltjes transform, $M, N$-consistent estimators of more complicated quantites relying on $r_F(x)$ and $r_H(x)$ as building blocks can be readily derived without any assumption on the actual distribution of the observations (other than zero-mean, bounded moments and circularity).

Specifically, it is proved in [Gir98] that the limiting real-valued Stieltjes transforms of $F_{\mathsf{M}}^M(\lambda)$ and $H_{\mathsf{M}}^M(\lambda)$ converge as $M, N \to \infty$, with $M/N \to c < +\infty$, as

$$r_F(x) \asymp \frac{1}{M} \operatorname{Tr}\left[\left(\mathbf{I}_M + \theta(x)\hat{\mathbf{R}}\right)^{-1}\right], \tag{1.83}$$

and

$$r_H(x) \asymp \mathbf{x}^H \left(\mathbf{I}_M + \theta(x)\hat{\mathbf{R}}\right)^{-1} \mathbf{x}, \tag{1.84}$$

respectively, where $\theta(x)$ is the unique positive solution to the following canonical equation, namely,

$$\theta(x)\left(1 - c + \frac{1}{N} \operatorname{Tr}\left[\left(\mathbf{I}_M + \theta(x)\hat{\mathbf{R}}\right)^{-1}\right]\right) = x. \tag{1.85}$$

In the theory of GSA, the RHS of (5.27) and (5.28) are regarded as $G_2$-estimator and $G_{25}$-estimator, respectively (see e.g. [Gir98, Chapter 14]). Clearly, if the sample-size increases and the observation dimension remains constant ($c \to 0$), the estimator is equivalent to its $N$-consistent counterpart from classical estimation theory. This approach was used in [Mes06c] (see also [Mes05, Chapter 4]) to estimate the (asymptotically) optimum parameter of a diagonally loaded minimum variance beamformer under finite sample-support.

### 1.3.3 Generalized Consistent Estimation

In the following, a systematic approach to obtain the previously introduced G-estimators is presented. It is based on the inversion of the intrinsic relationship between the asymptotic spectrum of $\mathbf{R}$ and the limiting Stieltjes transform of the SCM spectrum.

More precisely, observe that, from the asymptotic limit of the Stieltjes transform of the empirical spectral distribution function of $\hat{\mathbf{R}}$ (cf. Theorem 1.1), the following relations are straightforward for each $z \in \mathsf{C}^+$, namely,

$$\hat{w}(z) \asymp w(z), \tag{1.86}$$

$$\hat{f}(z) \asymp f(z), \tag{1.87}$$

as $M, N \to \infty$, with $c < +\infty$, where we have defined

$$\hat{w}(z) = 1 - c - cz\frac{1}{M}\operatorname{Tr}\left[\left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1}\right] \tag{1.88}$$

$$= 1 - \frac{c}{M}\sum_{m=1}^{M}\frac{\lambda_m\left(\hat{\mathbf{R}}\right)}{\lambda_m\left(\hat{\mathbf{R}}\right) - z}, \tag{1.89}$$

and

$$\hat{f}(z) = \frac{z}{\hat{w}(z)}, \tag{1.90}$$

respectively. In particular, the expression in (1.89), namely the estimator of the canonical equation in (1.47), will be of special use for the description of our generalized estimators.

The previous asymptotic equivalents are defined for each $z \in \mathsf{C}^+$. For the purposes of establishing our framework for the generalized consistent estimation of spectral functions of $\mathbf{R}$, an extended asymptotic convergence for $z \in \mathsf{R}$ will be needed. Indeed, according to the estimation procedure that we follow (to be described below), we will be interested in evaluating the limiting Stieltjes transform of the spectrum of $\hat{\mathbf{R}}$, noted here as $m_{\hat{\mathsf{R}}}(z)$ (that we recall is expressed in terms of the spectrum of $\mathbf{R}$) at $x \in \mathsf{R}$ outside the limiting eigenvalue support. Essentially, by expressing the searched functions of $\mathbf{R}$ in terms of a limiting Stieltjes transform associated with $\hat{\mathbf{R}}$, an $M, N$-consistent estimator of the original quantity can be immediately found in terms of the SCM. We develop on these ideas later in this section, but, first, consider the following extension of the limit of $m_{\hat{\mathsf{R}}}(z)$ on $\mathsf{R}$ as the solution, for $z = x + \mathrm{j}\,y$, and $y \to 0^+$, to

$$m(x) = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{\lambda_m(\mathbf{R})\left(1 - c - cxm(x)\right) - x}. \tag{1.91}$$

The roots $m = m(x)$ of the polynomial equation in (1.91) can be uniquely defined as follows. If $m \in \mathsf{C}^+$, then the root is unique and $x$ belongs to the interior of the eigenvalue density support. In this case, the strictly positive imaginary part of the limiting solution determines, by Property

(**P7**) in Section 1.1, a nonzero contribution to the eigenvalue density function. Otherwise, all the solutions of $m$ are real, such that $x$ lies outside the support, and there is a unique root such that

$$\frac{1}{M}\sum_{m=1}^{M}\left(\frac{\lambda_m(\mathbf{R})(1-c-cxm(x))}{\lambda_m(\mathbf{R})(1-c-cxm(x))-x}\right)^2 \leq \frac{1}{c}. \tag{1.92}$$

In particular, equality in (1.92) holds for values of $x$ belonging to the boundary of the density clusters defined in Section 1.1.3 from the unique real-valued solution in $f$ to $\Phi'(f)=0$.

Furthermore, regarding the evaluation of $f(z)$ for $z \in \mathsf{R}$, the value of $f(x)$ for $x$ outside the limiting eigenvalue density support can be similarly found as the unique root of the polynomial equation $\Phi(f)=x$, such that $\Phi'(f) \geq 0$, or, equivalently,

$$\frac{1}{M}\sum_{m=1}^{M}\left(\frac{\lambda_m(\mathbf{R})}{\lambda_m(\mathbf{R})-f}\right)^2 \leq \frac{1}{c},$$

with equality holding under the same conditions as discussed above.

For the sake of clarity of exposition, with some abuse of notation, we will denote by $m_{\mathsf{R}}(z)$ and $m_{\hat{\mathsf{R}}}(z)$ the Stieltjes transforms of the spectrum of $\mathbf{R}$ and $\hat{\mathbf{R}}$, respectively. According to the approach proposed by Girko, the spectral function of $\mathbf{R}$ to be estimated is first expressed in terms of the Stieltjes transform of the spectral distribution of $\mathbf{R}$, namely,

$$m_{\mathsf{R}}(z) = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{\lambda_m(\mathbf{R})-z}.$$

Thus, an $M,N$-consistent estimator of the original function can be found by just replacing $m_{\mathsf{R}}(z)$ by its G-estimator. Now, the $M,N$-consistent estimator of $m_{\mathsf{R}}(z)$ is obtained as follows. Observe that, from the limiting Stieltjes transform of $\hat{\mathbf{R}}$ in (1.22), i.e., the solution $m = m_{\hat{\mathsf{R}}}(z)$ for $M,N \to \infty$, with $M/N \to c < +\infty$, to

$$m_{\hat{\mathsf{R}}}(z) = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{\lambda_m(\mathbf{R})\left(1-c-czm_{\hat{\mathsf{R}}}(z)\right)-z},$$

we can write

$$\begin{aligned} m_{\hat{\mathsf{R}}}(z)\left(1-c-czm_{\hat{\mathsf{R}}}(z)\right) &= \frac{1}{M}\sum_{m=1}^{M}\frac{1}{\lambda_m(\mathbf{R})-\frac{z}{1-c-czm_{\hat{\mathsf{R}}}(z)}} \\ &= m_{\mathsf{R}}(\omega)\,|_{\omega=f(z)}, \end{aligned} \tag{1.93}$$

where, as previously defined,

$$f(z) = \frac{z}{1-c-czm_{\hat{\mathsf{R}}}(z)}.$$

Equation (1.93) provides an implicit relationship between the limiting distribution of the true eigenvalues (associated with $m_{\mathsf{R}}(\omega)$) and the asymptotic distribution of the sample eigenvalues (associated with a particular solution of $m_{\hat{\mathsf{R}}}(z)$). Thus, building upon the fact that $m_{\hat{\mathsf{R}}}(z)$ is readily $M,N$-consistently estimated by the Stieltjes transform of the empirical distribution function of the sample eigenvalues, i.e.,

$$m_{\hat{\mathsf{R}}}(z) \asymp \frac{1}{M} \operatorname{Tr}\left[\left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1}\right], \tag{1.94}$$

the idea behind G-estimation is to invert (1.93) in order to express $m_{\mathsf{R}}(\omega)$ in terms of $m_{\hat{\mathsf{R}}}(z)$. Specifically, assume that the equation $f(z) = \omega$ has a single solution in $z$, denoted as $z = f^{-1}(\omega)$. In this case, according to (1.93), $b_{\mathsf{R}}(\omega)$ can be expressed univocally as

$$m_{\mathsf{R}}(\omega) = m_{\hat{\mathsf{R}}}\left(f^{-1}(\omega)\right)\left(1 - c - cf^{-1}(\omega)\, m_{\hat{\mathsf{R}}}\left(f^{-1}(\omega)\right)\right), \tag{1.95}$$

and an $M,N$-consistent estimator of $m_{\mathsf{R}}(\omega)$ can be obtained by directly replacing $m_{\hat{\mathsf{R}}}(z)$ in (1.95) with the RHS of (1.94).

**Example 4** *Consider the estimation of the quantity* $\frac{1}{M}\operatorname{tr}\left[\mathbf{R}^{-1}\right]$. *An estimator simply replacing the true covariance matrix by its sample estimate does not appropriately approximate the quantity in the large observation dimension regime. In order to derive a G-estimation following the previous approach, note that the previous quantity can be expressed as* $m_{\mathsf{R}}(0)$. *Now, the equation* $f(z) = 0$ *has a unique solution (*$z = 0$*) and, consequently, a G-estimator may be constructed using (1.95) as* $m_{\hat{\mathsf{R}}}(z)\left(1 - c - czm_{\hat{\mathsf{R}}}(z)\right)|_{z=f^{-1}(0)=0} = (1-c)\frac{1}{M}\operatorname{tr}\left[\hat{\mathbf{R}}^{-1}\right]$.

**Example 5** *Assume now that the quantity* $M^{\mathsf{R}}_{(-2)} = \frac{1}{M}\operatorname{Tr}\left[\mathbf{R}^{-2}\right]$ *is to be estimated. First, note that the spectral function* $\frac{1}{M}\operatorname{Tr}\left[\mathbf{R}^{-2}\right]$ *can be written in terms of* $m_{\mathsf{R}}(\omega)$ *as*

$$\begin{aligned}
\frac{1}{M}\operatorname{Tr}\left[\mathbf{R}^{-2}\right] &= \frac{1}{M}\sum_{m=1}^{M}\frac{1}{\lambda_m^2(\mathbf{R})} \\
&= \frac{\partial}{\partial\omega}\left\{\frac{1}{M}\sum_{m=1}^{M}\frac{1}{\lambda_m(\mathbf{R}) - \omega}\right\}\Bigg|_{\omega=0} \\
&= \frac{\partial}{\partial\omega}\left\{m_{\mathsf{R}}(\omega)\right\}\Bigg|_{\omega=0}.
\end{aligned}$$

*Thus, as in the previous example, the* $M,N$*-consistent estimator can be constructed replacing* $m_{\mathsf{R}}(\omega)$ *with its G-estimator, i.e.,*

$$\begin{aligned}
M^{\hat{\mathsf{R}}}_{(-2)} &= \frac{\partial}{\partial\omega}\left\{m_{\hat{\mathsf{R}}}\left(\hat{f}^{-1}(\omega)\right)\left(1 - c - cf^{-1}(\omega)\, m_{\hat{\mathsf{R}}}\left(\hat{f}^{-1}(\omega)\right)\right)\right\}\Bigg|_{\omega=0} \\
&= (1-c)^2\frac{1}{M}\operatorname{Tr}\left[\hat{\mathbf{R}}^{-2}\right] - c(1-c)\left(\frac{1}{M}\operatorname{Tr}\left[\hat{\mathbf{R}}^{-1}\right]\right)^2,
\end{aligned}$$

*after applying the chain rule, the inverse function theorem and the fact that* $z = f(z)$ *has a single solution* $z = 0$ *whenever* $c < 1$.

In Chapter 3, a general expression for the $M, N$-consistent estimator of the negative eigenvalue moments of an observation covariance matrix $\mathbf{R}$, i.e., $M_{(-k)}^{\mathsf{R}}$, $k = 1, 2, \ldots$, is derived.

In the case of functions depending on the eigenvectors of the covariance matrix as well, an equivalent approach can be used. In particular, assume that the function to be estimated can be expressed in terms of the Stieltjes transform of a certain vector-valued distribution function of the eigenvalues and eigenvectors of $\mathbf{R}$ (cf. Section 1.1.4), say,

$$\upsilon_{\mathsf{R}}(z) \triangleq \mathbf{a}_1^H (\mathbf{R} - z\mathbf{I}_M)^{-1} \mathbf{a}_2,$$

where, without loss of generality, $\mathbf{a}_1$ and $\mathbf{a}_2$ are two nonrandom vectors with unit Euclidean norm. In this case, the procedure to derive a $M, N$-consistent estimator boils down to finding a G-estimator of $\upsilon_{\mathsf{R}}(\omega)$. Let $\upsilon_{\hat{\mathsf{R}}}(z)$ be defined as $\upsilon_{\hat{\mathsf{R}}}(z) \triangleq \mathbf{a}_1^H \left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1} \mathbf{a}_2$. From Theorem 1.5, we know that

$$\upsilon_{\hat{\mathsf{R}}}(z) \asymp \upsilon_{\mathsf{R}}(z),$$

or, equivalently,

$$\upsilon_{\hat{\mathsf{R}}}(z) \asymp \frac{f(z)}{z} \mathbf{a}_1^H (\mathbf{R} - f(z)\mathbf{I}_M)^{-1} \mathbf{a}_2,$$

as $M, N \to \infty$ with $M/N \to c < +\infty$. Thus, in order to estimate $\upsilon_{\mathsf{R}}(\omega)$, we can alternatively address the estimation of

$$\upsilon_{\mathsf{R}}(\omega) = \upsilon_{\mathsf{R}}(z)(1 - c - czm_{\mathsf{R}}(z)),$$

with $\omega = f(z)$. Consequently, if $\omega = f(z)$ has a unique solution, we have

$$\upsilon_{\mathsf{R}}(\omega) \asymp \upsilon_{\hat{\mathsf{R}}}(z)\left(1 - c - czm_{\hat{\mathsf{R}}}(z)\right)\Big|_{z=\hat{f}^{-1}(\omega)},$$

so that the RHS of () is then an $M, N$-consistent estimator of $\upsilon_{\mathsf{R}}(\omega)$.

**Example 6** *Assume that we need to estimate the quantity $\varphi(\mathbf{R}) = \left\{\mathbf{R}^{-1}\right\}_{i,j}$. First, observe that $\varphi(\mathbf{R})$ can be expressed as a function of $\upsilon_{\mathsf{R}}(\omega)$, with $\mathbf{a}_1 = \mathbf{e}_i$ and $\mathbf{a}_2 = \mathbf{e}_j$, as*

$$\begin{aligned}
\left\{\mathbf{R}^{-1}\right\}_{i,j} &= \mathbf{e}_i^H \mathbf{R}^{-1} \mathbf{e}_j \\
&= \sum_{m=1}^{M} \frac{\mathbf{e}_i^H \mathbf{q}_m(\mathbf{R}) \mathbf{q}_m^H(\mathbf{R}) \mathbf{e}_j}{\lambda_m(\mathbf{R})} \\
&= \sum_{m=1}^{M} \frac{\mathbf{e}_i^H \mathbf{q}_m(\mathbf{R}) \mathbf{q}_m^H(\mathbf{R}) \mathbf{e}_j}{\lambda_m(\mathbf{R}) - \omega}\Bigg|_{\omega=0} \\
&= \upsilon_{\mathsf{R}}(\omega)\big|_{\omega=0},
\end{aligned}$$

*where $\mathbf{e}_m$ is the $M$-dimensional unit-vector with $m$th element 1 and all others 0. Now, the $M, N$-consistent estimator can be constructed replacing $\upsilon_{\mathsf{R}}(\omega)$ with its G-estimator, i.e.,*

$$\begin{aligned}
\hat{\varphi}(\mathbf{R}) &= \upsilon_{\hat{\mathsf{R}}}(z)\left(1 - c - czm_{\hat{\mathsf{R}}}(z)\right)\Big|_{z=\hat{f}^{-1}(0)} \\
&= (1 - c)\,\upsilon_{\hat{\mathsf{R}}}(0) \\
&= (1 - c)\left\{\hat{\mathbf{R}}^{-1}\right\}_{i,j},
\end{aligned}$$

*after using the fact that $z = f(z)$ has a single solution $z = 0$ whenever $c < 1$.*

**Example 7** *(Whitening filter) Consider the estimation of the quantity $\varphi(\mathbf{R}) = \left\{\mathbf{R}^{-1/2}\right\}_{i,j}$. Using the following integral representaiton of the square-root, namely,*

$$\frac{1}{\sqrt{x}} = \frac{2}{\pi}\int_0^\infty \frac{1}{x+t^2}dt, \quad x > 0,$$

*the quantity $\varphi(\mathbf{R})$ can be expressed in terms of $\upsilon_{\mathsf{R}}(\omega)$, with $\mathbf{a}_1 = \mathbf{e}_i$ and $\mathbf{a}_2 = \mathbf{e}_j$, as*

$$\begin{aligned}\left\{\mathbf{R}^{-1}\right\}_{i,j} &= \frac{2}{\pi}\int_0^\infty \mathbf{e}_i^H\left(\mathbf{R}+t^2\mathbf{I}_M\right)\mathbf{e}_j\ dt \\ &= \frac{2}{\pi}\int_0^\infty \upsilon_{\mathsf{R}}\left(-t^2\right)\ dt.\end{aligned}$$

*If $c < 1$, observe that $f(z) = -t^2$ has a unique solution for any $t \in \mathsf{R}$. Hence, the $M,N$-consistent estimator can be constructed replacing $\upsilon_{\mathsf{R}}(\omega)$ with its G-estimator, i.e.,*

$$\begin{aligned}\hat{\varphi}(\mathbf{R}) &= \frac{2}{\pi}\int_0^\infty \upsilon_{\hat{\mathsf{R}}}(z)\left(1-c-czm_{\hat{\mathsf{R}}}(z)\right)\Big|_{z=\hat{f}^{-1}(-t^2)}\ dt \\ &\frac{2}{\pi}\int_0^\infty \sum_{m=1}^{M}\frac{\mathbf{e}_i^H\mathbf{q}_m\left(\hat{\mathbf{R}}\right)\mathbf{q}_m^H\left(\hat{\mathbf{R}}\right)\mathbf{e}_j}{\lambda_m\left(\hat{\mathbf{R}}\right)-\hat{f}^{-1}\left(-t^2\right)}\left(1-c-cz\frac{1}{M}\sum_{k=1}^{M}\frac{1}{\lambda_k\left(\hat{\mathbf{R}}\right)-\hat{f}^{-1}\left(-t^2\right)}\right)\ dt.\end{aligned}$$

The applicability of this approach is limited to situations in which the function $f(z) = \omega$ is invertible. Unfortunately, this equation does not happen to be invertible in many cases where, in fact, the equation might not have a solution at all. In this dissertation, we will approach the construction of $M,N$-consistent estimators of general vector-valued spectral functions of $\mathbf{R}$ appearing in optimum parameter estimation procedures in statistical signal processing by resorting to the characterization of the limiting Stieltjes transform of the SCM spectrum, without relying on an explicit function inversion.

## Appendix 1.A Proof of Proposition 1.1

In order to prove Proposition 1.1, we follow the lines of the proof of [Doz07b, Theorem 1.1]. Consider a matrix $\mathbf{\Xi} = [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_N] \in \mathbb{C}^{M \times N}$ such that $\mathbf{N} = \sigma^2 \mathbf{\Xi}$ and let us define $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_N]$. Furthermore, consider also a random matrix $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]$ such that

$$\hat{\mathbf{R}} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H = \hat{\mathbf{R}}_i + \frac{1}{N}\mathbf{y}_i\mathbf{y}_i^H,$$

where the matrix $\hat{\mathbf{R}}_i$ is defined as

$$\hat{\mathbf{R}}_i = \hat{\mathbf{R}} - \frac{1}{N}\mathbf{y}_i\mathbf{y}_i^H.$$

For the sake of notational convenience, we will use the following definitions in our derivations

$$\begin{aligned}
\mathbf{Q}(z) &= \left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1}, \\
\mathbf{Q}_i(z) &= \left(\hat{\mathbf{R}}_i - z\mathbf{I}_M\right)^{-1}, \\
\mathbf{P}(z) &= \left(\mathbf{\Omega} - z\mathbf{I}_M\right)^{-1}, \\
\mathbf{P}_i(z) &= \left(\mathbf{\Omega}_i - z\mathbf{I}_M\right)^{-1}.
\end{aligned}$$

where we have defined

$$\begin{aligned}
\mathbf{\Omega} &= \frac{1}{1+\sigma^2 c m(z)}\mathbf{\Psi} - \sigma^2\left(1 - c - czm(z)\right), \\
\mathbf{\Omega}_n &= \frac{1}{1+\sigma^2 c m_i(z)}\mathbf{\Psi} - \sigma^2\left(1 - c - czm_i(z)\right),
\end{aligned}$$

with $m_i(z)$ being defined in terms of $\hat{\mathbf{R}}_i$. As in the proof of [Doz07b, Theorem 1.1], we proceed by factoring the difference of inverses and expanding the middle factor as

$$\begin{aligned}
\mathbf{P}(z) - \mathbf{Q}(z) &= \mathbf{P}(z)\left(\hat{\mathbf{R}} - \mathbf{A}\right)\mathbf{Q}(z) \\
&= \mathbf{P}(z)\left(\frac{\sigma^2 c m_i(z)}{1+\sigma^2 c m_i(z)}\frac{1}{N}\mathbf{S}\mathbf{S}^H + \frac{1}{N}\sigma^2\mathbf{\Xi}\mathbf{S}^H + \frac{1}{N}\sigma^2\mathbf{S}\mathbf{\Xi}^H + \frac{1}{N}\sigma^2\mathbf{\Xi}\mathbf{\Xi}^H + \sigma^2 w(z)\right)\mathbf{Q}(z) \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbf{P}(z)\left(\frac{\sigma^2 c m_i(z)}{1+\sigma^2 c m_i(z)}\mathbf{s}_i\mathbf{s}_i^H + \sigma^2\boldsymbol{\xi}_i\mathbf{s}_i^H + \sigma^2\mathbf{s}_i\boldsymbol{\xi}_i^H + \sigma^2\boldsymbol{\xi}_i\boldsymbol{\xi}_i^H + \sigma^2 w(z)\right)\mathbf{Q}(z),
\end{aligned}$$

where $w(z) = 1 - c - czm(z)$, and, in the first equality, we have used the resolvent identity, i.e., $\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{B}^{-1}(\mathbf{A} - \mathbf{B})\mathbf{A}^{-1}$. Further, we apply the vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ and apply the matrix inversion lemma as

$$\mathbf{Q}(z) = \mathbf{Q}_i(z) - \frac{1}{N}\frac{\mathbf{Q}_i(z)\mathbf{y}_i\mathbf{y}_i^H\mathbf{Q}_i(z)}{1+\frac{1}{N}\mathbf{y}_i^H\mathbf{Q}_i(z)\mathbf{y}_i}, \tag{1.96}$$

to obtain

$$\begin{aligned}
\mathbf{a}_1^H\left(\mathbf{P}(z) - \mathbf{Q}(z)\right)\mathbf{a}_2 &= \mathbf{a}_1^H\left(\mathbf{\Psi} - z\mathbf{I}_M\right)^{-1}\mathbf{a}_2 - \mathbf{a}_1^H\left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1}\mathbf{a}_2 \\
&= \frac{1}{N}\sum_{i=1}^{N} W_i^{(1)} + W_i^{(2)} + W_i^{(3)} + W_i^{(4)} + W_i^{(5)},
\end{aligned}$$

where

$$W_i^{(1)} = \frac{1}{\alpha_i}\left(\frac{\sigma^2 cm(z)}{1+\sigma^2 cm(z)}\right)\left[(1+\gamma_i+\omega_i)\hat{\rho}_i - (\rho_i+\beta_i)\hat{\gamma}_i\right],$$
$$W_i^{(2)} = \frac{1}{\alpha_i}\left[(1+\gamma_i+\omega_i)\hat{\beta}_i - (\rho_i+\beta_i)\hat{\omega}_i\right],$$
$$W_i^{(3)} = \frac{1}{\alpha_i}\left[(1+\rho_i+\beta_i)\hat{\gamma}_i - (\gamma_i+\omega_i)\hat{\gamma}_i\right],$$
$$W_i^{(4)} = \frac{1}{\alpha_i}\left[(1+\rho_i+\beta_i)\hat{\omega}_i - (\gamma_i+\omega_i)\hat{\beta}_i\right],$$
$$W_i^{(5)} = \frac{1}{\alpha_i}\sigma^2\mathbf{a}_1^H\mathbf{P}(z)\mathbf{Q}(z)\mathbf{a}_2,$$

with the following notation, namely,

$$\rho_i = \tfrac{1}{N}\mathbf{s}_i^H\mathbf{Q}_i(z)\mathbf{s}_i, \qquad \hat{\rho}_i = \mathbf{a}_1^H\mathbf{P}(z)\mathbf{s}_i\mathbf{s}_i^H\mathbf{Q}_i(z)\mathbf{a}_2,$$
$$\omega_i = \sigma^2\tfrac{1}{N}\boldsymbol{\xi}_i^H\mathbf{Q}_i(z)\boldsymbol{\xi}_i, \hat{\omega}_i = \sigma^2\mathbf{a}_1^H\mathbf{P}(z)\boldsymbol{\xi}_i\boldsymbol{\xi}_i^H\mathbf{Q}_i(z)\mathbf{a}_2,$$
$$\beta_i = \sigma^2\tfrac{1}{N}\mathbf{s}_i^H\mathbf{Q}_i(z)\boldsymbol{\xi}_i, \ \hat{\beta}_i = \sigma^2\mathbf{a}_1^H\mathbf{P}(z)\boldsymbol{\xi}_i\mathbf{s}_i^H\mathbf{Q}_i(z)\mathbf{a}_2,$$
$$\gamma_i = \sigma^2\tfrac{1}{N}\boldsymbol{\xi}_i^H\mathbf{Q}_i(z)\mathbf{s}_i, \ \hat{\gamma}_i = \sigma^2\mathbf{a}_1^H\mathbf{P}(z)\mathbf{s}_i\boldsymbol{\xi}_i^H\mathbf{Q}_i(z)\mathbf{a}_2,$$

as well as

$$\alpha_i = 1 + \frac{1}{N}\mathbf{y}_i^H\mathbf{Q}_i(z)\mathbf{y}_i,$$

Furthermore, after simplification, we get

$$\mathbf{a}_1^H(\boldsymbol{\Psi} - z\mathbf{I}_M)^{-1}\mathbf{a}_2 - \mathbf{a}_1^H\left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1}\mathbf{a}_2 = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\alpha_i}\left(\hat{\omega}_i - W_i^{(5)}\right) \tag{1.97}$$
$$+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\alpha_i}\frac{1}{1+\sigma^2 cm(z)}\left(\sigma^2 cm(z) - \omega_i - \gamma_i\right)\hat{\rho}_i \tag{1.98}$$
$$+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\alpha_i}\left[\frac{1}{1+\sigma^2 cm(z)}(\rho_i+\beta_i)\hat{\gamma}_i + \hat{\gamma}_i\right] \tag{1.99}$$
$$+\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\alpha_i}\hat{\beta}_i, \tag{1.100}$$

Hence, in order to prove the result in (1.71), it is enough to prove that the quantities in (1.97)-(1.100) vanish almost surely. This is afforded in the following. Before proceeding further, note from Lemma A.8 the following bounds (see [Doz07b] for further details), namely, for any $i = 1, 2, \ldots, N$ and any $M$,

$$\frac{1}{|\alpha_i|} \le \frac{|z|}{\operatorname{Im}\{z\}}, \quad \frac{1}{|b(z)|} \le \frac{|z|}{\operatorname{Im}\{z\}} \tag{1.101}$$

and

$$\|\mathbf{Q}_i(z)\| \le \frac{1}{\operatorname{Im}\{z\}}, \quad \|\mathbf{P}(z)\| \le \frac{1}{\operatorname{Im}\{z\}}, \quad \|\mathbf{P}_i(z)\| \le \frac{1}{\operatorname{Im}\{z\}}, \tag{1.102}$$

**Proof of (1.97).** Let us first rewrite (1.97) as

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\alpha_i}\left(\hat{\omega}_i - W_i^{(5)}\right) = \frac{1}{N}\sum_{i=1}^{N}\frac{\sigma^2}{\alpha_i}\mathbf{a}_1^H\left(\mathbf{P}\left(z\right)-\mathbf{P}_i\left(z\right)\right)\boldsymbol{\xi}_i\boldsymbol{\xi}_i{}^H\mathbf{Q}_i\left(z\right)\mathbf{a}_2 \tag{1.103}$$

$$+\frac{1}{N}\sum_{i=1}^{N}\frac{\sigma^2}{\alpha_i}\left(\mathbf{a}_1^H\mathbf{P}_i\left(z\right)\boldsymbol{\xi}_i\boldsymbol{\xi}_i{}^H\mathbf{Q}_i\left(z\right)\mathbf{a}_2 - \mathbf{a}_1^H\mathbf{P}_i\left(z\right)\mathbf{Q}_i\left(z\right)\mathbf{a}_2\right) \tag{1.104}$$

$$+\frac{1}{N}\sum_{i=1}^{N}\frac{\sigma^2}{\alpha_i}\mathbf{a}_1^H\left(\mathbf{P}_i\left(z\right)-\mathbf{P}\left(z\right)\right)\mathbf{Q}_i\left(z\right)\mathbf{a}_2 \tag{1.105}$$

$$+\frac{1}{N}\sum_{i=1}^{N}\frac{\sigma^2}{\alpha_i}\mathbf{a}_1^H\mathbf{P}\left(z\right)\left(\mathbf{Q}_i\left(z\right)-\mathbf{Q}\left(z\right)\right)\mathbf{a}_2. \tag{1.106}$$

According to Lemma A.2 and from the absolutely boundness of $\sigma^2/\alpha_i$ (cf. Equation (1.101)), in order to prove the almost sure convergence to zero of (1.103) it is enough to show that

$$\max_{1\leq i\leq N}\mathsf{E}\left[\left|\mathbf{a}_1^H\left(\mathbf{P}\left(z\right)-\mathbf{P}_i\left(z\right)\right)\boldsymbol{\xi}_i\boldsymbol{\xi}_i{}^H\mathbf{Q}_i\left(z\right)\mathbf{a}_2\right|^p\right]\leq\frac{C}{N^{1+\delta}},$$

for some constants $C$, $\delta > 0$ and $p > 1$ not depending on $N$. To that effect, we can use Lemma A.3 with $\mathbf{C} = \mathbf{Q}_i\left(z\right)\mathbf{a}_2\mathbf{a}_1^H\left(\mathbf{P}\left(z\right)-\mathbf{P}_i\left(z\right)\right)$ as

$$\begin{aligned}\mathsf{E}\left[\left|\mathbf{u}_n^H\mathbf{C}\mathbf{u}_n - \mathrm{Tr}\left[\mathbf{C}\right]\right|^p\right] &\leq K_p\left(\mathsf{E}^{p/2}\left[|\xi|^4\right]+\mathsf{E}\left[|\xi|^{2p}\right]\right)\left(\mathbf{a}_2^H\mathbf{Q}_i^2\left(z\right)\mathbf{a}_2\right)^{p/2}\left(\mathbf{a}_1^H\left(\mathbf{P}\left(z\right)-\mathbf{P}_i\left(z\right)\right)^2\mathbf{a}_1\right)^{p/2}\\ &\leq C\left(\mathsf{E}^{p/2}\left[|\xi|^4\right]+\mathsf{E}\left[|\xi|^{2p}\right]\right)\frac{\|\mathbf{a}_2\|^p}{\mathrm{Im}^p\left\{z\right\}}\frac{\|\mathbf{a}_1\|^p}{N^p}.\end{aligned}$$

Hence, convergence is proved by choosing $p \geq 2$. Furthermore, note that similar reasoning can be used to show that (1.105) vanishes almost surely. On the other hand, Lemma (A.5) can be directly applied with $\mathbf{U}_i = \mathbf{Q}_i\left(z\right)$ and $\mathbf{C} = \mathbf{a}_2\mathbf{a}_1^H\mathbf{P}_i\left(z\right)$ in order to prove the convergence of (1.104). Regarding (1.106), using the Sherman-Morrison inversion formula as in (1.96), we can write

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\sigma^2}{\alpha_i^2}\frac{1}{N}\mathbf{a}_1^H\mathbf{P}\left(z\right)\mathbf{Q}_i(z)\mathbf{y}_i\mathbf{y}_i^H\mathbf{Q}_i(z)\mathbf{a}_2 = \frac{1}{N}\sum_{i=1}^{N}\frac{\sigma_n^4}{\alpha_i^2}\frac{1}{N}\mathbf{a}_1^H\mathbf{P}\left(z\right)\mathbf{Q}_i(z)\boldsymbol{\xi}_i\boldsymbol{\xi}_i{}^H\mathbf{Q}_i(z)\mathbf{a}_2 \tag{1.107}$$

$$+\frac{1}{N}\sum_{i=1}^{N}\frac{\sigma_n^3}{\alpha_i^2}\frac{1}{N}\mathbf{a}_1^H\mathbf{P}\left(z\right)\mathbf{Q}_i(z)\boldsymbol{\xi}_i\mathbf{s}_i^H\mathbf{Q}_i(z)\mathbf{a}_2 \tag{1.108}$$

$$+\frac{1}{N}\sum_{i=1}^{N}\frac{\sigma_n^3}{\alpha_i^2}\frac{1}{N}\mathbf{a}_1^H\mathbf{P}\left(z\right)\mathbf{Q}_i(z)\mathbf{s}_i\boldsymbol{\xi}_i{}^H\mathbf{Q}_i(z)\mathbf{a}_2 \tag{1.109}$$

$$+\frac{1}{N}\sum_{i=1}^{N}\frac{\sigma^2}{\alpha_i^2}\frac{1}{N}\mathbf{a}_1^H\mathbf{P}\left(z\right)\mathbf{Q}_i(z)\mathbf{s}_i\mathbf{s}_i^H\mathbf{Q}_i(z)\mathbf{a}_2. \tag{1.110}$$

Now, (1.107) tends clearly to zero by Lemma A.2 with $p \geq 2$ since, using Lemma A.4 with $\mathbf{A} = \mathbf{Q}_i(z)\mathbf{a}_2\mathbf{a}_1^H\mathbf{P}\left(z\right)\mathbf{Q}_i(z)$, we note that

$$\mathsf{E}\left[\left|\mathbf{a}_1^H\mathbf{P}\left(z\right)\mathbf{Q}_i(z)\boldsymbol{\xi}_i\boldsymbol{\xi}_i{}^H\mathbf{Q}_i(z)\mathbf{a}_2\right|^2\right] < +\infty.$$

Regarding (1.108) and, equivalently, (1.109), we can use the Cauchy-Schwarz inequality to write $(p = 2q)$

$$\frac{1}{N^p}\mathsf{E}\left[\left|\mathbf{a}_1^H\mathbf{P}(z)\mathbf{Q}_i(z)\boldsymbol{\xi}_i\mathbf{s}_i^H\mathbf{Q}_i(z)\mathbf{a}_2\right|^p\right] \leq \frac{1}{N^p}\mathsf{E}^{1/2}\left[\left|\boldsymbol{\xi}_i{}^H\mathbf{Q}_i(z)\mathbf{P}(z)\mathbf{a}_1\mathbf{a}_1^H\mathbf{P}(z)\mathbf{Q}_i(z)\boldsymbol{\xi}_i\right|^p\right]\left(\frac{\|\mathbf{a}_2\|^p}{\mathrm{Im}^p\{z\}}N^q\right).$$

Thus, using Lemma A.4 with $\mathbf{A} = \mathbf{Q}_i(z)\mathbf{P}(z)\mathbf{a}_1\mathbf{a}_1^H\mathbf{P}(z)\mathbf{Q}_i(z)$, the result can be proved by choosing $p > 2$. Finally, considering again (1.96) and (1.101), we may prove that (1.110) vanish almost surely by equivalently proving the following, namely,

$$\left|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\mathbf{a}_1^H\mathbf{P}(z)\left(\mathbf{Q}(z)+\frac{1}{N}\mathbf{Q}_i(z)\mathbf{y}_i\mathbf{y}_i^H\mathbf{Q}_i(z)\right)\mathbf{s}_i\mathbf{s}_i^H\left(\mathbf{Q}(z)+\frac{1}{N}\mathbf{Q}_i(z)\mathbf{y}_i\mathbf{y}_i^H\mathbf{Q}_i(z)\right)\mathbf{a}_2\right| \to 0.$$

To that effect, it is enough to show

$$\left|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\mathbf{a}_1^H\mathbf{P}(z)\mathbf{Q}(z)\mathbf{s}_i\mathbf{s}_i^H\mathbf{Q}(z)\mathbf{a}_2\right| \to 0, \tag{1.111}$$

since the result for the other three terms follows by using similar arguments as previously in the proof of the convergence of (1.107) to (1.109). Thus, observe that

$$\mathsf{E}\left[\left|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\mathbf{a}_1^H\mathbf{P}(z)\mathbf{Q}(z)\mathbf{s}_i\mathbf{s}_i^H\mathbf{Q}(z)\mathbf{a}_2\right|^p\right] = \frac{1}{N^p}\mathsf{E}\left[\left|\frac{1}{N}\mathbf{a}_1^H\mathbf{P}(z)\mathbf{Q}(z)\mathbf{S}\mathbf{S}^H\mathbf{Q}(z)\mathbf{a}_2\right|^p\right] \leq \frac{\|\mathbf{a}_1\|^p\|\mathbf{a}_2\|^p}{\mathrm{Im}\{z\}}.$$

Thus, the convergence follows from the Borel-Cantelli lemma by choosing $p \geq 2$.

**Proof of (1.98).** From (1.101) and

$$\left|\frac{1}{N}\mathbf{s}_i^H\mathbf{Q}_i(z)\mathbf{s}_i\right| \leq \frac{C}{\mathrm{Im}\{z\}},$$

we just need to prove that, almost surely,

$$\left|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\mathbf{s}_i^H\mathbf{Q}_i(z)\boldsymbol{\xi}_i\mathbf{a}_1^H\mathbf{P}(z)\mathbf{s}_i\boldsymbol{\xi}_i{}^H\mathbf{Q}_i(z)\mathbf{a}_2\right| \to 0, \tag{1.112}$$

and

$$\left|\frac{1}{N}\sum_{i=1}^{N}\mathbf{a}_1^H\mathbf{P}(z)\mathbf{s}_i\boldsymbol{\xi}_i{}^H\mathbf{Q}_i(z)\mathbf{a}_2\right| \to 0. \tag{1.113}$$

Let us first focus on (1.112), which can be clearly reduced to proving the following, namely,

$$\left|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\mathbf{s}_i^H\mathbf{Q}_i(z)\boldsymbol{\xi}_i\boldsymbol{\xi}_i{}^H\mathbf{Q}_i(z)\mathbf{a}_2\mathbf{a}_1^H\mathbf{P}(z)\mathbf{s}_i - \frac{1}{N}\mathbf{s}_i^H\mathbf{Q}_i^2(z)\mathbf{a}_2\mathbf{a}_1^H\mathbf{P}(z)\mathbf{s}_i\right| \to 0, \tag{1.114}$$

$$\left|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\mathbf{s}_i^H\mathbf{Q}_i^2(z)\mathbf{a}_2\mathbf{a}_1^H\mathbf{P}(z)\mathbf{s}_i\right| \to 0. \tag{1.115}$$

Now, the result in (1.114) is direct by Lemma A.5, whereas (1.115) can be proved similarly as the convergence of (1.110).

On the other hand, regarding the result in (1.113), let us first define $\eta_k = \mathbf{a}_1^H \mathbf{P}(z) \mathbf{s}_k \boldsymbol{\xi}_k{}^H \mathbf{Q}_k(z) \mathbf{a}_2$, and consider the filtration $\{\mathcal{F}_k\}$ generated by the random vectors $\{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k\}$. Now, we observe that the process $\{\eta_k\}$ is a martingale difference sequence adapted to $\{\mathcal{F}_k\}$. Then, we can use the Burkholder inequality and note that, on the one hand, applying the Cauchy-Schwarz inequality, we get

$$\mathsf{E}\left[|\eta_k|^2 |\mathcal{F}_{k-1}\right] \leq \mathsf{E}^{1/2}\left[\left|\mathbf{s}_k^H \mathbf{P}(z) \mathbf{a}_1 \mathbf{a}_1^H \mathbf{P}(z) \mathbf{s}_k\right|^2 |\mathcal{F}_{k-1}\right] \mathsf{E}^{1/2}\left[\left|\boldsymbol{\xi}_k{}^H \mathbf{Q}_k(z) \mathbf{a}_2 \mathbf{a}_2^H \mathbf{Q}_k(z) \boldsymbol{\xi}_k\right|^2 |\mathcal{F}_{k-1}\right],$$

which can be readily shown to be bounded by Lemma A.4. On the other hand, similarly,

$$\mathsf{E}\left[\left|\mathbf{a}_1^H \mathbf{P}(z) \mathbf{s}_k \boldsymbol{\xi}_k{}^H \mathbf{Q}_k(z) \mathbf{a}_2\right|^p\right] \leq \mathsf{E}^{1/2}\left[\left|\mathbf{s}_k^H \mathbf{P}(z) \mathbf{a}_1 \mathbf{a}_1^H \mathbf{P}(z) \mathbf{s}_k\right|^p\right] \mathsf{E}^{1/2}\left[\left|\boldsymbol{\xi}_k{}^H \mathbf{Q}_k(z) \mathbf{a}_2 \mathbf{a}_2^H \mathbf{Q}_k(z) \boldsymbol{\xi}_k\right|^p\right] < +\infty$$

Then, we finally have that

$$\mathsf{E}\left[\left|\frac{1}{N}\sum_{k=1}^{N} \eta_k\right|^p\right] \leq \frac{C'}{N^{p/2}} + \frac{C''}{N^{p-1}},$$

so that the result can be readily proved by applying the Borel-Cantelli lemma with $p > 2$.

**Proof of (1.99) and (1.100).** Considering (1.101), we prove separately

$$\left|\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\boldsymbol{\xi}_i{}^H \mathbf{Q}_i(z) \boldsymbol{\xi} - \frac{1}{N}\operatorname{Tr}[\mathbf{Q}(z)]\right) \mathbf{a}_1^H \mathbf{P}(z) \mathbf{s}_i \mathbf{s}_i^H \mathbf{Q}_i(z) \mathbf{a}_2\right| \to 0, \tag{1.116}$$

and

$$\left|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{N}\boldsymbol{\xi}_i{}^H \mathbf{Q}_i(z) \mathbf{s}_i \mathbf{a}_1^H \mathbf{P}(z) \mathbf{s}_i \mathbf{s}_i^H \mathbf{Q}_i(z) \mathbf{a}_2\right| \to 0. \tag{1.117}$$

Regarding (1.116), we note that $\mathsf{E}\left[\left|\mathbf{a}_1^H \mathbf{P}(z) \mathbf{s}_i \mathbf{s}_i^H \mathbf{Q}_i(z) \mathbf{a}_2\right|^p\right] < +\infty$ and check

$$\max_{1\leq i\leq N} \mathsf{E}\left[\left|\frac{1}{N}\boldsymbol{\xi}_i{}^H \mathbf{Q}_i(z) \boldsymbol{\xi} - \frac{1}{N}\operatorname{Tr}[\mathbf{Q}(z)]\right|^p\right] \leq \frac{C}{N^{1+\delta}}, \tag{1.118}$$

for some constants $C$, $\delta > 0$ and $p > 1$ not depending on $N$. Indeed, (1.118) holds and is shown in [Doz07b, pp. 688]. To see that, observe that we can write the expectation as

$$\mathsf{E}\left[\left|\frac{1}{N}\boldsymbol{\xi}_i{}^H \mathbf{Q}_i(z) \boldsymbol{\xi} - \frac{1}{N}\operatorname{Tr}[\mathbf{Q}(z)]\right|^p\right] \leq \frac{1}{N^p}\left\{\mathsf{E}\left[\left|\boldsymbol{\xi}_i{}^H \mathbf{Q}_i(z) \boldsymbol{\xi} - \operatorname{Tr}[\mathbf{Q}_i(z)]\right|^p\right] + \mathsf{E}\left[\left|\operatorname{Tr}[\mathbf{Q}(z) - \mathbf{Q}_i(z)]\right|^p\right]\right\}. \tag{1.119}$$

The fist term in the RHS of (1.119) can is bounded by Lemma A.3, whereas for the second one we may use the fact, from Lemma A.9, that

$$\left|\operatorname{Tr}\left[\left(\hat{\mathbf{R}}_i - z\mathbf{I}_M\right)^{-1} - \left(\hat{\mathbf{R}}_i + \frac{1}{N}\mathbf{y}_i\mathbf{y}_i^H - z\mathbf{I}_M\right)^{-1}\right]\right| \leq \left|\frac{\frac{1}{N}\mathbf{y}_i^H\left(\hat{\mathbf{R}}_i - z\mathbf{I}_M\right)^{-2}\mathbf{y}_i}{1 + \frac{1}{N}\mathbf{y}_i^H\left(\hat{\mathbf{R}}_i - z\mathbf{I}_M\right)^{-1}\mathbf{y}_i}\right| \leq \frac{1}{\operatorname{Im}\{z\}}.$$

Hence, the convergence in (1.116) can be proved by using $p \geq 2$. On the other hand, regarding (1.117), using again $\mathsf{E}\left[\left|\mathbf{a}_1^H \mathbf{P}(z)\mathbf{s}_i\mathbf{s}_i^H \mathbf{Q}_i(z)\mathbf{a}_2\right|^p\right] < +\infty$, we just have to show that $(p = 2q)$

$$\max_{1\leq i\leq N} \frac{1}{N^q}\mathsf{E}\left[\left|\frac{1}{N}\boldsymbol{\xi}_i{}^H \mathbf{Q}_i(z)\mathbf{s}_i\mathbf{s}_i^H \mathbf{Q}_i(z)\boldsymbol{\xi}_i\right|^q\right] \leq \frac{C}{N^{1+\delta}},$$

for some constants $C$, $\delta > 0$ and $p > 1$ not depending on $N$, which is also proved in [Doz07b, pp. 689]. Indeed, the expectation is bounded by Lemma A.3 for $q \geq 2$, and so is the result in (1.117) proved according to Lemma A.2 by choosing $p \geq 4$.

## Appendix 1.B Proof of Corollary 1.3

The result in (1.77) follows directly from Proposition 1.1, since, by analytic continuation, the asymptotic limit of the Stieltjes transform can be evaluated at any $x \in \mathsf{R} - \{0\}$ (see also [Doz07a, Theorem 2.1]).

On the other hand, in order to obtain (1.78), let us first define

$$b(z) = 1 + \sigma^2 cm(z), \tag{1.120}$$

and

$$t(z) = b(z)\left(\sigma^2(1-c) - zb(z)\right). \tag{1.121}$$

Then, noting that

$$\mathbf{u}_1^H\left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-2}\mathbf{u}_1\Big|_{z=\xi} = -\frac{\partial}{\partial z}\left\{\mathbf{u}_1^H\left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)_k^{-1}\mathbf{u}_1\right\}\Big|_{z=\xi},$$

we just need to differente the limiting Stieltjes transform as

$$m'(z) = b'(z)\frac{1}{M}\operatorname{Tr}\left[(\boldsymbol{\Psi} + t(z)\mathbf{I}_M)^{-1}\right] - b(z)t'(z)\frac{1}{M}\operatorname{Tr}\left[(\boldsymbol{\Psi} + t(z)\mathbf{I}_M)^{-2}\right].$$

On the one hand, we have

$$b'(z) = cm'(z),$$

and

$$\begin{aligned} t'(z) &= b'(z)(1 - c - zb(z)) - b(z)\left(b(z) + zb'(z)\right) \\ &= b'(z)(1 - c - 2zb(z)) - b^2(z). \end{aligned}$$

On the other hand, from (1.120), we have that

$$\begin{aligned} b'(z) &= b'(z)\frac{1}{N}\operatorname{Tr}\left[(\boldsymbol{\Psi} + t(z)\mathbf{I}_M)^{-1}\right] - t'(z)b(z)\frac{1}{N}\operatorname{Tr}\left[(\boldsymbol{\Psi} + t(z)\mathbf{I}_M)^{-2}\right] \\ &= b'(z)\frac{1}{N}\operatorname{Tr}\left[(\boldsymbol{\Psi} + t(z)\mathbf{I}_M)^{-1}\right] - \left(b'(z)(1 - c - 2zb(z)) - b^2(z)\right)b(z)\frac{1}{N}\operatorname{Tr}\left[(\boldsymbol{\Psi} + t(z)\mathbf{I}_M)^{-2}\right] \\ &= b'(z)\left(\frac{1}{N}\operatorname{Tr}\left[(\boldsymbol{\Psi} + t(z)\mathbf{I}_M)^{-1}\right] - b(z)(1 - c - 2zb(z))\frac{1}{N}\operatorname{Tr}\left[(\boldsymbol{\Psi} + t(z)\mathbf{I}_M)^{-2}\right]\right) \\ &\quad + b^3(z)\frac{1}{N}\operatorname{Tr}\left[(\boldsymbol{\Psi} + t(z)\mathbf{I}_M)^{-2}\right]. \end{aligned}$$

Therefore, we can write

$$b'(z) = \frac{b^3(z)\frac{1}{N}\operatorname{Tr}\left[(\boldsymbol{\Psi}+t(z)\mathbf{I}_M)^{-2}\right]}{1-\frac{1}{N}\operatorname{Tr}\left[(\boldsymbol{\Psi}+t(z)\mathbf{I}_M)^{-1}\right]+b(z)(1-c-2zb(z))\frac{1}{N}\operatorname{Tr}\left[(\boldsymbol{\Psi}+t(z)\mathbf{I}_M)^{-2}\right]},$$

so that,

$$m'(z) = \frac{b^3(z)\frac{1}{M}\operatorname{Tr}\left[(\boldsymbol{\Psi}+t(z)\mathbf{I}_M)^{-2}\right]}{1-\frac{1}{N}\operatorname{Tr}\left[(\boldsymbol{\Psi}+t(z)\mathbf{I}_M)^{-1}\right]+b(z)(1-c-2zb(z))\frac{1}{N}\operatorname{Tr}\left[(\boldsymbol{\Psi}+t(z)\mathbf{I}_M)^{-2}\right]}.$$

Consequently, we finally get

$$\mathbf{v}^H\left(\hat{\mathbf{R}}-z\mathbf{I}_M\right)^{-2}\mathbf{v}\Big|_{z=\xi} = \zeta(z)\,\mathbf{u}_1^H(\boldsymbol{\Psi}+t(z)\mathbf{I}_M)^{-2}\mathbf{u}_1\Big|_{z=\xi},$$

where we have defined

$$\zeta(z) = \frac{b^3(z)}{1-\frac{1}{N}\operatorname{Tr}\left[(\boldsymbol{\Psi}+t(z)\mathbf{I}_M)^{-1}\right]+b(z)(1-c-2zb(z))\frac{1}{N}\operatorname{Tr}\left[(\boldsymbol{\Psi}+t(z)\mathbf{I}_M)^{-2}\right]}.$$

## Appendix 1.C Proof of Proposition 1.2

Let us define $\gamma_1 = \mathsf{SNR}_1^{1/2}$. Clearly, in order to prove (1.79) in Proposition 1.2, it is enough to show that the following quantity vanishes almost surely, as $M,N\to\infty$, $M/N\to c<+\infty$ , namely,

$$\gamma_1\left(\frac{\mathbf{u}_1^H\left(\hat{\mathbf{R}}-\xi_1\mathbf{I}_M\right)^{-1}\mathbf{u}_1}{\left(\mathbf{u}_1^H\left(\hat{\mathbf{R}}-\xi_1\mathbf{I}_M\right)^{-2}\mathbf{u}_1\right)^{1/2}}\frac{\mathbf{v}_1^H\left(\hat{\mathbf{C}}-\xi_2\mathbf{I}_K\right)^{-1}\mathbf{v}_1}{\left(\mathbf{v}_1^H\left(\hat{\mathbf{C}}-\xi_2\mathbf{I}_K\right)^{-2}\mathbf{v}_1\right)^{1/2}}-\frac{U_1(\xi_1)}{(U_2(\xi_1))^{1/2}}\frac{V_1(\xi_2)}{(V_2(\xi_2))^{1/2}}\right),$$

or, equivalently, the almost surely convergence to zero of

$$\gamma_1\frac{\mathbf{v}_1^H\left(\hat{\mathbf{C}}-\xi_2\mathbf{I}_K\right)^{-1}\mathbf{v}_1}{\left(\mathbf{v}_1^H\left(\hat{\mathbf{C}}-\xi_2\mathbf{I}_K\right)^{-2}\mathbf{v}_1\right)^{1/2}}\left(\frac{\mathbf{u}_1^H\left(\hat{\mathbf{R}}-\xi_1\mathbf{I}_M\right)^{-1}\mathbf{u}_1}{\left(\mathbf{u}_1^H\left(\hat{\mathbf{R}}-\xi_1\mathbf{I}_M\right)^{-2}\mathbf{u}_1\right)^{1/2}}+\frac{U_1(\xi_1)}{(U_2(\xi_1))^{1/2}}\right), \tag{1.122}$$

and

$$\gamma_1\frac{U_1(\xi_1)}{(U_2(\xi_1))^{1/2}}\left(\frac{\mathbf{v}_1^H\left(\hat{\mathbf{C}}-\xi_2\mathbf{I}_K\right)^{-1}\mathbf{v}_1}{\left(\mathbf{v}_1^H\left(\hat{\mathbf{C}}-\xi_2\mathbf{I}_K\right)^{-2}\mathbf{v}_1\right)^{1/2}}-\frac{V_1(\xi_2)}{(V_2(\xi_2))^{1/2}}\right). \tag{1.123}$$

Indeed, since all quantities in the sums (1.122) and (1.123) are uniformly bounded for all $K,M,N$, this can be readily shown using Lemma A.2 and the Cauchy-Schwarz inequality.

# Chapter 2

# Generalized Consistent Reduced-Rank LMMSE Filtering

## 2.1 Summary

An improved estimator of the optimal reduced-rank linear minimum mean-square error (MMSE) filter is derived that is consistent for arbitrarily high-dimensional observations. The new filter construction generalizes traditional filter realizations based on directly replacing the theoretical covariance matrix by its sample estimate, and being consistent when all dimensions in the model but the number of samples remain bounded. Our solution not only generalizes the conventional estimator, but also turns out to appropriately characterize finite sample-size situations defined in practice by a limited number of samples per observation dimension. The proposed implementation results from an enhanced consistent estimation of the MMSE filter coefficients on the reduced-dimensional subspace. Results are based on the theory of spectral analysis of large-dimensional random matrices. In particular, we build on the analytical description of the asymptotic spectrum of sample-covariance-type matrices in the limiting regime defined as both the number of samples and the observation dimension grow without bound at the same rate. As a result, the implementation of the linear MMSE estimator based on the proposed filter is empirically shown via simulations to present a superior performance under finite sample-size scenarios, avoiding the breakdown on the mean-square error as the selected rank increases.

## 2.2 Introduction

The theory of linear minimum mean-square error (LMMSE) estimation has found a wide variety of applications in the fields of communications, such as in channel estimation, equalization and symbol detection, and array processing, as in beamforming and radar/sonar. Indeed, a large number of these applications can be interpreted in terms of a parameter estimation problem, typically approached by a linear filtering operation acting upon a set of multidimensional observations. Moreover, in many cases, the underlying structure of the observable signals is linear in the parameter to be estimated. In this chapter, we are concerned with the general and fundamental problem of discrete-time linear filtering of noisy signals aimed at the estimation of a linearly described unknown random parameter. The MMSE solution to this problem is well-known to rely on the second-order statistics of the observed process, namely the observation covariance matrix, via an inverse operation.

In practice, the lack of true covariance information leads to a LMMSE filter implementation based on the empirical statistics of the received data samples. This fact suggests the use of the sample matrix inversion algorithm (SMI), consisting in simply replacing the theoretical covariance matrix by its sample estimate, namely the sample covariance matrix (SCM). However, two problems related to the SMI realization of the optimum LMMSE estimator may be readily identified: the computational complexity and the sample-support requirements. These problems are quickly aggravated as the observation dimension becomes higher. On the one hand, an increasingly large covariance matrix needs to be inverted. On the other hand, a particularly limited sample-support would especially contribute to a severe degradation of the performance of the estimator implementation. Therefore, the number of samples required to properly approximate the theoretical covariance matrix is considerably increased.

Existing methods proposed in the literature to mitigate the limitations of filter implementations under finite sample-size situations are categorized into two broad families. In the array processing literature, diagonal loading has been extensively analyzed and applied as a natural complement of the SMI technique [Abr81, Tre02, Li05]. This regularization approach renders it feasible to invert originally ill-posed sample covariance matrices and allows for an efficient reduction of the number of adaptive degrees of freedom (and hence the effective sample-size requirements) by adding to the SCM a small constant loading factor. Similarly, reduced-rank methods, primarily developed with the aim of reducing the computational complexity of the subsequent processing by approximating the optimal solution in a lower-dimensional subspace, have also been widely proposed in the literature as a means to effectively lessen the sample-size requirements. Reduced-rank linear estimation and filtering (cf. [Sch91, Section 8.4][Joh93]) have been clasically discussed for the detection of CDMA signals [Mos96, Pad99, Guo99, Guo00, Gra01, Hon02], originally for situations where the processing gain is much larger than the signal subspace dimension. In array and radar signal processing applications, rank-reduction methods have been used in order to enable an accurate calculation of the filter coefficients with a relatively small amount of observed data under complexity constraints (see e.g. [Gol97a, Gol97b, Gol97c, Zul98, Gol99, Gue00, Pec00] and the much earlier work [Erm93]). Beamforming algorithms derived from an optimization space of reduced dimension are summarized in [Tre02, Section 6.8], whereas contributions to the theory of adaptive filters can be found in [Bur02, Xia05]. Additionally, reduced-dimensional approaches were also investigated in [Str96], where a subspace tracking algorithm based on power iterations was applied to the adaptive filtering problem, as well as in [YH01] for the estimation of reduced-rank multivariate linear regressions (see also references therein). Finally, other applications include channel equalization [Cho00], Jammer suppression in satellite systems [Myr00] and model order reduction and quantization [Sch91, Sch98, Sch06].

Early applications of reduced-rank methods to engineering problems (see e.g. [Tuf82, Tuf93]) consisted of directly replacing the covariance matrix by a lower-rank approximation obtained from its principal components analysis[1] (PCA) [Hot46]. In fact, the approximation of the covariance of an observed input random vector obtained from the truncation of its eigen-expansion can be readily identified as the covariance matrix associated with the MMSE estimator of the random observation on the reduced-dimensional subspace [Chu05]. However, this approach is merely based on the covariance matrix (or its sample estimate) and takes no further signal structure into account. In order to obtain an approximation of the MMSE estimator of the unknown parameter, a reduced-rank filter should be instead obtained that minimizes a mean-square error measure involving not only the covariance matrix, but also the particular signal structure of the observations. In this sense, a reduced-rank Wiener filter is derived based on projected instances of the received observations onto a lower-dimensional subspace. Different reduced-rank methods can be identified depending on the selected projection subspace. The cross-spectral method (CSM), originally reported in [Gol97a], proposed to select the subspace basis through the enumeration of the covariance eigenvectors maximizing the correlation with the intended signal signature. This procedure takes into consideration the structure of the signal model in order to further minimize the approximation mean-square error (MSE). The multistage Wiener filter (MSWF) was conceived in [Gol98] as an extension of the CSM and consists of different stages that are concatenated in the form of a chain of generalized sidelobe cancelers (GSC) [Tre02] in order to successively explore a number of orthogonal directions of maximal correlation with the desired signal. An important breakthrough in the understanding of the MSWF was made in [Gol98], where the projection space is recognized as the Krylov subspace spanned by the covariance matrix of the input signals and the cross-correlation between the observations and the desired signal. As it turns out, the aforementioned Krylov subspace is the optimum linear projection space in the sense of minimizing the approximation MSE [Saa96]. This fact allowed the research community to rapidly establish the connection between this direct approach and well-known iterative algorithms. In particular, the conjugate gradient method (CG) can be employed in order to provide numerically stable solutions through the recursive construction of an orthonormal basis for the Krylov subspace (see e.g. [Vor03b] and also [Sch03a], and refer-

[1] In signal processing, PCA is known as Karhunen-Loéve transform (KLT) [Kar46], and both are immediately related to the eigendecomposition of the covariance matrix.

ences herein, for the engineering literature). Finally, the performance of the previous projection methods under finite sample-support can be related to the behaviour of Tikhonov regularization techniques (diagonal loading) via their representation in the spectral domain in terms of the so-called filter-factors [Han97].

In practice, the performance of the MMSE estimator is highly determined by the accuracy of the available empirical statistics, i.e., the error incurred in the estimation of the covariance matrix from a finite number of observations. This error will be especially large in high-dimensional, relatively low sample-support scenarios. Despite its unquestionable interest in practice, little analytical insight can still be drawn from the broad literature about estimation problems characterized by the availability of a limited number of samples of arbitrarily high dimension. In this chapter, a class of reduced-rank LMMSE estimators uniquely based on the SCM is derived that are consistent under more general conditions than the conventional implementation. In particular, our estimator generalizes conventional filter realizations, consistent in the classical sence, by guaranteeing consistency even for arbitrarily high-dimensional observations. To that effect, the proposed filtering structure minimizes the empirical mean-square error for an arbitrary number of samples per observation dimension. This is accomplished by approximating the empirical performance measure in terms of the spectrum of the SCM in a doubly-asymptotic regime defined when both the number of samples and their dimension grow without bound at the same rate. This asymptotic regime is in agreement with the characterization of finite sample-support situations, namely in the sense that the ratio between the actual sample-size and the existing filtering degrees of freedom is allowed to remain finite. For our purposes, we resort to the theory of the spectral analysis of large random matrices (or random matrix theory) [Tul04], that studies the limiting behaviour of the spectrum of certain random matrix models as their dimension grow large with a given (finite) aspect ratio. In particular, our approach is rooted in Girko's general statistical analysis (GSA) of observations of large dimension [Gir98] (or G-analysis), aimed at (asymptotically) approximating a certain class of functions of the spectrum of covariance-type matrices by (asymptotically) equivalent spectral functions of the SCM. An analogous approach has been used in [Mes06a] for the estimation of the eigenvalues of covariance matrices and their associated subspaces from their sample estimates under the assumption of observations of arbitrarily high dimension. Similarly conceived work, reporting a doubly-consistent G-estimator of the optimum loading factor of a diagonally-loaded Capon beamformer, was published in [Li05] (see also [Mes06c]).

The article is organized as follows. Section 2.4 outlines the theory of optimum linear MMSE estimation on a subspace of reduced dimension. Extending on the classical implementation, an alternative approach is introduced at the end of the section leading to a filter construction that is consistent for arbitrarily high observations. This approach is used to derive an improved realization of the reduced-rank LMMSE estimator. In Section 2.5, an asymptotically equivalent expression of the key elements describing the filter performance is derived that is used in Section

2.6 to obtain the consistent optimal reduced-rank filter. The performance of the new estimator under non-asymptotic conditions is evaluated in Section 2.7 via numerical simulations. Finally, after some concluding remarks in Section 2.8, pertinent proofs and derivations are provided in the appendices.

## 2.3 Reduced-rank approaches to low sample-support

In this section we review existing rank-reduction techniques for linear filtering. Motivated by the two problems addressed above, these reduced-rank methods perform a certain filtering operation on an instance of the received signal that is projected onto a lower-dimensional subspace. Equivalently, the observed data may be thought as being compressed to a lower dimensional subspace upon the action of a pre-filtering matrix whose columns are the basis vectors of the reduced-rank subspace. Specifically, let $\mathbf{S}_D \in \mathbb{C}^{M \times D}$ be a matrix whose columns form a basis for a particular $D$-dimensional subspace, where $D < M$, and $M$ is the observation dimension. The transformed lower-dimensional received signal is then

$$\tilde{\mathbf{y}}(n) = \mathbf{S}_D^H \mathbf{y}(n).$$

The objective is now to obtain a filter $\tilde{\mathbf{w}}$ that minimizes the reduced-rank MSE[2]. This is accomplished by the (reduced-rank) Wiener filter solution, given by

$$\tilde{\mathbf{w}} = \tilde{\mathbf{R}}^{-1}\tilde{\mathbf{s}} = \left(\mathbf{S}_D^H \mathbf{R} \mathbf{S}_D\right)^{-1} \mathbf{S}_D^H \mathbf{s}, \tag{2.1}$$

where $\tilde{\mathbf{R}} = \mathbf{S}_D^H \mathbf{R} \mathbf{S}_D$ and $\tilde{\mathbf{s}} = \mathbf{S}_D^H \mathbf{s}$. On the other hand, the reduced-rank ($D$-dimensional) approximation of the $M$-tap linear MVDR/MMSE filter is

$$\mathbf{w}_{RR} = \mathbf{S}_D \tilde{\mathbf{w}} = \mathbf{S}_D \left(\mathbf{S}_D^H \mathbf{R} \mathbf{S}_D\right)^{-1} \mathbf{S}_D^H \mathbf{s}. \tag{2.2}$$

Finally, the reduced-rank MMSE is

$$\mathsf{MMSE} = 1 - \tilde{\mathbf{s}}^H \tilde{\mathbf{R}}^{-1} \tilde{\mathbf{s}} = 1 - \mathbf{s}^H \mathbf{S}_D \left(\mathbf{S}_D^H \mathbf{R} \mathbf{S}_D\right)^{-1} \mathbf{S}_D^H \mathbf{s}. \tag{2.3}$$

Next, we discuss briefly the reduced-dimensional filtering operation in different projection spaces.

### 2.3.1 Eigen-based methods

The first approach to dimension-reduction was based on principal components analysis (PCA). According to this method, a pre-filtering matrix composed of the eigenvectors belonging to the

[2] Note that the mean squared error as regarded here refers to the Bayesian risk

principal eigenvalues of the observation covariance matrix is applied to the received samples in order to obtain a transformed signal of lower dimensionality. Consider an observation covariance matrix with the following structure, namely, $\mathbf{R} = \mathbf{S}\mathbf{P}\mathbf{S}^H + \sigma^2 \mathbf{I}_M$. This definition exemplifies perfectly the PCA method and applies directly to a typical array signal processing application or a synchronous DS-CDMA system, where $\mathbf{S} \in \mathbb{C}^{M \times K}$ is the signatures matrix and $\mathbf{P} \in \mathbb{R}^{K \times K}$ is diagonal and contains the powers associated to the $K$ sources or users. The columns of $\mathbf{S}$ (assumed linearly independent) span the $K$-dimensional signal-subspace (actually signal-plus-interference), assuming $K < M$. The eigenvectors of $\mathbf{R}$ that correspond to the smallest $M - K$ eigenvalues (equal to $\sigma^2$) span the noise-subspace, which is orthogonal to the signal-subspace. Basically, rank-reduction based on PCA relies on the separation of the signal and noise subspaces associated with the covariance matrix and selects the signal components with largest power without distinguishing between desired and interference signals. Thus, by choosing $D \geq K$, this method retains full-rank MMSE performance. However, the performance can degrade considerably as $D$ decreases below the dimension of the signal subspace, since there is no guarantee that the associated subspace will retain most of the desired signal energy.

In order to overcome this problem, the cross-spectral (CS) method was introduced in [Gol97a, Gol97b] as an extension of the linearly constrained (MMSE) Wiener beamformer in array signal processing. In applications where a desired signal is not explicitly available, such as in spatial-reference filtering techniques (unlike temporal-reference methods), the Wiener filter can be formulated as an optimization problem with multiple linear constraints. The MVDR filter may be then regarded as the special case of a single constraint.

In the array processing literature, two equivalent representations may be considered [Sch91]. First, the direct-form processor performs a conventional filtering operation referred to as elemen-space processing. On the other hand, when the number of sensors is large, it may be approppriate to create a set of beams as a preliminary step to further processing (beamspace-processing). Depicted in Figure 2.1, this second representation is referred to as the generalized sidelobe-canceller (GSC) and is especially convenient to identify the possibility of rank-reduction by implementing a transformation from the element-space (related to fixed weights in Figure 2.1) to the beamspace (related to adjustable weights in Figure 2.1) such that the number of auxiliary beams is less than the number of sensors or antennas. The CS method was conceived as an extension of the GSC that selects the columns of the blocking matrix (denoted by $\mathbf{B}$ in Figure 2.1) as the set of eigenvectors yielding the largest CS metric, chosen to minimize the MSE considering not only the statistics of the observation but also its relation to the desired signal (i.e. the parameter signature). In particular, this measure is used to determine the smallest number of degrees of freedom $D$ that are needed in order to estimate linearly with little loss (minimum MSE) a scalar random process from a set of $M$ correlated complex random processes. According to the CS method, and considering the definition of $\mathbf{S}_D$ for eigendecomposition-based methods, a set of $D$ eigenvectors $\{\mathbf{q}_d\}$ of the covariance matrix $\mathbf{R}$ is to be chosen in order to

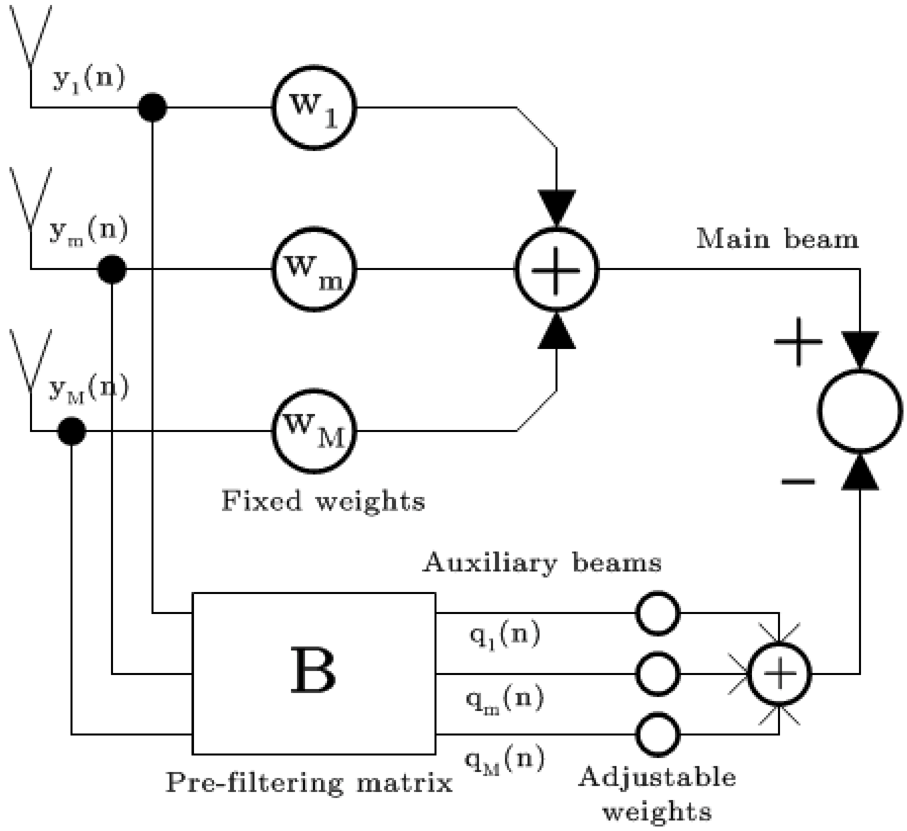


Figure 2.1: Block diagram of Generalized Sidelobe Canceller

minimize (2.3):

$$\begin{aligned}\mathsf{MMSE} &= 1 - \mathbf{s}^H \mathbf{S}_D \left(\mathbf{S}_D^H \mathbf{R} \mathbf{S}_D\right)^{-1} \mathbf{S}_D^H \mathbf{s} \\ &= 1 - \mathbf{s}^H \mathbf{S}_D \mathbf{\Lambda}_D^{-1} \mathbf{S}_D^H \mathbf{s} \\ &= 1 - \sum_{d \in \mathcal{C}} \frac{\left|\mathbf{q}_d^H \mathbf{s}\right|^2}{\lambda_d},\end{aligned}$$

where $\mathcal{C}$ is the index set of the selected eigenvectors and $\mathbf{\Lambda}_D$ is a diagonal matrix containing the corresponding set of eigenvalues $\{\lambda_d\}$. Thus, as opposed to PCA methods, this technique can perform well for $D < K$ since it takes into account the energy in the subspace contributed by the desired signature. Both PCA and CS methods admit a representation as a direct-form or GSC-form processor, although their performance under rank-reduction is different depending on the implementation.

### 2.3.2 Multistage Wiener filter

The multistage Wiener filter (MSWF) proposed in [Gol98] was motivated by a continuation of the original representation of the CS method as a GSC processor (see Figure 2.1). Contrary to eigendecomposition-based methods, both direct-form and GSC-form representations of the MSWF are equivalent for any selected rank (see [Zul98] and discussion in Appendix B of [Gol98]). After the first beamspace processor, different stages are concatenated in the form of a chain of GSCs in order to construct recursively the pre-filtering matrix. From the output of the blocking matrix at each stage, a new basis vector is obtained that is orthogonal to the previous filter in the sequence (i.e. in its nullspace). At each stage, a matched filter is chosen to guarantee that its output is maximally correlated with the output of the previous stage (whitening innovation). Consequently, it is straightforward to show that the optimum sequence of filters can be obtained as

$$\mathbf{s}_i = \frac{\left(\prod_{k=1}^{i-1} \mathbf{B}_k\right) \mathbf{R}\mathbf{s}_{i-1}}{\left\| \left(\prod_{k=1}^{i-1} \mathbf{B}_k\right) \mathbf{R}\mathbf{s}_{i-1} \right\|}, \tag{2.4}$$

where $\mathbf{s}_0 = \mathbf{s}$ and, in the case of orthonormal filters, the $k$th blocking matrix is defined as $\mathbf{B}_k = \mathbf{I}_M - \mathbf{s}_k \mathbf{s}_k^H$. Accordingly, using the MSWF representation as a chain of GSCs, the pre-filtering matrix can be written as

$$\begin{aligned} \mathbf{S}_D &= \left[ \mathbf{s}_1 \, \mathbf{B}_1 \mathbf{s}_2 \cdots \left(\prod_{k=1}^{D} \mathbf{B}_k\right) \mathbf{s}_D \right] \\ &= \left[ \mathbf{s}_1 \, \mathbf{s}_2 \cdots \mathbf{s}_D \right], \end{aligned}$$

where the second equality holds due to the orthonormality impossed to the filters and the corresponding definition of the blocking matrices. It is shown in [Gol98] that the MSWF recursion in (4.15) tri-diagonalizes the covariance matrix $\mathbf{R}$ at any stage. In other words, the projected covariance matrix (i.e. $\mathbf{S}_D^H \mathbf{R} \mathbf{S}_D$) takes the form of a tri-diagonal matrix.

An important breakthrough in understanding of the MSWF was made in [Hon01], where the inherent connection with Krylov subspaces is identified. Honig et. al. showed that the $D$-rank MSWF solution lies on the $D$-dimensional Krylov subspace spanned by the covariance matrix of the observations and the parameter signature vector, which is defined as the column space of the (Krylov) matrix $\left[ \mathbf{s} \, \mathbf{R}\mathbf{s} \cdots \mathbf{R}^{M-1} \mathbf{s} \right]$. In the following, this subspace will be referred to as $\mathcal{K}_D(\mathbf{R}, \mathbf{s})$. Thus, although solely motivated by statistical reasoning, the reduced-rank MSWF solution can be regarded as an approximation of the Wiener filter obtained by forcing the solution to lie on $\mathcal{K}_D(\mathbf{R}, \mathbf{s})$ or, equivalently, as an approximation in this subspace of the solution of a system of linear equations, namely $\mathbf{R}\mathbf{w} = \mathbf{s}$. More descriptively, the inverse of the covariance matrix $\mathbf{R}^{-1}$ is approximated by a matrix polynomial of order $D-1$ (see further the next section). Having established this, the solution for the filter sequence may be obtained by directly computing an orthonormal basis for $\mathcal{K}_D(\mathbf{R}, \mathbf{s})$. In fact, the recursive algorithm in

(4.15) is nothing else than the well-known Gram-Schmidt Arnoldi algorithm [Gol96] to compute an orthonormal basis of the Krylov subspace spanned by $\mathbf{R}$ and $\mathbf{s}$. This algorithm is known as the GMRES algorithm when applied to solve a system of linear equations [Tre97]. For an arbitrary square matrix $\mathbf{R}$, the Arnoldi algorithm returns a transformed Hessenberg matrix $\mathbf{S}_D^H \mathbf{R} \mathbf{S}_D$. On the other hand, if $\mathbf{R}$ is Hermitian, as it is the case here, the algorithm outputs a Hermitian Hessenberg matrix or, equivalently, Hermitial tri-diagonal matrix. Note that it is in agreement with the fact that the MSWF tri-diagonalizes the covariance matrix. Moreover, the fact that $\mathbf{R}$ is Hermitian can be exploited to obtain the sequence of filters by the Lanczos algorithm. In [Joh00], an order-recursive version of the MSWF based on the Lanczos algorithm is presented that updates the reduced-rank filter and MSE at each step. Finally, since $\mathbf{R}$ is also positive definite, the Conjugate Gradient (CG) algorithm [Tre97] can be applied to compute the solution for the orthonormal basis in an iterative fashion. In brief, the CG algorithm can be identified as a solver of quadratic (non-linear) optimization problems which, unlike the Newton's method that optimizes a quadratic function in a single but computationally-intensive step, may take up to $M$ low-complexity iterations to find the optimal solution. The CG algorithm returns the exact solution of the linear system $\mathbf{Rw} = \mathbf{s}$, after $M$ iterations. However, contrary to direct methods (e.g. Gaussian elimination), the algorithm may be terminated after $D < M$ iterations to obtain an approximate solution of the system in $\mathcal{K}_D(\mathbf{R}, \mathbf{s})$.

### 2.3.3 Polynomial expansion linear filtering

Another natural approach to dimension-reduction in the line of the above interpretation of the reduced-rank MSWF solution can be derived from a result of the following theorem.

**Theorem 2.1** *(Cayley-Hamilton) Let* $p_{\mathsf{R}}(\lambda) = \det(\mathbf{R} - \lambda \mathbf{I}_M) = \sum_{l=0}^{M} a_l \lambda^l$ *be the characteristic polynomial of* $\mathbf{R}$. *Then*

$$p_{\mathsf{R}}(\mathbf{R}) = \sum_{l=0}^{M} a_l \mathbf{R}^l = 0.$$

Proof. *See [Hor85].* ■

Hence, it follows that if $\det(\mathbf{R}) \neq 0$, its inverse $\mathbf{R}^{-1}$ can be expressed as a polynomial in $\mathbf{R}$ of order $M-1$. Then, the fundamental filter solution may be rewritten using the Cayley-Hamilton (CH) theorem in terms of a polynomial expansion as

$$\mathbf{w}_{\mathsf{CH}} = \mathbf{R}^{-1}\mathbf{s} = \sum_{l=0}^{M-1} \varpi_l \mathbf{R}^l \mathbf{s}, \tag{2.5}$$

where $\varpi_l = -a_{l+1}/a_0$. A $D$-rank solution clearly lying in $\mathcal{K}_D(\mathbf{R}, \mathbf{s})$ can be obtained now by constraining the polynomial to be of order $D-1$. In this case, the coefficients must be found according to some optimality criterion, usually to minimize MSE or, equivalently, maximize SINR. From (3.10), the SINR for the CH-based filter can be expressed as

$$\mathsf{SINR} = \left( \frac{\sum_{i=0}^{D-1} \sum_{j=0}^{D-1} \omega_i \omega_j \mathbf{s}^H \mathbf{R}^{i+j+1} \mathbf{s}}{\left| \sum_{i=0}^{D-1} \omega_i \mathbf{s}^H \mathbf{R}^i \mathbf{s} \right|^2} - 1 \right)^{-1} = \left( \frac{\boldsymbol{\omega}^H \mathbf{B} \boldsymbol{\omega}}{\boldsymbol{\omega}^H \mathbf{A} \boldsymbol{\omega}} - 1 \right)^{-1},$$

where $\boldsymbol{\omega} \in \mathsf{R}^D$ is a column vector with the polynomial coefficients to be optimized, $\mathbf{A} = \mathbf{v}\mathbf{v}^H$, where $[\mathbf{v}]_k = \mathbf{s}^H \mathbf{R}^{k-1} \mathbf{s}$, and the elements of the matrix $\mathbf{B}$ are $[\mathbf{B}]_{k,l} = \mathbf{s}^H \mathbf{R}^{k+l+1} \mathbf{s}$. Thus, the vector of coefficients maximizing the output SINR is straightforwardly found as the eigenvector associated with the minimum eigenvalue of the generalized eigenproblem $\mathbf{B}\boldsymbol{\omega}_o = \lambda_o \mathbf{A} \boldsymbol{\omega}_o$, or, equivalently, $\boldsymbol{\omega}_o = \mathbf{B}^{-1}\mathbf{v}$. Using this expression of the optimal vector of coefficients $\boldsymbol{\omega}_o$ a reduced-rank filter is obtained that is equivalent to the solution given in (2.2), i.e.

$$\mathbf{w}_{\mathsf{RR}} = \sum_{i=0}^{D-1} \omega_i \mathbf{R}^i \mathbf{s} = \mathbf{S}_D \boldsymbol{\omega}_o = \mathbf{S}_D \left( \mathbf{S}_D^H \mathbf{R} \mathbf{S}_D \right)^{-1} \mathbf{S}_D^H \mathbf{s} = \mathbf{S}_D \mathbf{B}^{-1} \mathbf{v}.$$

Although the scalar estimation solution based on the CH theorem is clearly equivalent to the MSWF, polynomial expansion linear filtering was independently introduced in a previous contribution [Mos96]. A much earlier work on adaptive arrays [Erm93] (see reference [3] for the original contribution in Russian) proposed the representation of the inverse of the covariance matrix in terms of a power basis related to the minimum polynomial of $\mathbf{R}$.

### 2.3.4 Auxiliary-vector algorithm

The auxiliary-vector (AV) method [Pad97] parallels also beamspace processing techniques as the GSC but, founded solely on statitical signal processing principles, does not attempt any matrix inversion, eigendecomposition or diagonalization. An extension to complex linear spaces is presented in [Pad99], where, as in the previous earlier work, the issue of reduced-rank filter design for small-sample-support adaptation was directly discussed. Basically, a set of (unit-norm) auxiliary vectors, required to be mutually orthonormal and also orthogonal to the signature vector, are chosen to maximize the magnitude of the statistical cross-correlation between the output of the previous filter in the sequence and the projection of the input data onto the auxiliary vector itself. A recursive conditional optimization (one-by-one) of the filter taps is performed based on scalar auxiliary vector weights that are chosen to minimize the new filter output variance. Given that the true covariance matrix is available, the sequence of filters converges to the optimum filter. When $\mathbf{R}$ is substituted by $\hat{\mathbf{R}}$ in the recursively generated sequence of filters, the corresponding filter estimators offer a means for effective control over the filter estimator bias versus covariance tradeoff. Starting from the zero-variance, high-bias (for non-white inputs) matched-filter estimate, the solution evolves all the way up to the unbiased, yet high-variance for short data record sizes. In [Che02], the AV-algorithm is shown to be equivalent to the MSWF or, correspondingly to the CH-based solution. In [Pad01], the orthogonality condition among auxiliary vectors is relaxed, and therefore the length of the generated sequence of filters is not limited to the vector space dimension. This extension is shown in [Pad01] to outperform the previous existing reduced-rank methods. Finally, in [Qia03], a procedure to select the best AV estimator in the sequence directly from the input data is presented.

### 2.3.5 Iterative methods

Before concluding this section, we introduce a class of methods resembling the iterative algorithms for the solution of linear systems that are related to reduced-rank approximations of the optimum solution.

In order to approximate the linear MVDR/MMSE solution, we address the iterative solution of $\mathbf{Rw} = \mathbf{s}$. The Jacobi iteration [Axe94] is given by

$$\mathbf{w}_{k+1} = \mathbf{s} + (\mathbf{I} - \mathbf{R})\,\mathbf{w}_k, \tag{2.6}$$

with $\mathbf{w}_0 = \mathbf{s}$. Furthermore, the solution after $k$ iterations can be shown by induction to admit the following compact equivalent expression:

$$\mathbf{w}_k = \sum_{i=0}^{k} (\mathbf{I} - \mathbf{R})^i\,\mathbf{s}.$$

This is equivalent to implementing the matrix inversion operation via the series expansion

$$\mathbf{R}^{-1} = \lim_{P\to\infty} \sum_{p=1}^{P} (\mathbf{I} - \mathbf{R})^{p-1},$$

whose convergence is guaranteed if the covariance matrix $\mathbf{R}$ has spectral radius smaller than two. Obviously, this condition turns out to be quite restrictive in many cases of interest and fast convergence is in practice hardly guaranteed. In order to improve the convergence capabilities of the Jacobi iteration, a sequence of parameters can be introduced in (2.6) as

$$\begin{aligned}\mathbf{w}_{k+1} &= \tau_k\mathbf{s} + (\mathbf{I} - \tau_k\mathbf{R})\,\mathbf{w}_k \\ &= \mathbf{w}_k + \tau_k\,(\mathbf{s} - \mathbf{R}\mathbf{w}_k)\,.\end{aligned} \tag{2.7}$$

These parameters are usually chosen to improve the convergence speed of the method. The filter expression in (2.7) is regarded as first-order iterative solution. If $\tau_k = \tau$ for all $k$, the method is called stationary. In this case, a closed-form expression for the filter after $k$ iterations is also found as

$$\mathbf{w}_k = \tau \sum_{i=0}^{k} (\mathbf{I} - \tau\mathbf{R})^i\,\mathbf{s}. \tag{2.8}$$

Finally, a direct connection to gradient-like methods for unconstrained optimization problems can be readily established. Consider as a (quadratic) cost-function the expression of MSE for the filtering estimation problem at hand, namely

$$\mathsf{MSE} = \mathsf{E}\left\{\left|b - \mathbf{w}^H \mathbf{y}\right|^2\right\} = 1 - \mathbf{w}^H \mathbf{s} - \mathbf{s}^H \mathbf{w} + \mathbf{w}^H \mathbf{R} \mathbf{w}.$$

The steepest-descent method (SDM) [Lue84] gives the following recursion to minimize MSE

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mu \frac{\partial \mathsf{MSE}}{\partial \mathbf{w}_k^*},$$

where $\mu$ is a positive real-valued constant, usually referred to as step-size, and the gradient takes the expression of $\frac{\partial \mathsf{MSE}}{\partial \mathbf{w}_k^*} = \mathbf{R}\mathbf{w}_k - \mathbf{s}$. Hence, the filter expression after $k+1$ iterations can be written as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu \left(\mathbf{s} - \mathbf{R}\mathbf{w}_k\right),$$

which is equivalent to the stationary version of the first-order iterative solution, where the parameter $\tau$ is replaced by the step-size $\mu$. The introduction of a variable step-size leads to the original iteration in (2.7). As before, the solution after $k$ iterations can be expressed in this case as the following non-recursive (one-shot) filter

$$\mathbf{w}_k = \sum_{i=1}^{k} \mu_i \prod_{j=i+1}^{k} \left(\mathbf{I} - \mu_j \mathbf{R}\right) \mathbf{s}.$$

As the MSWF and CH-based solutions, the previous implementations of the optimum filter based on iterative methods for solving linear systems or, more generally, unconstrained optimization problems can be shown to lie on $\mathcal{K}_D(\mathbf{R}, \mathbf{s})$ after $D$ iterations. In particular, for stationary (or fixed step-size) methods, the solution after $D$ iterations can be rewritten replacing each term of the sum in (2.8) by its binomial series expansion as

$$\mathbf{w}_D = \sum_{i=0}^{D} \sum_{j=0}^{i} \binom{i}{j} (-1)^j \tau^{j+1} \mathbf{R}^j \mathbf{s},$$

which clearly lies on $\mathcal{K}_D(\mathbf{R}, \mathbf{s})$. Regarding the optimality of the solution minimizing the quadratic function $\mathsf{MSE}(\mathbf{w})$, it is worth noting that if the vicinity of the cost function minimum has the shape of a long, narrow valley, the CG algorithm reaches the (global) optimal solution in far fewer steps than would be the case using the SDM.

Filtering schemes based on iterative solutions of linear systems and their convergence properties were extensively studied in [Gra01] in the context of multiuser detection in DS-CDMA systems. Departing from a symbol-mached filter model of the received CDMA signal, the conventional (linear) parallel interference cancellation (PIC) receiver is shown to be equivalently formulated in terms of the code-correlation matrix as a Jacobi iteration. Furthermore, this receiver using soft decisions to compute the data estimates of all users in parallel is shown to converge to the decorrelator receiver. Motivated by a reduction of the bias in the decision statistic, the previous PIC receiver is improved by only partially cancelling the interference from data estimates for each user, at each stage. The linear multistage partial PIC (PPIC) receiver is equivalent to a first-order stationary iterative method, where the parameter is regarded as the partial cancellation factor (PCF) (see [Tri00], where the PCF is optimized to minimize the bit-error-rate (BER), and references therein). By including the knowledge of the noise variance in the iteration (complete covariance knowledge), an iterative receiver is obtained that converges to the linear MMSE multiuser detector. Guo et al. in [Guo99, Guo00] identified the connection between the PPIC and the SDM for updating adaptive filter tap-weights to minimize MSE. Furthermore, they showed that a number of stages equal to the dimension of the signal subspace is merely required for the equivalent one-shot filter to be identical to the linear MMSE filter. Following the principles of the SDM, they derived the corresponding one-shot cancellation filter of a weighted linear multistage PIC with variable step-size, and devised techniques for optimizing the choice of weights with respect to the MSE for a given number of stages.

## 2.4 Reduced-Rank Linear MMSE Estimation

The structure of a vast number of statistical estimation problems arising in signal processing and communications can be properly described by defining the observable variable according to the following linear system model

$$\mathbf{y}(n) = x(n)\,\mathbf{h} + \mathbf{n}(n)\,, \tag{2.9}$$

where the signal waveform $x(n)$ represents the unknown parameter to be estimated, received in a colored noise $\mathbf{n}(n) \in \mathsf{C}^M$ after being operated upon by the known signature vector $\mathbf{h} \in \mathsf{C}^M$. In this chapter, as conventionally assumed, $x(n)$ and $\mathbf{n}(n)$ are two independent and jointly distributed wide-sense stationary random processes. Furthermore, both signal and noise processes have mean zero and covariance $\mathsf{E}\left[x(m)\,x(n)^H\right] = \delta_{m,n}P$ and $\mathsf{E}\left[\mathbf{n}(m)\,\mathbf{n}(n)^H\right] = \delta_{m,n}\mathbf{R}_N$, respectively, where $\delta_{m,n}$ is the Kronecker delta function, $P$ is the signal-of-interest power and $\mathbf{R}_N$ is the covariance matrix of the background noise and interference component. For the sake of ease of notation, and withough loss of generality, we assume $P = 1$, so that the magnitude of the power associated with each parameter is modeled within the corresponding signature vector. In particular, the observation $\mathbf{y}(n) \in \mathsf{C}^M$ may be modeling the matched filter output sufficient statistic for the received unknown symbols in, for instance, a CDMA multiuser detector, where the columns of $\mathbf{h}$ is the effective user signature associated with a certain desired user; an array processor, where $\mathbf{h}$ contains the angular frequency information (steering vectors) related to the intended source, represented by the entries of $x(n)$.

The MMSE estimator $\hat{x}(n)$ of the parameter $x(n)$ is found as the output of the linear transformation by $\mathbf{w} \in \mathsf{C}^M$ of the observations that minimizes the mean-square error measure, namely,

$$\mathsf{MSE}(\mathbf{w}) = \mathsf{E}\left[\|x(n) - \hat{x}(n)\|^2\right] = \mathsf{E}\left[\left\|x(n) - \mathbf{w}^H \mathbf{y}(n)\right\|^2\right].$$

The linear MMSE filter solution is obtained as [Sch91]

$$\mathbf{w}_{\mathsf{MMSE}} = \mathbf{R}^{-1}\mathbf{h},$$

where $\mathbf{R} \in \mathbb{C}^{M \times M}$ is the covariance matrix of the observed process, i.e.,

$$\mathbf{R} = \mathsf{E}\left\{(\mathbf{y} - \mathsf{E}\{\mathbf{y}\})(\mathbf{y} - \mathsf{E}\{\mathbf{y}\})^H\right\} = \mathbf{h}\mathbf{h}^H + \mathbf{R}_N. \tag{2.10}$$

Accordingly, the achieved MMSE is

$$\mathsf{MSE}(\mathbf{w}) = 1 - 2\,\mathrm{Re}\left\{\mathbf{w}^H\mathbf{h}\right\} + \mathbf{w}^H\mathbf{R}\mathbf{w} = 1 - \mathbf{h}^H\mathbf{R}^{-1}\mathbf{h} + \mathsf{MSE}_{\mathsf{excess}}(\mathbf{w}_{\mathsf{MMSE}}, \mathbf{w}), \tag{2.11}$$

where $\mathsf{MSE}_{\mathsf{excess}}(\mathbf{w}_{\mathsf{MMSE}}, \mathbf{w}) = (\mathbf{w}_{\mathsf{MMSE}} - \mathbf{w})^H \mathbf{R}(\mathbf{w}_{\mathsf{MMSE}} - \mathbf{w})$ is the excess mean-square error. Note that minimizing $\mathsf{MSE}(\mathbf{w})$ is equivalent to minimizing the distance $\mathsf{MSE}_{\mathsf{excess}}(\mathbf{w}_{\mathsf{MMSE}}, \mathbf{w})$. An alternative performance measure particularly spread across the communications literature is the so-called signal-to-interference-plus-noise ratio (SINR), defined as

$$\mathsf{SINR}(\mathbf{w}) = \frac{\left|\mathbf{w}^H\mathbf{h}\right|^2}{\mathbf{w}^H\mathbf{R}_N\mathbf{w}} = \left(\frac{\mathbf{w}^H\mathbf{R}\mathbf{w}}{\left|\mathbf{w}^H\mathbf{h}\right|^2} - 1\right)^{-1}. \tag{2.12}$$

Clearly, the problem of maximizing $\mathsf{SINR}(\mathbf{w})$ is equivalent to that of minimizing $\mathsf{MSE}(\mathbf{w})$, and both optimum solutions are related as

$$\mathsf{SINR}(\mathbf{w}_{\mathsf{MMSE}}) = \frac{1 - \mathsf{MSE}(\mathbf{w}_{\mathsf{MMSE}})}{\mathsf{MSE}(\mathbf{w}_{\mathsf{MMSE}})},$$

with a maximum (MSINR) at the output of the MMSE filter equal to $\mathbf{h}^H\mathbf{R}_N^{-1}\mathbf{h}$.

Assume we are now interested in approximating the MMSE estimator on a subspace of arbitrary dimension. The rationale behind this approach is often a reduction of the computational complexity of the filter implementation. For instance, most standard iterative algorithms searching for the Wiener filter in a recursive yet less computationally expensive manner are indeed based on a succession of such approximations. Equivalently, a similar procedure referred to as beamspace processing [Tre02, Section 3.10] is employed in array processing in order to reduce the number of adaptive channels in applications where the number of interfering sources is much less than the number of adaptive weights (adaptive degrees of freedom).

More specifically, let the Wiener solution be written using the Cayley-Hamilton theorem [Hor85] as

$$\mathbf{w} = \mathbf{R}^{-1}\mathbf{h} \tag{2.13}$$

$$= \sum_{m=0}^{M-1} \upsilon_m \mathbf{R}^m \mathbf{h} = \mathbf{M}\boldsymbol{\upsilon}, \tag{2.14}$$

where $\mathbf{M} \in \mathsf{C}^{M\times M}$ is a *Krylov matrix* defined as

$$\mathbf{M} = \left[\mathbf{h}\,\mathbf{R}\mathbf{h}\cdots\,\mathbf{R}^{M-1}\mathbf{h}\right], \tag{2.15}$$

and the entries of the vector $\boldsymbol{\upsilon} = [\upsilon_1, \cdots, \upsilon_M]^T$ are $\upsilon_m = -a_{m+1}/a_0$, with $\{a_m\}_{m=0,\ldots,M}$ being the set of coefficients of the characteristic polynomial of $\mathbf{R}$[3]. Clearly, the optimum MMSE filter lies on the column space of $\mathbf{M}$, namely the *Krylov space* generated by the covariance matrix and the parameter signature. More interestingly, the coefficients linearly combining the columns of $\mathbf{M}$ can be interpreted as a set of new degrees of freedom that can be possibly exploited to further reduce the computational effort. In the literature, a truncated version of the optimum filter representation in (2.14) has been extensively reported, namely,

$$\tilde{\mathbf{w}}_D = \sum_{d=1}^{D} \omega_d \mathbf{R}^{d-1}\mathbf{h}, \tag{2.16}$$

where the coefficients of the polynomial expansion $\{\omega_d\}_{d=1\ldots D}$ are found to minimize $\mathsf{MSE}(\tilde{\mathbf{w}}_D)$. In practice, a dimension (or rank), equal to $D << M$, namely spanning a $D$-dimensional Krylov subspace (denoted in the sequel as $\mathcal{K}_D$) can be used with little performance loss and a substantial complexity reduction. On the other hand, it is a well-known fact in the literature that a reduced-rank Wiener filter equivalent to that in (2.16) can be alternatively obtained through properly orthogonalizing the total space enclosing the optimum solution (see [Gol98]).

The optimality of the previous procedure in the MMSE sense is considered next. Formally stated, the problem of optimum reduced-rank linear filtering under the MMSE criterion can be decoupled into the optimal selection of both the estimation subspace and the reduce-rank filter. The solution to these two problems can be found as follows.

Let $\mathbf{S}_D \in \mathsf{C}^{M\times D}$ be an arbitrary full-rank matrix and assume the estimation subspace $\mathcal{H}_D$ is defined as the column space of $\mathbf{S}_D$. Then, the filter minimizing $\mathsf{MSE}(\mathbf{w})$ on $\mathcal{H}_D$, namely,

$$\mathbf{w}_D = \arg\min_{\mathbf{w}\in\mathcal{H}_D} \mathsf{E}\left[\left\|x(n) - \mathbf{w}^H\mathbf{y}(n)\right\|^2\right],$$

can be found as the optimum linear combination of the columns of $\mathbf{S}_D$ in the MMSE sense, i.e., $\mathbf{w}_D = \mathbf{S}_D\boldsymbol{\omega}_D$, where

$$\boldsymbol{\omega}_D = \arg\min_{\boldsymbol{\omega}} \mathsf{E}\left[\left|x(n) - \boldsymbol{\omega}^H\mathbf{S}_D^H\mathbf{y}(n)\right|^2\right] \tag{2.17}$$

is the vector of optimum coefficients of such combination. Then, the filter minimizing $\mathsf{MSE}(\mathbf{w})$ over $\mathcal{H}_D$ is given by

$$\mathbf{w}_D = \mathbf{S}_D\left(\mathbf{S}_D^H\mathbf{R}\mathbf{S}_D\right)^{-1}\mathbf{S}_D^H\mathbf{h}. \tag{2.18}$$

Note that the filter in (2.18) represents the *best approximation* to $\mathbf{w}_{\mathsf{MMSE}}$ on $\mathcal{H}_D$ in the norm inducing the distance $\mathsf{MSE}_{\mathsf{excess}}(\mathbf{w}_{\mathsf{MMSE}}, \mathbf{w})$ in (2.11) [Saa96]. Furthermore, observe that

[3] Note that the coefficients $\{a_m\}_{m=0,\ldots,M}$ can be also obtained in terms of (traces of) the powers of $\mathsf{R}$ from the application of Newton's identities [Hor85] to the characteristic polynomial of the covariance matrix.

(2.18) represents a canonical transformation for extracting a reduced-rank approximation to the solution of the system $\mathbf{R}\mathbf{w} = \mathbf{h}$, establishing a general framework for the derivation of most existing iterative techniques for solving linear systems of equations [Gra01].

On the other hand, regarding the estimation subspace, the following lemma establishes the optimality of the Krylov matrix in (2.15) as the one whose columns provide the sequence of expanding subspaces minimizing the MSE for an arbitrary rank $D$. (The full-rank MMSE is assumed to be uniquely achieved in the case $D = K$.)

**Lemma 2.1** *Let $\mathcal{H}_1, \mathcal{H}_2, \ldots$ be a sequence of estimation subspaces of increasing dimension. Under the assumption $\mathcal{H}_1$ , $\mathrm{span}\{\mathbf{h}\}$[4], and over all possible choices of $\mathcal{H}_D$, the ($D$-rank) MMSE achieved on the $D$th subspace in the sequence is minimized on $\mathcal{H}_D = \mathcal{K}_D$, i.e.,*

$$\mathcal{K}_D = \arg\min_{\mathcal{H}_D} \min_{\mathbf{w} \in \mathcal{H}_D} \mathsf{MSE}(\mathbf{w}) .$$

**Proof.** Note that, given $\mathcal{H}_1$ , $\mathrm{span}\{\mathbf{h}\}$, the optimum $\mathcal{H}_D$ can be found as the $D$-th element of a sequence of subspaces minimizing $\mathsf{MSE}(\mathbf{w})$ and expanding one dimension at a time. Indeed, as the dimension of the approximation subspace increases, the norm of the mean-square error decreases most rapidly in the opposite direction of its gradient, namely

$$-\frac{\partial}{\partial \mathbf{w}^*} \mathsf{MSE}(\mathbf{w}) = \mathbf{h} - \mathbf{R}\mathbf{w}.$$

Thus, in order for the expanded subspace to minimize the objective function, we must ensure

$$-\frac{\partial}{\partial \mathbf{w}^*} \mathsf{MSE}(\mathbf{w}) \in \mathcal{H}_{D+1},$$

at the reduced-rank filter $\mathbf{w}$ achieving the MMSE in $\mathcal{H}_D$. However, since $-\frac{\partial}{\partial \mathbf{w}^*} \mathsf{MSE}(\mathbf{w}) \in \mathrm{span}\{\mathbf{h}, \mathbf{R}\mathbf{w}\}$, it is straightforward to see by iterating the induction process started at $\mathcal{H}_1 = \mathrm{span}\{\mathbf{h}\}$ that $\mathcal{H}_{D+1} = \mathbf{R}\mathcal{H}_D$ and, accordingly, $\mathcal{H}_{D+1} = \mathcal{K}^{D+1}$. ■

In conclusion, the optimum reduced-rank LMMSE filter is given by the expression in (2.18), where the columns of the subspace matrix $\mathbf{S}_D$ spans the Krylov subspace defined in Lemma 2.1. Furthermore, the MMSE achieved on the reduced-rank subspace is

$$\mathsf{MSE}(\mathbf{w}_D) = 1 - \mathbf{h}^H \mathbf{S}_D \left(\mathbf{S}_D^H \mathbf{R} \mathbf{S}_D\right)^{-1} \mathbf{S}_D^H \mathbf{h}.$$

A further major motivation for the reduced-rank filtering approach, namely the one by which we are driven in this chapter, is to avoid the performance degradation associated with filter implementations due to a limited sample-size. Indeed, in practically affordable realizations of the optimum reduced-rank filter based on a collection of received observations, the theoretical covariance matrix defining the optimum Krylov approximation subspace derived in Lemma 2.1 is

[4] This choice is usually motivated by the matched filter solution.

not available. Instead, a sample estimate is mandatorily to be employed, which leads unavoidably to a decrease in the filter performance quality. The direct substitution in (2.18) of the true covariance matrix by the SCM has been extensively regarded in the literature as the filter construction implementing the optimum approximation of the reduced-rank LMMSE filter based on the received observations. Certainly, this filter estimator can be readily shown to be consistent in the classical sense, i.e., it converges stochastically to the true optimum filter as the sample-size tends to infinity whereas all other dimensions in the signal model remain constant. In fact, aditionally to the complexity reduction mentioned above associated with the suboptimal reduced-rank solution, such an implementation has already proved to enhance the performance in finite sample-size scenarios.

Interestingly enough, the subspace coefficients (i.e., the degrees of freedom afforded by the reduced-rank approach) can be alternatively used to provide a further improved approximant of the optimum solution under limited sample-support constraints. Indeed, if all dimensions are used to approximate the optimum filter (i.e., $D = M$), the Krylov matrix constructed by replacing $\mathbf{R}$ in the optimum $\mathbf{S}_D$ in Lemma 2.1 with the SCM ($N \geq M$), has clearly full-rank with probability one[5]. Consequently, the range of $\hat{\mathbf{S}}_M$ defines, in principle[6], the total space of dimension $M$ embedding the optimum filter. In particular, regarding the (practically meaningful) case in which $D < M$, we concentrate on filter constructions of the form $\mathbf{w} = \hat{\mathbf{S}}_D \boldsymbol{\omega}$ and focus on the set of coefficients minimizing the following empirical performance measure, namely,

$$\breve{\boldsymbol{\omega}}_D = \arg\min_{\boldsymbol{\omega}} \mathsf{MSE}\left(\hat{\mathbf{S}}_D \boldsymbol{\omega}\right), \tag{2.19}$$

where we have defined

$$\hat{\mathbf{S}}_D = \left[\mathbf{h}\, \hat{\mathbf{R}}\mathbf{h} \cdots \hat{\mathbf{R}}^{D-1}\mathbf{h}\right],$$

with $\hat{\mathbf{R}}$ being the SCM, i.e.,

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{y}(n)\,\mathbf{y}^H(n). \tag{2.20}$$

Hence, the filter construction designed using the extra degrees of freedom available as in (2.19) takes the form

$$\breve{\mathbf{w}}_D = \hat{\mathbf{S}}_D \breve{\boldsymbol{\omega}}_D = \hat{\mathbf{S}}_D \left(\hat{\mathbf{S}}_D^H \mathbf{R} \hat{\mathbf{S}}_D\right)^{-1} \hat{\mathbf{S}}_D^H \mathbf{h}, \tag{2.21}$$

attaining the following minimum of the mean-square error performance measure, namely,

$$\mathsf{MSE}\left(\breve{\mathbf{w}}_D\right) = 1 - \mathbf{h}^H \hat{\mathbf{S}}_D \left(\hat{\mathbf{S}}_D^H \mathbf{R} \hat{\mathbf{S}}_D\right)^{-1} \hat{\mathbf{S}}_D^H \mathbf{h}. \tag{2.22}$$

[5] Note that in practice, almost surely, the signature vector $\mathsf{h}$ has a non-zero component in every direction defining the eigenspace of $\hat{\mathsf{R}}$.

[6] As a conclusion of the power method [Mey00], the Krylov matrix $\mathsf{M}$ becomes quickly rank-deficient as its columns get linearly dependent converging towards the dominant eigenvector of $\mathsf{R}$. In practice, an orthogonal basis for the Krylov subspace is rather selected in one way or another in order to improve the numerical stability.

The expression in (2.22) constitutes a lower-bound on the mean-square error achieved in practice by any possible filter implementations of the form $\mathbf{w} = \hat{\mathbf{S}}_D \boldsymbol{\omega}$. Observe that the optimum filter coefficients in (2.19) can be expressed as

$$\breve{\boldsymbol{\omega}}_D = \breve{\mathbf{B}}^{-1} \hat{\mathbf{v}}, \tag{2.23}$$

where, for $k, l = 1, \ldots, D$, we have defined $\left[\breve{\mathbf{B}}\right]_{k,l} = \mathbf{h}^H \hat{\mathbf{R}}^{k-1} \mathbf{R} \hat{\mathbf{R}}^{l-1} \mathbf{h}$ and $[\hat{\mathbf{v}}]_k = \mathbf{h}^H \hat{\mathbf{R}}^{k-1} \mathbf{h}$. Therefore, the problem of properly approximating the optimum coefficients $\breve{\boldsymbol{\omega}}_D$ is equivalent to that of finding a good approximation of the quantities $\left\{\mathbf{h}^H \hat{\mathbf{R}}^i \mathbf{R} \hat{\mathbf{R}}^j \mathbf{h}\right\}_{i,j=0,\ldots,D-1}$, involving both the true covariance matrix as well as its sample estimate.

In this chapter, we take a step further into the construction of optimal approximations of the reduced-rank LMMSE filter and try to derive an improved solution based on properly estimating the entries of the matrix $\breve{\mathbf{B}}$. Since the finite-dimensional statistical analysis of these quantities is rather intricate, we will approach the problem via their approximation in an appropriate asymptotic regime. Note that a limiting regime based on large-sample theory provides us with asymptotically exact candidates only for an infinite number of samples, as it is assumed whenever the SCM is used to directly replace the true covariance matrix. However, the sample-size can never be considered in practice to be infinitely larger than any other dimension in the signal model. Hence, we consider instead an approximation regime that better models a finite sample-size situation (characterized by the fact that both the number of observations available and their dimension are comparable in magnitude) and focus on a doubly-asymptotic regime assuming that both dimension and size of the sample increase without bound at the same rate.

In order to derive a solution according to the framework introduced above, we make use of random matrix theory to obtain a set of (asymptotically optimum) approximants of the previous key quantities in terms of the SCM. In particular, we follow a two-step strategy by first providing an asymptotically equivalent expression of the entries of $\breve{\mathbf{B}}$ (involving both the true covariance matrix and its sample estimate) as a function of the spectrum of $\mathbf{R}$, and afterwards obtain a consistent estimator of these spectral functions based on the knowledge of $\hat{\mathbf{R}}$.

## 2.5 Asymptotic Analysis of the Proposed Filter Structure

Much work has been published over the last years on the asymptotic weighting [Mül01, Hac04, Li04b, Cot05] and performance analysis [Hon01, Lou03] of reduced-rank CDMA receivers minimizing the MSE based on the Krylov subspace introduced above (see also [Pan07] for PCA eigensubspaces). In most cases, the analysis relies on the statistical modeling of the actual structure of the observation covariance matrix in terms of a random transmission channel. In the previous works, both the filter weights and the asymptotic performance of the receiver are obtained as a function of the distributions of the signatures and noise, as well as some other intrinsic system parameters, such as the filter length and the signal subspace dimension. Here, on the contrary, we deal with the practical problem of designing MMSE estimators based on empirical statistics when the only available knowledge about the signal structure is the signature associated with the unknown parameter. To that effect, we find the set of optimum (subspace filter) coefficients minimizing the MSE over an approximation subspace constructed from the received samples.

As an intermediate step, in this section we derive the asymptotic expression of the scalar quantites $\left\{\mathbf{h}^H\hat{\mathbf{R}}^i\mathbf{R}\hat{\mathbf{R}}^j\mathbf{h}\right\}_{i,j=0,\ldots,D-1}$ in terms of only the true covariance matrix $\mathbf{R}$ and under the assumption that not only the sample-size $N$ but also the observation dimension $M$ increases without bound with a fixed ratio between them ($M/N \to c$). The rationale behind this strategy is twofold. First, an expression characterizing the large-system performance of the reduced-rank LMMSE estimator for a limited sample-size per filtering degrees of freedom is obtained as a function of the set of weights, the number of samples per observation dimension and the covariance matrix of the input signals. On the other hand, the latter functions of $\mathbf{R}$ can be consistently estimated using the spectrum of the sample covariance matrix $\hat{\mathbf{R}}$ (cf. Section 2.6), consequently providing a practical procedure for the calculation of the optimum MMSE filter weights. Under this asymptotic framework, the Stieltjes transform from random matrix theory allows us to characterize the asymptotic distribution of the eigenvalues of $\hat{\mathbf{R}}$ in terms of the limiting eigenvalue distribution of $\mathbf{R}$. Throughout the chapter, the following statistical assumptions of purely technical interest regarding the signal model are used:

(**As1**) The observation vectors $\mathbf{y}(n)$ can be statistically modeled as $\mathbf{y}(n) = \mathbf{R}^{1/2}\mathbf{u}(n)$, where $\mathbf{u}(n) \in \mathbb{C}^M$, $n = 1, \ldots, N$, is a collection of i.i.d. random vectors, whose entries have zero mean real and imaginary parts with variance $1/2$ and bounded higher moments.

(**As2**) The matrix $\mathbf{R}$ has uniformly bounded spectral radius for all $M$.

Thus, for the subsequent analysis, the SCM in (2.20) will be modeled according to (**As1**) as

$$\hat{\mathbf{R}} = \frac{1}{N}\mathbf{R}^{1/2}\mathbf{U}\mathbf{U}^H\mathbf{R}^{1/2}, \tag{2.24}$$

where the matrix $\mathbf{U} \in \mathbb{C}^{M\times N}$ is constructed using as columns the vectors $\mathbf{u}(n)$, $n = 1, \ldots, N$.

We are now ready to proceed with the presentation of the asymptotic approximation of the term $\mathbf{h}^H\hat{\mathbf{R}}^i\mathbf{R}\hat{\mathbf{R}}^j\mathbf{h}$ in the reduced-rank empirical performance measure involving only the true covariance matrix. The basic idea behind our approach builds on the following identity, that can be obtained directly from the Cauchy integral formula in two complex variables, namely

$$\mathbf{h}^H\hat{\mathbf{R}}^i\mathbf{R}\hat{\mathbf{R}}^j\mathbf{h} = \frac{1}{(2\pi\,\mathrm{j})^2}\oint_\Gamma\oint_\Gamma z_1^i z_2^j\ \eta(z_1, z_2)\ dz_1 dz_2, \tag{2.25}$$

where

$$\eta(z_1, z_2) = \mathbf{h}^H\left(\hat{\mathbf{R}} - z_1\mathbf{I}_M\right)^{-1}\mathbf{R}\left(\hat{\mathbf{R}} - z_2\mathbf{I}_M\right)^{-1}\mathbf{h}, \tag{2.26}$$

and the region $\Gamma$ is defined by a simply closed contour enclosing all the eigenvalues of $\hat{\mathbf{R}}$. From (2.25) and the dominated convergence theorem, the problem of obtaining an asymptotic equivalent of $\mathbf{h}^H\hat{\mathbf{R}}^i\mathbf{R}\hat{\mathbf{R}}^j\mathbf{h}$ reduces to that of finding the limiting expression of the integrand in (2.25). Indeed, the integrand in (2.25) can be shown to be almost surely bounded for all large $M$,$N$ over a compact subset containing the integration contour. Thus, we are allowed to invoke the dominated convergence theorem and integrate directly the limiting expression in order to obtain the asymptotic value of $\mathbf{h}^H\hat{\mathbf{R}}^i\mathbf{R}\hat{\mathbf{R}}^j\mathbf{h}$. An asymptotic equivalent of $\eta(z_1, z_2)$ in (2.26) is provided by the next theorem.

**Theorem 2.2** *Let $\hat{\mathbf{R}} = \mathbf{R}^{1/2}\mathbf{\Xi}\mathbf{\Xi}^H\mathbf{R}^{1/2}$ with $\mathbf{R}^{1/2}$ the positive square-root of the deterministic Hermitian positive definite $M \times M$ matrix $\mathbf{R}$, with uniformly bounded eigenvalues $\lambda_1 \leq \ldots \leq \lambda_M$. Let also $\mathbf{\Xi}$ denote an $M \times N$ complex random matrix, such that the real and imaginary parts of the entries of $\sqrt{N}\mathbf{\Xi}$ are independent and identically distributed with mean zero, variance $1/2$ and bounded moments. Finally, consider also an $M$-dimensional deterministic complex vector $\mathbf{h}$ with uniformly bounded Euclidean norm for all $M$. Then, for any $z_1, z_2 \in \mathbb{C}^+$, $\eta(z_1, z_2) \asymp \bar{\eta}(z_1, z_2)$, where*

$$\bar{\eta}(z_1, z_2) = \frac{f(z_1) - f(z_2)}{z_1 - z_2}\mathbf{h}^H\left(w(z_1)\mathbf{R} - z_1\mathbf{I}_M\right)^{-1}\mathbf{R}\left(w(z_2)\mathbf{R} - z_2\mathbf{I}_M\right)^{-1}\mathbf{h}, \tag{2.27}$$

*and we have further defined*

$$f(z) = \frac{z}{w(z)}, \quad w(z) = 1 - c - czb(z), \tag{2.28}$$

*with $b(z) = b$ being the unique solution to the following equation in the set $\{b \in \mathbb{C} : -(1-c)/z + cb \in \mathbb{C}^+\}$:*

$$b = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{\lambda_m(1 - c - czb) - z}.$$

**Proof.** See Appendix A. ■

**Corollary 2.1 (Asymptotic reduced-rank empirical MMSE)** *In the doubly-asymptotic regime defined above, the MMSE lower-bound in (2.22) converges as*

$$\mathsf{MSE}(\hat{\mathbf{w}}_D) \asymp \overline{\mathsf{MSE}} = 1 - \bar{\mathbf{v}}^H\bar{\mathbf{B}}^{-1}\bar{\mathbf{v}},$$

*where, for $k, l = 1, \ldots, D$,*

$$\left[\bar{\mathbf{B}}\right]_{k,l} = \frac{1}{(2\pi\mathrm{j})^2}\oint_\Gamma\oint_\Gamma z_1^k z_2^l\ \bar{\eta}(z_1, z_2)\ dz_1 dz_2,$$

*and $[\bar{\mathbf{v}}]_k$ is the asymptotic limit of $\mathbf{h}^H\hat{\mathbf{R}}^k\mathbf{h}$ (cf. Chapter 4).*

Observe that the limiting performance measure found by substituting, for $i, j = 0 \dots D-1$, every $\mathbf{h}^H \hat{\mathbf{R}}^i \mathbf{R} \hat{\mathbf{R}}^j \mathbf{h}$ in (2.22) by the asymptotic equivalents obtained in this section extends previous performance analyses to the case of sample observations of comparably large size and dimension. Under the conventional assumption of a number of samples infinitely larger than the observation dimension (i.e., $c = 0$), our general asymptotic performance measure coincides with that in the existing literature.

In the following section, an estimator of the quantities $\left\{ \mathbf{h}^H \hat{\mathbf{R}}^i \mathbf{R} \hat{\mathbf{R}}^j \mathbf{h} \right\}_{i,j=0,\dots,D-1}$, or (asymptotically) equivalently of their limits above, is derived using the asymptotic description of the spectrum of sample covariance matrices, that is consistent even when the number of samples per observation dimension remains bounded.

## 2.6 Consistent MMSE Estimator

Conventionally, the construction of LMMSE filters from a collection of received observations is based on directly replacing the true covariance matrix with the SCM. Estimators obtained following this procedure are consistent as the number of samples grows large without bound. However, filter operation conditions in practical finite sample-size situations are characterized by a limited number of samples per observation dimension. Under this practical contraint, the conventional estimator of the subspace coefficients $\breve{\boldsymbol{\omega}}_D$ in (2.23) is no longer consistent. Here, in order to improve the estimation performance under realistic conditions, we propose a more general estimator of $\breve{\boldsymbol{\omega}}_D$ that is consistent for arbitrarily high-dimensional observations (i.e., for a bounded sample-support per degree-of-freedom).

In order to construct the proposed generalized LMMSE filter, we depart from the asymptotic analysis in Section 2.5. The asymptotic approximations based on the (unknown) true covariance matrix obtained in the previous section are motivated by the fact that certain spectral functions of $\mathbf{R}$ may be arbitrarily well approximated in the previous asymptotic regime by quantities depending only on the spectrum of the sample covariance matrix $\hat{\mathbf{R}}$. In particular, approximations of this type can be found by resorting to the theory of general statistical analysis developed by Girko in e.g. [Gir98, Chapter 14]. G-analysis provides a systematic approach to derive estimators of a certain class of spectral functions of the theoretical covariance matrix that are consistent even in situations characterized by a collection of arbitrarily large-dimensional observations (namely ressembling the double-limiting regime hereby considered). As a consequence, G-estimators can be identified as $M, N$-consistent as opposed to traditional $N$-consistent estimators. The following theorem extends the theory of GSA with an $M, N$-consistent estimator of the set of spectral functions in (2.25) derived in Section 2.5.

**Theorem 2.3** *Let* $\hat{\mathbf{R}}$, $\mathbf{R}$ *and* $\mathbf{h}$ *be defined as in Theorem 2.3. Then,*

$$\mathbf{h}^H \hat{\mathbf{R}}^i \mathbf{R} \hat{\mathbf{R}}^j \mathbf{h} \asymp \frac{1}{c^2} \sum_{k=1}^{M} \sum_{l=1}^{M} \frac{\mu_k^i \mu_l^j}{\eta_k \eta_l} \sum_{m=1}^{M} \frac{\hat{\lambda}_m \left| \mathbf{h}^H \hat{\mathbf{e}}_m \right|^2}{\left( \hat{\lambda}_m - \mu_k \right) \left( \hat{\lambda}_m - \mu_l \right)}, \tag{2.29}$$

*where* $\hat{\lambda}_m$ *and* $\hat{\mathbf{e}}_m$ *are the mth eigenvalue and mth eigenvector of* $\hat{\mathbf{R}}$*, respectively,*

$$\eta_p = \frac{1}{M} \sum_{m=1}^{M} \frac{\hat{\lambda}_m}{\left( \hat{\lambda}_m - \mu_p \right)^2},$$

*and* $\mu_p$*,* $p = 1, \ldots, M$*, are the real-valued solutions of the following equation in* $\mu$

$$\frac{1}{M} \sum_{m=1}^{M} \frac{\hat{\lambda}_m}{\hat{\lambda}_m - \mu} = \frac{1}{c}.$$

**Proof.** See Appendix G. ■

**Remark 2.1** *The RHS of (2.29) is a strongly consistent estimator of the quantity in the LHS.*

Therefore, the proposed implementation of the reduced-rank LMMSE filter can be constructed as $\breve{\mathbf{w}}_D = \hat{\mathbf{S}}_D \breve{\boldsymbol{\omega}}_D$, where

$$\breve{\boldsymbol{\omega}}_D = \breve{\mathbf{B}}^{-1} \hat{\mathbf{v}},$$

and, for $k, l = 1, \ldots, D$, $\left[ \breve{\mathbf{B}} \right]_{k,l}$ is defined by the consistent estimator of $\mathbf{h}^H \hat{\mathbf{R}}^{k-1} \mathbf{R} \hat{\mathbf{R}}^{l-1} \mathbf{h}$ given by (2.29).

## 2.7 Evaluation of non-asymptotic estimator performance

In this section, the goodness of the new generalized estimator is evaluated under non-asymptotic conditions via numerical simulations. Essentially, it will be empirically shown that particularly the fact that the proposed implementation generalizes the traditional estimator (by proving to be consistent even for arbitrarily high-dimensional observations) traduces to an improved performance also in the finite sample-size regime.

Throughout the following exposition, the proposed realizable construction of the optimum reduced-rank LMMSE filter is compared with the conventional counterpart obtained by directly replacing the covariance matrix with its sample estimate. In all the simulations, the covariance matrix is modeled according to (2.10), whereas the SCM in (2.20) is modeled as in (2.24). The dimension of the signal subspace is chosen as a fraction of the observation dimension, namely $K = \alpha M$. Furthermore, all vector signatures are generated as realizations of a complex random vector with i.i.d. entries having real and imaginary parts of mean zero and unit variance.

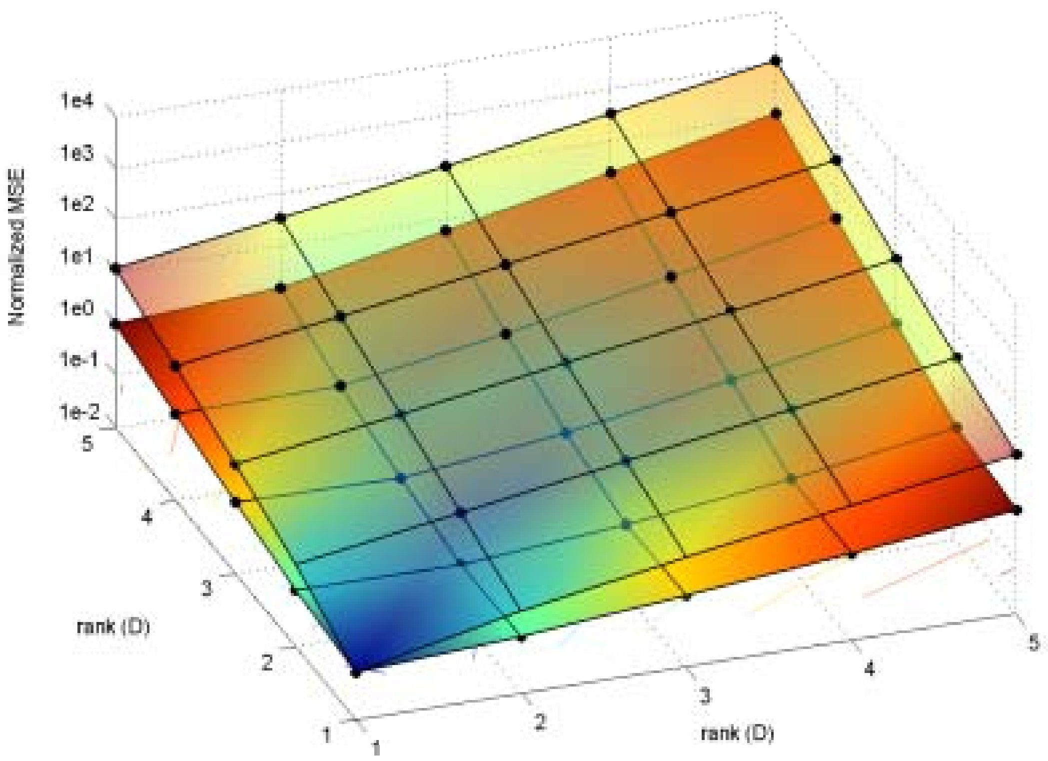


Figure 2.2: Normalized mean squared error incurred in the proposed and conventional estimation of the individual entries of $\mathbf{B}$.

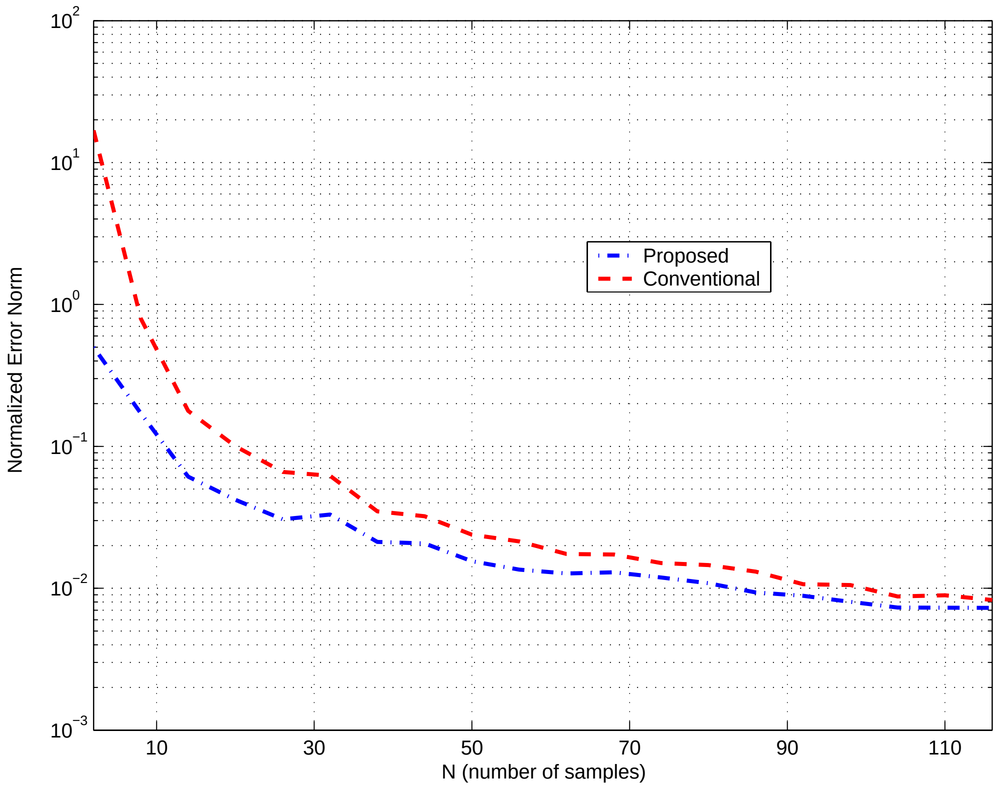


Figure 2.3: Estimator convergence for an increasing number of samples

Finally, the noise power is fixed to $\sigma_n^2 = 1$, and the desired and the $K-1$ nuisance signatures are received with a power of 10 dB and 15 dB, respectively, over the noise floor. In order to assess the comparative performance of the $M,N$-consistent estimator $\breve{\mathbf{B}}$, we first evaluate the error incurred by using both approximations when estimating the individual entries of the matrix $\breve{\mathbf{B}}$. Figure 2.2 shows the normalized mean-square error achieved by the proposed (bottom) and the conventional (top) estimators, namely

$$\left| \frac{\left[\breve{\mathbf{B}}\right]_{i,j} - \left[\breve{\mathbf{B}}\right]_{i,j}}{\left[\mathbf{B}\right]_{i,j}} \right|^2 \text{ and } \left| \frac{\left[\breve{\mathbf{B}}\right]_{i,j} - \left[\hat{\mathbf{B}}\right]_{i,j}}{\left[\mathbf{B}\right]_{i,j}} \right|^2,$$

respectively. The averaged error over 100 realizations is depicted for the case of $M = 8$, $N = 10$ and $K = 6$. On the other hand, the consistency of both estimators as the number of samples grows large is illustrated in Figure 2.3, where the normalized squared Frobenius norm of the error matrix for the proposed and conventional estimators of the matrix $\breve{\mathbf{B}}$, namely

$$\frac{\left\| \breve{\mathbf{B}} - \breve{\mathbf{B}} \right\|_F^2}{\left\| \mathbf{B} \right\|_F^2} \text{ and } \frac{\left\| \breve{\mathbf{B}} - \hat{\mathbf{B}} \right\|_F^2}{\left\| \mathbf{B} \right\|_F^2},$$

respectively, is averaged over 300 realizations with $M = 4$, $K = 3$ and $D = 2$, and all signatures received 5 dB over the noise floor. Clearly, a higher convergence rate can be appreciated

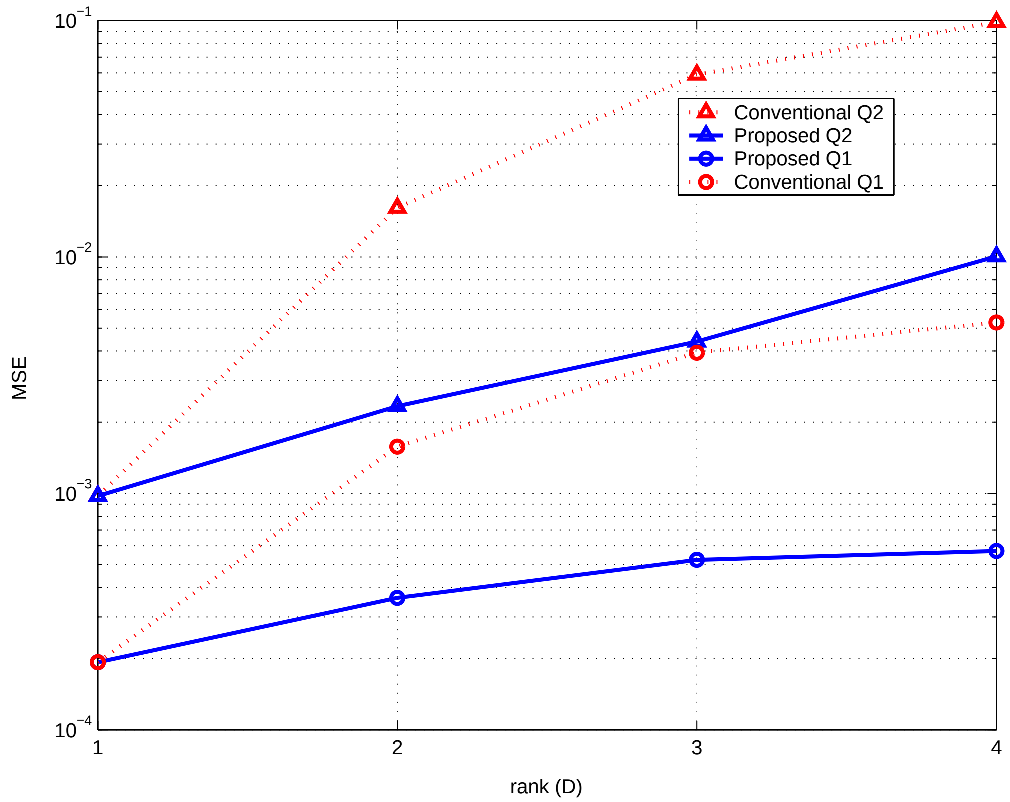


Figure 2.4: Averaged MSE of the estimation of $Q_1(\mathbf{w})$ and $Q_2(\mathbf{w})$

regarding the proposed estimator.

It is a well-known fact in the literature that the optimum linear reduced-rank estimator minimizing the MSE simultaneously achieves the maximum output SINR. Indeed, observe that also the filter in (2.21) does attain the optimum tradeoff between minimizing MSE and maximizing SINR (MSINR). In practice, under non-asymptotic conditions, the approximation error incurred by the estimator of $\check{\mathbf{B}}$ prevents the MMSE and MSINR criteria to be simultaneously attained. In order to assess the effect of the errors in the estimation of the individual entries of $\check{\mathbf{B}}$ in the practically achieved estimation accuracy, it will be in order to define $Q_1(\mathbf{w}) = \mathbf{w}^H\mathbf{h}$ and $Q_2(\mathbf{w}) = \mathbf{w}^H\mathbf{R}\mathbf{w}$. Accordingly, we can rewrite both performance measures as

$$\mathsf{MSE}(\mathbf{w}) = 1 - 2\,\mathrm{Re}\{Q_1(\mathbf{w})\} + Q_2(\mathbf{w}), \qquad \mathsf{SINR}(\mathbf{w}) = \left(\frac{Q_2(\mathbf{w})}{Q_1^2(\mathbf{w})} - 1\right)^{-1}. \tag{2.30}$$

Clearly, as argued above, if $\mathbf{w} = \check{\mathbf{w}}_D$, $Q_1(\check{\mathbf{w}}_D) = Q_2(\check{\mathbf{w}}_D) = \hat{\mathbf{v}}^H\check{\mathbf{B}}^{-1}\hat{\mathbf{v}}$ and both performance measures are simultaneously optimized, whereas, for an estimated filter construction, the achieved MMSE is asymmetrically traded off against the achieved MSINR obtained at the output of the reduced-rank filter. Before evaluating the effect of the approximation error of the previous quantities on the performance tradeoff, the distance (in terms of the mean square-error)

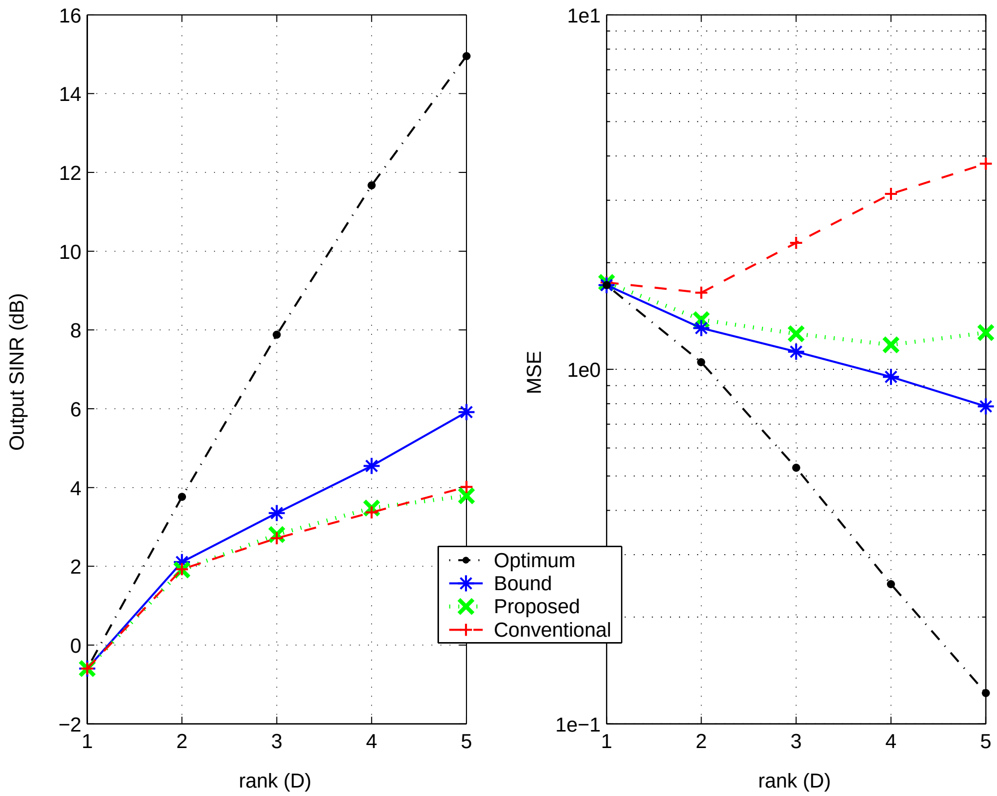


Figure 2.5: Conventional and proposed filter performance tradeoff ($M = 30$ and $N = 40$).

from each quantity to its actual value meeting the MMSE and MSINR criteria simultaneously is illustrated in Figure 2.4 for the case $M = 16$ and $N = 20$ averaged over 100 realizations. An increasing error around one order of magnitude larger can be observed for the conventional estimators of $Q_1(\check{\mathbf{w}}_D)$ and $Q_2(\check{\mathbf{w}}_D)$.

The proposed and the conventional reduced-rank LMMSE estimators are compared in terms of the achieved performance. In particular, curves representing the MSE and SINR performance of both estimators are depicted versus rank selection along with the performance bound attained by $\hat{\mathbf{w}}_D$ and the reduced-rank optimum (clairvoyant) case. Figure 2.5 shows the achieved MSE and SINR in the case of $M = 30$ and $N = 40$. Even though both the conventional and the proposed implementation of the optimum reduced-rank filter attain an improved output SINR of around 5 dB with respect to the matched filter, the conventional realization experiences a sharp degradation in terms of MSE, whereas the proposed filter is even able to reduce the error. The filter performance for a rank-deficient SCM, specifically for $M = 30$ and $N = 20$ ($c > 1$), is shown in Figure 2.6. In this figure, an even superior enhancement in terms of MSE performance gap (namely twice as high as in the previous figure) can be appreciated. Finally, an improved estimation accuracy can be observed for the large-signal case in Figure 2.7,where the performance of the proposed estimator closely resembles the bound performance.

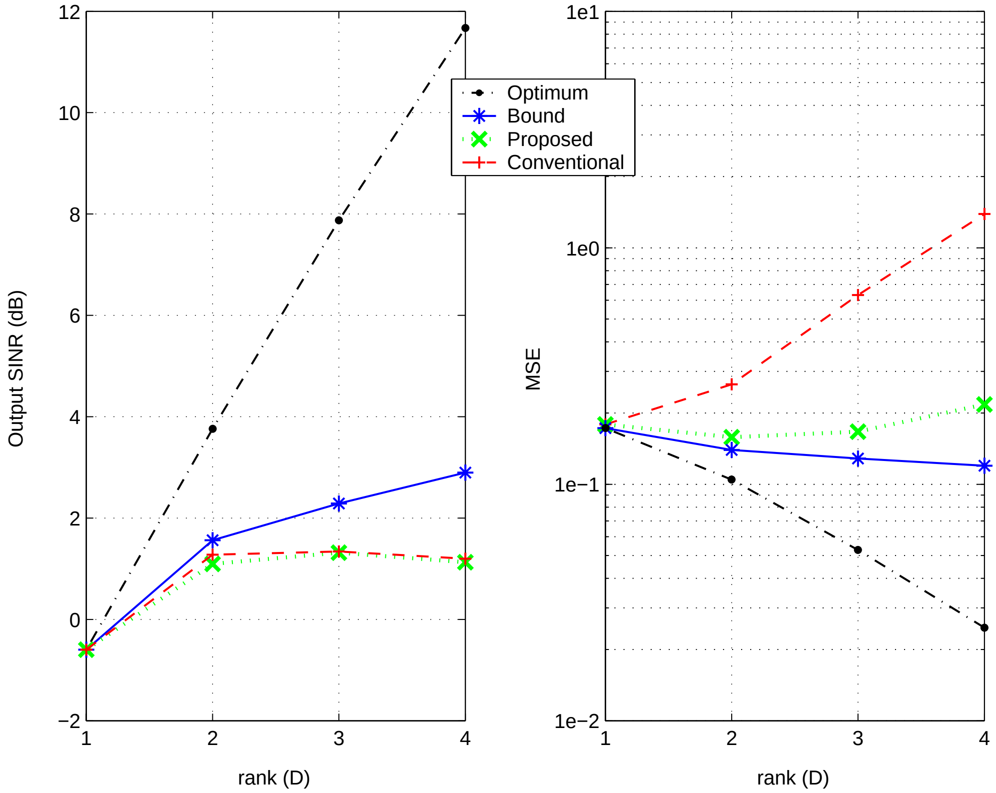


Figure 2.6: Conventional and proposed filter performance tradeoff ($M = 30$ and $N = 20$).

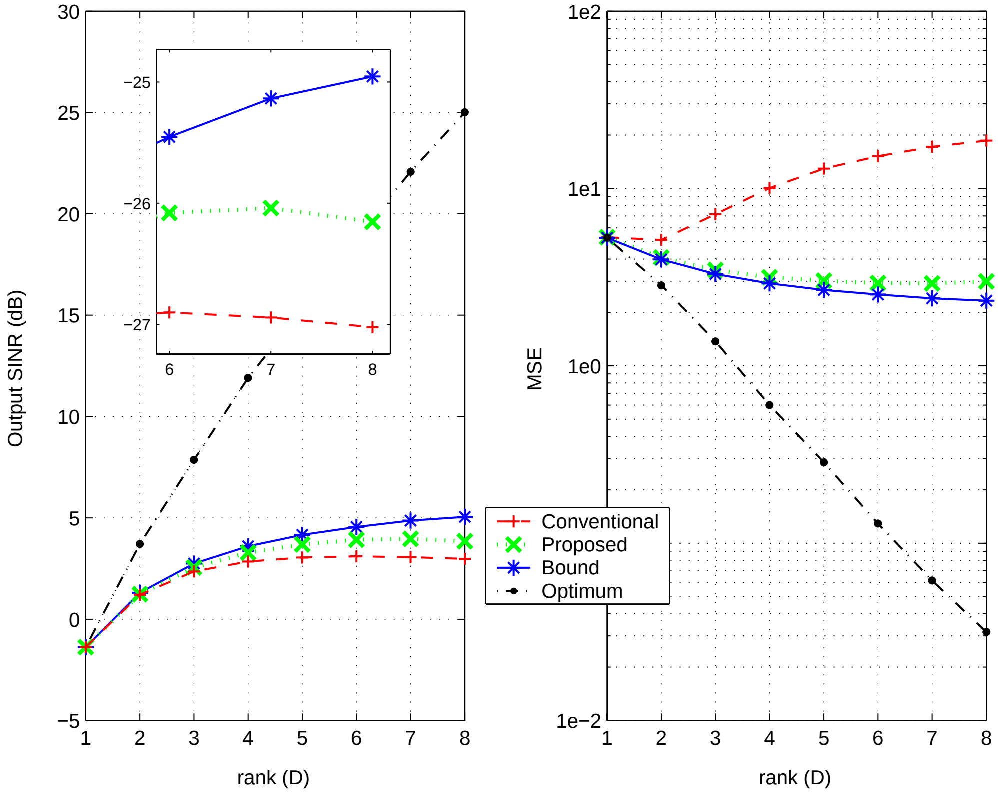


Figure 2.7: Conventional and proposed filter performance tradeoff ($M = 80$ and $N = 100$).

The $M, N$-consistency associated with the subspace coefficient estimator introduced in the previous section allows for a reasonable filter performance under fairly less stringent conditions on the number of samples per observation dimension, thereby outperforming the conventional $M$-consistent reduced-rank LMMSE filter in finite sample-size situations. The last and final section summarizes the contribution of the chapter.

## 2.8 Conclusion

The construction of optimum approximations of the LMMSE estimator on a reduced-dimensional subspace is a fundamental problem in many disciplines and, particularly, in communication theory and signal processing applications. Realizable implementations of the reduced-rank LMMSE filter derived as a function of the unknown second-order statistics of the input observations are traditionally based on the direct substitution of the theoretical covariance matrix for its sample estimate, namely the SCM. Given a matrix with column space spanning the approximation subspace, we use the coefficients linearly combining its columns as a set of degrees of freedom available for filter design. Building upon the conventional implementation of the optimal Krylov subspace matrix in terms of the SCM, an improved construction of the reduced-rank LMMSE filter is proposed that is consistent under more general conditions than the conventional implementation. In particular, the proposed estimator generalizes conventional filter realizations, consistent in the classical sence, by guaranteeing consistency even for arbitrarily high-dimensional observations. Our results are based on a general description in the previous doubly-asymptotic regime of the spectrum of SCM-type random matrix models that is also applicable to the undersampled case. Accordingly, this asymptotic regime is in perfect agreement with realistic deployment settings in practice, characterized by a bounded ratio between sample and system sizes, in the sense that it allows for the possibility of the number of received observations and the number of filtering degrees of freedom being comparable in magnitude. As a result, the proposed generalized LMMSE estimator is shown via numerical simulation to present a superior performance under finite sample-size situations avoiding the degradation in terms of MSE performance as the selected rank value increases.

## Appendix 2.A Proof of Proposition 2.2

Throughout the following proof, all limiting expressions must be understood in the asymptotic regime defined when both $M$ and $N$ go to infinity at the same rate fixed by the constant $c = M/N$. In order to proof Theorem 2.2, we proceed by decomposing $\eta\left(z_1, z_2\right)$ as a sum involving $\bar{\eta}\left(z_1, z_2\right)$ and a finite number of quantities vanishing almost surely as $M, N \to \infty$. Consider a random matrix $\mathbf{Y} = \left[\mathbf{y}_1, \ldots, \mathbf{y}_2\right]$ such that

$$\hat{\mathbf{R}} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H = \hat{\mathbf{R}}_n + \frac{1}{N}\mathbf{y}_n\mathbf{y}_n^H = \hat{\mathbf{R}}_{mn} + \frac{1}{N}\mathbf{y}_m\mathbf{y}_m^H + \frac{1}{N}\mathbf{y}_n\mathbf{y}_n^H,$$

where the matrices $\hat{\mathbf{R}}_n$ and $\hat{\mathbf{R}}_{mn}$ are defined as

$$\hat{\mathbf{R}}_n = \hat{\mathbf{R}} - \frac{1}{N}\mathbf{y}_n\mathbf{y}_n^H;\ \hat{\mathbf{R}}_{mn} = \hat{\mathbf{R}} - \frac{1}{N}\mathbf{y}_n\mathbf{y}_n^H - \frac{1}{N}\mathbf{y}_m\mathbf{y}_m^H.$$

For the sake of notational convenience, we will use the following definitions in our derivations

$$\mathbf{Q}\left(z\right) = \left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1}, \tag{2.31}$$

$$\mathbf{Q}_n\left(z\right) = \left(\hat{\mathbf{R}}_n - z\mathbf{I}_M\right)^{-1}, \tag{2.32}$$

$$\mathbf{P}\left(z\right) = \left(w\left(z\right)\mathbf{R} - z\mathbf{I}_M\right)^{-1}. \tag{2.33}$$

First, we observe that

$$\begin{aligned}\eta\left(z_1, z_2\right) &= \mathbf{h}^H\mathbf{Q}\left(z_1\right)\mathbf{R}\mathbf{Q}\left(z_2\right)\mathbf{h} \\ &= \mathbf{h}^H\left(\mathbf{Q}\left(z_1\right) - \mathbf{P}\left(z_1\right)\right)\mathbf{R}\left(\mathbf{Q}\left(z_2\right) - \mathbf{P}\left(z_2\right)\right)\mathbf{h} + \mathbf{h}^H\mathbf{P}\left(z_1\right)\mathbf{R}\mathbf{P}\left(z_2\right)\mathbf{h} \\ &\quad + \mathbf{h}^H\left(\mathbf{Q}\left(z_1\right) - \mathbf{P}\left(z_1\right)\right)\mathbf{R}\mathbf{P}\left(z_2\right)\mathbf{h} + \mathbf{h}^H\mathbf{P}\left(z_1\right)\mathbf{R}\left(\mathbf{Q}\left(z_2\right) - \mathbf{P}\left(z_2\right)\right)\mathbf{h}.\end{aligned}$$

Define the vectors $\mathbf{h}_1 = \mathbf{P}_1^H\mathbf{h}$ and $\mathbf{h}_2 = \mathbf{P}_2\mathbf{h}$. Then, using the fact that (cf. Section 1.1)

$$\begin{aligned}\mathbf{h}^H\mathbf{Q}\left(z_1\right)\mathbf{R}\mathbf{h}_1 &\asymp \mathbf{h}^H\mathbf{P}\left(z_1\right)\mathbf{R}\mathbf{h}_1 \\ \mathbf{h}_2^H\mathbf{R}\mathbf{Q}\left(z_2\right)\mathbf{h} &\asymp \mathbf{h}_2^H\mathbf{R}\mathbf{P}\left(z_2\right)\mathbf{h},\end{aligned}$$

we can write

$$\begin{aligned}\eta\left(z_1, z_2\right) &\asymp \mathbf{h}^H\left(\mathbf{Q}\left(z_1\right) - \mathbf{P}\left(z_1\right)\right)\mathbf{R}\left(\mathbf{Q}\left(z_2\right) - \mathbf{P}\left(z_2\right)\right)\mathbf{h} + \mathbf{h}_1^H\mathbf{R}\mathbf{h}_2 \\ &= \mathbf{h}_1^H\left(w(z_1)\mathbf{R} - \hat{\mathbf{R}}\right)\mathbf{Q}\left(z_1\right)\mathbf{R}\mathbf{Q}\left(z_2\right)\left(w(z_2)\mathbf{R} - \hat{\mathbf{R}}\right)\mathbf{h}_2 + \mathbf{h}_1^H\mathbf{R}\mathbf{h}_2 \\ &= \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\mathbf{h}_1^H\left(w(z_1)\mathbf{R} - \mathbf{y}_n\mathbf{y}_n^H\right)\mathbf{Q}\left(z_1\right)\mathbf{R}\mathbf{Q}\left(z_2\right)\left(w(z_2)\mathbf{R} - \mathbf{y}_m\mathbf{y}_m^H\right)\mathbf{h}_2 + \mathbf{h}_1^H\mathbf{R}\mathbf{h}_2,\end{aligned} \tag{2.34}$$

where, in the first equality, we have used the resolvent identity, i.e., $\mathbf{A}^{-1} - \mathbf{B}^{-1} = \mathbf{B}^{-1}\left(\mathbf{A} - \mathbf{B}\right)\mathbf{A}^{-1}$.

Define further

$$\hat{\omega}_n(z) = \frac{1}{1 + \frac{1}{N}\mathbf{y}_n^H \mathbf{Q}_n(z)\mathbf{y}_n}.$$

Observe now that the RHS of (2.34) can be equivalently written as

$$\eta(z_1, z_2) \asymp \mathbf{h}_1^H \mathbf{R} \mathbf{h}_2 \tag{2.35}$$

$$+ \left( w(z_1) - \frac{1}{N}\sum_{n=1}^{N} \hat{\omega}_n(z_1) \right) \mathbf{h}_1^H \mathbf{R}\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\mathbf{R}\mathbf{h}_2 \left( w(z_2) - \frac{1}{N}\sum_{m=1}^{N} \hat{\omega}_n(z_2) \right) \tag{2.36}$$

$$+ \left( w(z_1) - \frac{1}{N}\sum_{n=1}^{N} \hat{\omega}_n(z_1) \right) \mathbf{h}_1^H \mathbf{R}\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2) \frac{1}{N}\sum_{m=1}^{N} \left( \hat{\omega}_n(z_2)\mathbf{R} - \mathbf{y}_m\mathbf{y}_m^H \right) \mathbf{h}_2 \tag{2.37}$$

$$+ \frac{1}{N}\sum_{n=1}^{N} \mathbf{h}_1^H \left( \hat{\omega}_n(z_1)\mathbf{R} - \mathbf{y}_n\mathbf{y}_n^H \right) \mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\mathbf{R}\mathbf{h}_2 \left( w(z_2) - \frac{1}{N}\sum_{m=1}^{N} \hat{\omega}_n(z_2) \right) \tag{2.38}$$

$$+ \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N} \mathbf{h}_1^H \left( \hat{\omega}_n(z_1)\mathbf{R} - \mathbf{y}_n\mathbf{y}_n^H \right) \mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2) \left( \hat{\omega}_n(z_2)\mathbf{R} - \mathbf{y}_m\mathbf{y}_m^H \right) \mathbf{h}_2.$$

We can prove that the expressions in (2.36), (2.37) and (2.38) vanish asymptotically almost surely. Indeed, the convergence for (2.36) is readily given by the fact that (cf. Section 1.1)

$$\frac{1}{N}\sum_{n=1}^{N} \hat{\omega}_n(z_1) \asymp w(z), \tag{2.39}$$

since, clearly,

$$\left\| \left[ \hat{\mathbf{R}} - z\mathbf{I}_M \right]^{-1} \right\| \leq \frac{1}{\operatorname{Im}\{z\}} < +\infty, \tag{2.40}$$

and, consequently,

$$\left| \mathbf{h}_1^H \mathbf{R}\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\mathbf{R}\mathbf{h}_2 \right| \leq \|\mathbf{h}_1\| \|\mathbf{h}_2\| \|\mathbf{R}\|^3 \frac{1}{\operatorname{Im}\{z_1\}} \frac{1}{\operatorname{Im}\{z_2\}} < +\infty.$$

Furthermore, the following convergence result is proved in Appendix B, namely

$$\left( w(z_1) - \frac{1}{N}\sum_{n=1}^{N} \hat{\omega}_n(z_1) \right) \operatorname{Tr}\left[ \mathbf{\Theta}\mathbf{R}\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2) \frac{1}{N}\sum_{m=1}^{N} \left( \hat{\omega}_n(z_2)\mathbf{R} - \mathbf{y}_m\mathbf{y}_m^H \right) \right] \asymp 0, \tag{2.41}$$

where $\mathbf{\Theta} \in \mathbb{C}^{M\times M}$ is an arbitrary matrix with uniformly bounded spectral radius. Using (2.41) with $\mathbf{\Theta} = \mathbf{h}_2\mathbf{h}_1^H$, the convergence of (2.37), and equivalently (2.38), can be readily stated. In conclusion, we have

$$\eta(z_1, z_2) \asymp \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N} \mathbf{h}_1^H \left( \hat{\omega}_n(z_1)\mathbf{R} - \mathbf{y}_n\mathbf{y}_n^H \right) \mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2) \left( \hat{\omega}_n(z_2)\mathbf{R} - \mathbf{y}_m\mathbf{y}_m^H \right) \mathbf{h}_2 + \mathbf{h}_1^H \mathbf{R}\mathbf{h}_2.$$

Moreover, using the matrix inversion lemma to write

$$\mathbf{Q}(z) = \mathbf{Q}_n(z) - \frac{1}{N}\frac{\mathbf{Q}_n(z)\mathbf{y}_n\mathbf{y}_n^H\mathbf{Q}_n(z)}{1+\frac{1}{N}\mathbf{y}_n^H\mathbf{Q}_n(z)\mathbf{y}_n}, \tag{2.42}$$

we see that

$$\frac{1}{N}\sum_{n=1}^{N}\left(\hat{\omega}_n(z_1)\mathbf{R} - \mathbf{y}_n\mathbf{y}_n^H\right)\mathbf{Q}(z_1) = \frac{1}{N}\sum_{n=1}^{N}\frac{\mathbf{R}\mathbf{Q}(z_1) - \mathbf{y}_n\mathbf{y}_n^H\mathbf{Q}_n(z_1)}{1+\frac{1}{N}\mathbf{y}_n^H\mathbf{Q}_n(z_1)\mathbf{y}_n}. \tag{2.43}$$

Using (2.43), we observe that

$$\eta(z_1,z_2) \asymp \mathbf{h}_1^H\mathbf{R}\mathbf{h}_2$$

$$\frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\hat{\omega}_n(z_1)\hat{\omega}_m(z_2)\mathbf{h}_1^H\mathbf{R}\left(\mathbf{Q}(z_1)-\mathbf{Q}_n(z_1)\right)\mathbf{R}\left(\mathbf{Q}(z_2)-\mathbf{Q}_m(z_2)\right)\mathbf{R}\mathbf{h}_2 \tag{2.44}$$

$$+\frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\hat{\omega}_n(z_1)\hat{\omega}_m(z_2)\mathbf{h}_1^H\left(\mathbf{R}-\mathbf{y}_n\mathbf{y}_n^H\right)\mathbf{Q}_n(z_1)\mathbf{R}\left(\mathbf{Q}(z_2)-\mathbf{Q}_m(z_2)\right)\mathbf{R}\mathbf{h}_2 \tag{2.45}$$

$$+\frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\hat{\omega}_n(z_1)\hat{\omega}_m(z_2)\mathbf{h}_1^H\mathbf{R}\left(\mathbf{Q}(z_1)-\mathbf{Q}_n(z_1)\right)\mathbf{R}\mathbf{Q}_m(z_2)\left(\mathbf{R}-\mathbf{y}_m\mathbf{y}_m^H\right)\mathbf{h}_2 \tag{2.46}$$

$$+\frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\hat{\omega}_n(z_1)\hat{\omega}_m(z_2)\mathbf{h}_1^H\left(\mathbf{R}-\mathbf{y}_n\mathbf{y}_n^H\right)\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_m(z_2)\left(\mathbf{R}-\mathbf{y}_m\mathbf{y}_m^H\right)\mathbf{h}_2. \tag{2.47}$$

In Appendices C and D, the following asymptotic equivalents are stated, respectively,

$$\frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\hat{\omega}_n(z_1)\hat{\omega}_m(z_2)\,\mathrm{Tr}\left[\mathbf{\Theta}\mathbf{R}\left(\mathbf{Q}(z_1)-\mathbf{Q}_n(z_1)\right)\mathbf{R}\left(\mathbf{Q}(z_2)-\mathbf{Q}_m(z_2)\right)\mathbf{R}\right] \asymp 0 \tag{2.48}$$

and

$$\frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N}\hat{\omega}_n(z_1)\hat{\omega}_m(z_2)\,\mathrm{Tr}\left[\mathbf{\Theta}\left(\mathbf{R}-\mathbf{y}_n\mathbf{y}_n^H\right)\mathbf{Q}_n(z_1)\mathbf{R}\left(\mathbf{Q}(z_2)-\mathbf{Q}_m(z_2)\right)\mathbf{R}\right] \asymp 0, \tag{2.49}$$

where $\mathbf{\Theta} \in \mathbb{C}^{M\times M}$ is an arbitrary matrix with uniformly bounded spectral radius. With $\mathbf{\Theta} = \mathbf{h}_2\mathbf{h}_1^H$, the results in (2.48) and (2.49) can be used to show that both (2.44) and (2.45), and equivalently (2.46), respectively, vanish with probability one. In addition, a further result (cf. Appendix E)

$$\frac{1}{N^2}\sum_{m=1}^{N}\sum_{n=1}^{N}\hat{\omega}_n(z_1)\hat{\omega}_n(z_2)\,\mathrm{Tr}\left[\mathbf{\Theta}\left(\mathbf{R}-\mathbf{y}_m\mathbf{y}_m^H\right)\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_m(z_2)\left(\mathbf{R}-\mathbf{y}_m\mathbf{y}_m^H\right)\right]$$

$$\asymp cw(z_1)\,w(z_2)\,\mathrm{Tr}\left[\mathbf{\Theta}\mathbf{R}\right]\frac{1}{M}\mathrm{Tr}\left[\mathbf{R}\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\right], \tag{2.50}$$

where $\mathbf{\Theta} \in \mathbb{C}^{M \times M}$ is an arbitrary matrix with uniformly bounded spectral radius, can be applied with $\mathbf{\Theta} = \mathbf{h}_2 \mathbf{h}_1^H$ to write the following asymptotic equivalent

$$\eta(z_1, z_2) \asymp \mathbf{h}_1^H \mathbf{R} \mathbf{h}_2 + c w(z_1) w(z_2) \mathbf{h}_1^H \mathbf{R} \mathbf{h}_2 \frac{1}{M} \operatorname{Tr}[\mathbf{R}\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)]. \tag{2.51}$$

Now,

$$\begin{aligned}
&\frac{1}{M}\operatorname{Tr}[\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\mathbf{R}] = \\
&= \frac{1}{M}\operatorname{Tr}[(\mathbf{Q}(z_1) - \mathbf{P}(z_1))\mathbf{R}(\mathbf{Q}(z_2) - \mathbf{P}(z_2))\mathbf{R}] + \frac{1}{M}\operatorname{Tr}[\mathbf{P}(z_1)\mathbf{R}\mathbf{P}(z_2)\mathbf{R}] \\
&+ \frac{1}{M}\operatorname{Tr}[(\mathbf{Q}(z_1) - \mathbf{P}(z_1))\mathbf{R}\mathbf{P}(z_2)\mathbf{R}] + \frac{1}{M}\operatorname{Tr}[\mathbf{P}(z_1)\mathbf{R}(\mathbf{Q}(z_2) - \mathbf{P}(z_2))\mathbf{R}].
\end{aligned} \tag{2.52}$$

Both quantites in (2.52) are asymptotically equivalent to zero with probability one. To see this, note that the singular value decomposition of $\mathbf{R}\mathbf{P}(z_2)\mathbf{R}$ can be used to rewrite

$$\frac{1}{M}\operatorname{Tr}[(\mathbf{Q}(z_1) - \mathbf{P}(z_1))\mathbf{R}\mathbf{P}(z_2)\mathbf{R}] = \frac{1}{M}\sum_{k=1}^{M} \mathbf{v}_k^H (\mathbf{Q}(z_1) - \mathbf{P}(z_1)) \mathbf{s}_k, \tag{2.53}$$

where $\{\mathbf{s}_k\}$ and $\{\mathbf{v}_k\}$, $k = 1, \ldots, M$, are the sets of $M$ left and right singular vectors, including, with some abuse of notation, the magnitude of the singular values of the decomposed matrix. Note that $\{\mathbf{s}_k\}$ and $\{\mathbf{v}_k\}$ have by assumption uniformly bounded Euclidean norm for all $M$. The RHS of (2.53) can be shown to asymptotically vanish almost surely by Lemma A.2, since it can be bounded as (cf. Appendix F)

$$\max_{1 \le k \le N} \mathsf{E}\left[\left|\mathbf{v}_k^H (\mathbf{Q}(z_1) - \mathbf{P}(z_1)) \mathbf{u}_k\right|^p\right] \le \frac{C}{N^{1+\delta}}, \tag{2.54}$$

for some constants $C$, $\delta > 0$ and $p > 1$ not depending on $N$. In conclusion,

$$\frac{1}{M}\operatorname{Tr}[\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\mathbf{R}] \asymp \frac{1}{M}\operatorname{Tr}[(\mathbf{Q}(z_1) - \mathbf{P}(z_1))\mathbf{R}(\mathbf{Q}(z_2) - \mathbf{P}(z_2))\mathbf{R}] + \frac{1}{M}\operatorname{Tr}[\mathbf{P}(z_1)\mathbf{R}\mathbf{P}(z_2)\mathbf{R}].$$

Now, using the resolvent identity, the first term can be rewritten as

$$\begin{aligned}
u(z_1, z_2) &= \frac{1}{M}\operatorname{Tr}[(\mathbf{Q}(z_1) - \mathbf{P}(z_1))\mathbf{R}(\mathbf{Q}(z_2) - \mathbf{P}(z_2))\mathbf{R}] \\
&= \frac{1}{M}\operatorname{Tr}\left[\mathbf{P}(z_1)\left(w(z_1)\mathbf{R} - \hat{\mathbf{R}}\right)\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\left(w(z_2)\mathbf{R} - \hat{\mathbf{R}}\right)\mathbf{P}(z_2)\mathbf{R}\right] \\
&= \frac{1}{N}\frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{N}\frac{1}{M}\operatorname{Tr}\left[\mathbf{R}_1\left(w(z_1)\mathbf{R} - \mathbf{y}_n\mathbf{y}_n^H\right)\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\left(w(z_2)\mathbf{R} - \mathbf{y}_m\mathbf{y}_m^H\right)\mathbf{R}_2\right].
\end{aligned}$$

where we have defined $\mathbf{R}_1 = \mathbf{R}^{1/2}\mathbf{P}(z_1)$ and $\mathbf{R}_2 = \mathbf{P}(z_2)\mathbf{R}^{1/2}$. Similarly as in (2.35), we observe that

$$u(z_1, z_2) = \frac{1}{M} \operatorname{Tr}\left[\mathbf{R}_1 \left(w(z_1) - \frac{1}{N}\sum_{n=1}^{N} \hat{\omega}_n(z_1)\right) \mathbf{R}\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2) \left(w(z_2) - \frac{1}{N}\sum_{m=1}^{N} \hat{\omega}_n(z_2)\right) \mathbf{R}\mathbf{R}_2\right] \tag{2.55}$$

$$+ \frac{1}{N}\sum_{m=1}^{N} \frac{1}{M} \operatorname{Tr}\left[\mathbf{R}_1 \left(w(z_1) - \frac{1}{N}\sum_{n=1}^{N} \hat{\omega}_n(z_1)\right) \mathbf{R}\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2) \left(\hat{\omega}_m(z_2)\mathbf{R} - \mathbf{y}_m\mathbf{y}_m^H\right) \mathbf{R}_2\right] \tag{2.56}$$

$$+ \frac{1}{N}\sum_{n=1}^{N} \frac{1}{M} \operatorname{Tr}\left[\mathbf{R}_1 \left(\hat{\omega}_n(z_1)\mathbf{R} - \mathbf{y}_n\mathbf{y}_n^H\right) \mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2) \left(w(z_2) - \frac{1}{N}\sum_{m=1}^{N} \hat{\omega}_n(z_2)\right) \mathbf{R}\mathbf{R}_2\right] \tag{2.57}$$

$$+ \frac{1}{N}\frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{N} \frac{1}{M} \operatorname{Tr}\left[\mathbf{R}_1 \left(\hat{\omega}_n(z_1)\mathbf{R} - \mathbf{y}_n\mathbf{y}_n^H\right) \mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2) \left(\hat{\omega}_m(z_2)\mathbf{R} - \mathbf{y}_m\mathbf{y}_m^H\right) \mathbf{R}_2\right]. \tag{2.58}$$

The first term in the RHS of the previous expression is asymptotically zero almost surely. Indeed, the convergence of (2.55) can be stated again straightforwardly using (2.39) and the fact that the quantity $\frac{1}{M}\operatorname{Tr}\left[\mathbf{R}_1\mathbf{R}\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\mathbf{R}\mathbf{R}_2\right]$ is clearly bounded. The expression in (2.56) and, equivalently, (2.57) vanishes also asymptotically with probability one (see Appendix B with $\mathbf{\Theta} = \mathbf{R}_2\mathbf{R}_1$ for a proof).

Hence, the next asymptotic equivalent of $u(z_1, z_2)$ follows

$$\frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N} \frac{1}{M} \operatorname{Tr}\left[\mathbf{R}_1 \left(\hat{\omega}_n(z_1)\mathbf{R} - \mathbf{y}_n\mathbf{y}_n^H\right) \mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2) \left(\hat{\omega}_m(z_2)\mathbf{R} - \mathbf{y}_m\mathbf{y}_m^H\right) \mathbf{R}_2\right].$$

Using as before (2.43), we equivalently write $u(z_1, z_2)$ asymptotically as

$$\frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N} \hat{\omega}_n(z_1)\hat{\omega}_m(z_2) \frac{1}{M} \operatorname{Tr}\left[\mathbf{R}_1\mathbf{R}\left(\mathbf{Q}(z_1) - \mathbf{Q}_n(z_1)\right)\mathbf{R}\left(\mathbf{Q}(z_2) - \mathbf{Q}_m(z_2)\right)\mathbf{R}\mathbf{R}_2\right] \tag{2.59}$$

$$+ \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N} \hat{\omega}_n(z_1)\hat{\omega}_m(z_2) \frac{1}{M} \operatorname{Tr}\left[\mathbf{R}_1\left(\mathbf{R} - \mathbf{y}_n\mathbf{y}_n^H\right)\mathbf{Q}_n(z_1)\mathbf{R}\left(\mathbf{Q}(z_2) - \mathbf{Q}_m(z_2)\right)\mathbf{R}\mathbf{R}_2\right] \tag{2.60}$$

$$+ \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N} \hat{\omega}_n(z_1)\hat{\omega}_m(z_2) \frac{1}{M} \operatorname{Tr}\left[\mathbf{R}_1\mathbf{R}\left(\mathbf{Q}(z_1) - \mathbf{Q}_n(z_1)\right)\mathbf{R}\mathbf{Q}_m(z_2)\left(\mathbf{R} - \mathbf{y}_m\mathbf{y}_m^H\right)\mathbf{R}_2\right] \tag{2.61}$$

$$+ \frac{1}{N^2}\sum_{n=1}^{N}\sum_{m=1}^{N} \hat{\omega}_n(z_1)\hat{\omega}_m(z_2) \frac{1}{M} \operatorname{Tr}\left[\mathbf{R}_1\left(\mathbf{R} - \mathbf{y}_n\mathbf{y}_n^H\right)\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_m(z_2)\left(\mathbf{R} - \mathbf{y}_m\mathbf{y}_m^H\right)\mathbf{R}_2\right]. \tag{2.62}$$

The quantities (2.59) to (2.61) vanish with probability one. The proof follows from (2.41) and (2.48) using $\mathbf{\Theta} = \mathbf{R}_2\mathbf{R}_1$ (cf. Appendices C and D). Regarding (2.62), using (2.50) with

$\mathbf{\Theta} = \mathbf{R}_2\mathbf{R}_1$, it is straighforward to show that

$$u(z_1, z_2) \asymp cw(z_1)\, w(z_2) \frac{1}{M} \operatorname{Tr}[\mathbf{R}_2\mathbf{R}_1\mathbf{R}] \frac{1}{M} \operatorname{Tr}[\mathbf{R}\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)].$$

Hence,

$$\frac{1}{M} \operatorname{Tr}[\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\mathbf{R}] \asymp \frac{1}{M} \operatorname{Tr}[\mathbf{R}_2\mathbf{R}_1\mathbf{R}] + cw(z_1)w(z_2)\frac{1}{M} \operatorname{Tr}[\mathbf{R}_2\mathbf{R}_1\mathbf{R}] \frac{1}{M} \operatorname{Tr}[\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\mathbf{R}],$$

and, consequently,

$$\frac{1}{M} \operatorname{Tr}[\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\mathbf{R}] \asymp \frac{\frac{1}{M} \operatorname{Tr}[\mathbf{P}(z_2)\mathbf{R}\mathbf{P}(z_1)\mathbf{R}]}{1 - cw(z_1)w(z_2)\frac{1}{M} \operatorname{Tr}[\mathbf{P}(z_2)\mathbf{R}\mathbf{P}(z_1)\mathbf{R}]}.$$

Inserting the asymptotic equivalent of $\frac{1}{M} \operatorname{Tr}[\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\mathbf{R}]$ into (2.51), we conclude

$$\eta(z_1, z_2) \asymp \frac{\mathbf{h}^H\mathbf{P}(z_1)\mathbf{R}\mathbf{P}(z_2)\mathbf{h}}{1 - cw(z_1)w(z_2)\frac{1}{M} \operatorname{Tr}[\mathbf{P}(z_2)\mathbf{R}\mathbf{P}(z_1)\mathbf{R}]}. \tag{2.63}$$

Regarding the denominator, we may decompose the quantity $w(z_1)w(z_2)\frac{1}{M} \operatorname{Tr}[\mathbf{P}(z_2)\mathbf{R}\mathbf{P}(z_1)\mathbf{R}]$ in partial fractions as

$$\begin{aligned} w(z_1)w(z_2)\frac{1}{M}\sum_{m=1}^{M} \frac{\lambda_m^2}{(w(z_1)\lambda_m - z_1)(w(z_2)\lambda_m - z_2)} &= \\ = w(z_1)w(z_2)\frac{1}{M}\sum_{m=1}^{M} \frac{\lambda_m^2}{z_1 w(z_2) - z_2 w(z_1)} &\left[\frac{w(z_1)}{w(z_1)\lambda_m - z_1} - \frac{w(z_2)}{w(z_2)\lambda_m - z_2}\right] \\ &= \frac{1}{f(z_1) - f(z_2)}\frac{1}{M}\sum_{m=1}^{M}\left[\frac{\lambda_m^2}{\lambda_m - f(z_1)} - \frac{\lambda_m^2}{\lambda_m - f(z_2)}\right], \end{aligned}$$

with the aim of finally writing

$$\begin{aligned} &1 - cw(z_1)w(z_2)\frac{1}{M} \operatorname{Tr}[\mathbf{P}(z_2)\mathbf{R}\mathbf{P}(z_1)\mathbf{R}] \\ &= \frac{1}{f(z_1) - f(z_2)}\left[\left(f(z_1) - c\frac{1}{M}\sum_{m=1}^{M}\frac{\lambda_m^2}{\lambda_m - f(z_1)}\right) - \left(f(z_2) - c\frac{1}{M}\sum_{m=1}^{M}\frac{\lambda_m^2}{\lambda_m - f(z_2)}\right)\right] \\ &= \frac{1}{f(z_1) - f(z_2)}\left[\left(z_1 - \frac{c}{M}\sum_{m=1}^{M}\lambda_m\right) - \left(z_2 - \frac{c}{M}\sum_{m=1}^{M}\lambda_m\right)\right] \\ &= \frac{z_1 - z_2}{f(z_1) - f(z_2)}. \end{aligned} \tag{2.64}$$

Plugging (2.64) into the denominator of (2.63) yields the result we wanted to prove.

## Appendix 2.B Proof of (2.41)

Define

$$a_n^{(N)} = \frac{1}{N}\sum_{n=1}^{N} \operatorname{Tr}\left[\mathbf{\Theta}\left(\hat{\omega}_n(z_1)\mathbf{R} - \mathbf{y}_n\mathbf{y}_n^H\right)\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\mathbf{R}\right],$$

and

$$b_n^{(N)} = w(z_1) - \frac{1}{N}\sum_{n=1}^{N}\frac{1}{1+\frac{1}{N}\mathbf{y}_n^H\mathbf{Q}_n(z_1)\mathbf{y}_n}.$$

We give a proof based on the direct application of Lemma A.1 to the quantity $a_n^{(N)}b_n^{(N)}$. Since $b_n^{(N)} \asymp 0$ (cf. Equation (2.39)), it is enough to show that $a_n^{(N)}$ is stochastically bounded almost surely [Fel66], namely $a_n^{(N)} \asymp t$, where $t$ is a random variable bounded for all $M, N$.

An alternative proof based on martingale differences can be obtained according to Lemma A.2 by showing that

$$\max_{1\le n\le N} \mathsf{E}\left[\left|a_n^{(N)}b_n^{(N)}\right|^p\right] \le \frac{C}{N^{1+\delta}},$$

for some constants $C$, $\delta > 0$ and $p > 1$ not depending on $N$. To that effect, we just have to apply the Cauchy-Schwarz inequality and show, on the one hand,

$$\max_{1\le n\le N} \mathsf{E}\left[\left|b_n^{(N)}\right|^{2p}\right] \le \frac{C}{N^{1+\delta}}, \tag{2.65}$$

and, on the other, $\mathsf{E}\left[\left|a_n^{(N)}\right|^{2p}\right] < +\infty$. The inequality in (2.65) can be proved to hold using Burkholder's inequality. The boundness of the last expectation will result from the following derivations.

We next prove $a_n^{(N)} \asymp t$, with $t$ being a random variable bounded for all $M, N$. According to (2.43), we may write $a_n^{(N)}$ as

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\omega}_n(z_1)\operatorname{Tr}\left[\mathbf{\Theta}\mathbf{R}\left(\mathbf{Q}(z_1)-\mathbf{Q}_n(z_1)\right)\mathbf{R}\mathbf{Q}(z_2)\mathbf{R}\right] \tag{2.66}$$

$$+\frac{1}{N}\sum_{n=1}^{N}\hat{\omega}_n(z_1)\operatorname{Tr}\left[\mathbf{\Theta}\left(\mathbf{R}-\mathbf{y}_n\mathbf{y}_n^H\right)\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}(z_2)\mathbf{R}\right]. \tag{2.67}$$

We first prove the following statement regarding (2.66), namely

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\omega}_n(z_1)\operatorname{Tr}\left[\mathbf{\Theta}\mathbf{R}\left(\mathbf{Q}(z_1)-\mathbf{Q}_n(z_1)\right)\mathbf{R}\mathbf{Q}(z_2)\mathbf{R}\right] \asymp 0. \tag{2.68}$$

For that purpose, we note first that, using (2.42), the LHS of (2.68) can be written as

$$\sum_{n=1}^{N}\omega_n^2(z_1)\frac{1}{N}\operatorname{Tr}\left[\mathbf{\Theta}\mathbf{R}\mathbf{Q}_n(z_1)\mathbf{y}_n\mathbf{y}_n^H\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}(z_2)\mathbf{R}\right].$$

Since we know that

$$\left|\hat{\omega}_n(z)\right| = \left|\frac{1}{1+\frac{1}{N}\mathbf{y}_n^H\mathbf{Q}_n(z)\mathbf{y}_n}\right| < \frac{|z|}{\left|\operatorname{Im}\{z\}\right|} < +\infty, \tag{2.69}$$

in order to establish (2.68), it is enough from Lemma A.2 to show that

$$\max_{1\le n\le N}\frac{1}{N^p}\mathsf{E}\left[\left|\mathbf{y}_n^H\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}(z_2)\mathbf{R}\mathbf{\Theta}\mathbf{R}\mathbf{Q}_n(z_1)\mathbf{y}_n\right|^p\right] \le \frac{C}{N^{1+\delta}},$$

for some constants $C$, $\delta > 0$ and $p > 1$ not depending on $N$. To that effect, we prove that the expectation is bounded for $p > 1$. First, applying (2.42) and the Jensen inequality, we write the expectation as

$$\mathsf{E}\left[\left|\mathbf{y}_n^H\mathbf{Q}_n(z_1)\mathbf{R}\left(\mathbf{Q}_n(z_2)-\frac{1}{N}\frac{\mathbf{Q}_n(z_2)\mathbf{y}_n\mathbf{y}_n^H\mathbf{Q}_n(z_2)}{1+\frac{1}{N}\mathbf{y}_n^H\mathbf{Q}_n(z_2)\mathbf{y}_n}\right)\mathbf{R\Theta R}\mathbf{Q}_n(z_1)\mathbf{y}_n\right|^p\right]$$
$$\leq 2^{p-1}\left\{\mathsf{E}\left[\left|\mathbf{y}_n^H\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_n(z_2)\mathbf{R\Theta R}\mathbf{Q}_n(z_1)\mathbf{y}_n\right|^p\right]\right.$$
$$\left.+\mathsf{E}\left[\left|\hat{\omega}_n\left(z_2\right)\frac{1}{N}\mathbf{y}_n^H\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_n(z_2)\mathbf{y}_n\mathbf{y}_n^H\mathbf{Q}_n(z_2)\mathbf{R\Theta R}\mathbf{Q}_n(z_1)\mathbf{y}_n\right|^p\right]\right\}.$$

The first term can be readily shown to be bounded by using Lemma A.4. As for the second one, considering (2.69) and using the Cauchy-Schwarz inequality, we may write

$$\mathsf{E}\left[\left|\hat{\omega}_n\left(z_2\right)\frac{1}{N}\mathbf{y}_n^H\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_n(z_2)\mathbf{y}_n\mathbf{y}_n^H\mathbf{Q}_n(z_2)\mathbf{R\Theta R}\mathbf{Q}_n(z_1)\mathbf{y}_n\right|^p\right]$$
$$\leq\left(\frac{|z_2|}{\mathrm{Im}\left\{z_2\right\}}\right)^p\mathsf{E}^{1/2}\left[\left|\mathbf{y}_n^H\mathbf{Q}_n(z_2)\mathbf{R\Theta R}\mathbf{Q}_n(z_1)\mathbf{y}_n\right|^{2p}\right]\mathsf{E}^{1/2}\left[\left|\frac{1}{N}\mathbf{y}_n^H\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_n(z_2)\mathbf{y}_n\right|^{2p}\right].$$

Now, both expectations can be seen to be bounded by directly applying Lemma A.4.

We next proceed to prove the convergence of (2.67) to a bounded random quantity. First, using the matrix inversion lemma as in (2.42), we rewrite the LHS of (2.67) as

$$\frac{1}{N}\sum_{n=1}^{N}\mathrm{Tr}\left[\mathbf{\Theta}\left(\mathbf{R}-\mathbf{y}_n\mathbf{y}_n^H\right)\mathbf{Q}_n(z_1)\mathbf{R}\left(\mathbf{Q}_n\left(z_2\right)-\frac{1}{N}\frac{\mathbf{Q}_n\left(z_2\right)\mathbf{y}_n\mathbf{y}_n^H\mathbf{Q}_n\left(z_2\right)}{1+\frac{1}{N}\mathbf{y}_n^H\mathbf{Q}_n(z_2)\mathbf{y}_n}\right)\mathbf{R}\right]. \tag{2.70}$$

Further, it follows from Lemma A.5 that, asymptotically, (2.70) can be equivalently written as

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\omega}_n\left(z_2\right)\frac{1}{N}\mathrm{Tr}\left[\mathbf{\Theta}\left(\mathbf{y}_n\mathbf{y}_n^H-\mathbf{R}\right)\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_n\left(z_2\right)\mathbf{y}_n\mathbf{y}_n^H\mathbf{Q}_n\left(z_2\right)\mathbf{R}\right]. \tag{2.71}$$

On the one hand, considering (2.69) and using Lemma A.2 together with Lemma A.4, we find that

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\omega}_n\left(z_2\right)\frac{1}{N}\mathrm{Tr}\left[\mathbf{\Theta R}\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_n\left(z_2\right)\mathbf{y}_n\mathbf{y}_n^H\mathbf{Q}_n\left(z_2\right)\mathbf{R}\right]\asymp 0.$$

Thus, in the asymptotic regime considered here, (2.71) is equivalent to

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\omega}_n\left(z_2\right)\frac{1}{N}\mathbf{y}_n^H\mathbf{Q}_n\left(z_2\right)\mathbf{R\Theta}\mathbf{y}_n\mathbf{y}_n^H\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_n\left(z_2\right)\mathbf{y}_n. \tag{2.72}$$

On the other hand, using again the fact that $\hat{\omega}_n\left(z_2\right)$ is absolutely bounded and (see next)

$$\frac{1}{N}\sum_{n=1}^{N}\xi_n\mathbf{y}_n^H\mathbf{Q}_n\left(z_2\right)\mathbf{R\Theta}\mathbf{y}_n\asymp 0, \tag{2.73}$$

where we have defined

$$\xi_n = \frac{1}{N}\mathbf{y}_n^H \mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_n(z_2)\mathbf{y}_n - \frac{1}{N}\operatorname{Tr}\left[\mathbf{R}\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_n(z_2)\right],$$

we obtain the following asymptotic equivalent of (2.72), namely

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\omega}_n(z_2)\,\mathbf{y}_n^H\mathbf{Q}_n(z_2)\,\mathbf{R}\boldsymbol{\Theta}\mathbf{y}_n\frac{1}{N}\operatorname{Tr}\left[\mathbf{R}\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_n(z_2)\right]. \tag{2.74}$$

To see (2.73), we just need to show that

$$\max_{1\le n\le N}\mathsf{E}\left[\left|\xi_n\mathbf{y}_n^H\mathbf{Q}_n(z_2)\,\mathbf{R}\boldsymbol{\Theta}\mathbf{y}_n\right|^p\right] \le \frac{C}{N^{1+\delta}},$$

for some constants $C$, $\delta > 0$ and $p > 1$ not depending on $N$. Using the Cauchy-Schwarz inequality, we find that

$$\mathsf{E}\left[\left|\xi_n\mathbf{y}_n^H\mathbf{Q}_n(z_2)\,\mathbf{R}\boldsymbol{\Theta}\mathbf{y}_n\right|^p\right] \le \mathsf{E}^{1/2}\left[\left|\mathbf{y}_n^H\mathbf{Q}_n(z_2)\,\mathbf{R}\boldsymbol{\Theta}\mathbf{y}_n\right|^{2p}\right]\mathsf{E}^{1/2}\left[\left|\xi_n\right|^{2p}\right].$$

Now, Lemma A.4 can be used to show the first expectation to be bounded. Regarding the second expectation, we can apply Lemma A.3 in order to obtain

$$\mathsf{E}\left[\left|\xi_n\right|^{2p}\right] \le K_{2p}\left\|\mathbf{C}\right\|^{2p}\mathsf{E}\left[\left|\xi\right|^{4p}\right]\left\{\frac{1}{N^p}+\frac{1}{N^{2p-1}}\right\},$$

where we have used $\left\|\mathbf{C}^p\right\|_W^2 \le \left\|\mathbf{C}^p\right\|^2 = \left\|\mathbf{C}\right\|^{2p}$ and, by Jensen's inequality, $\mathsf{E}^p\left[\left|\xi\right|^4\right] \le \mathsf{E}\left[\left|\xi\right|^{4p}\right]$ for any $p > 1$. Accordingly, we just need to choose $p \ge 2$ to prove the result in (2.73), since, by assumption,

$$\left\|\mathbf{C}\right\| \le \frac{1}{\left|\operatorname{Im}\{z_1\}\right|}\frac{1}{\left|\operatorname{Im}\{z_2\}\right|}\left\|\mathbf{R}\right\|^2 < +\infty.$$

Finally, we may use the fact that

$$\operatorname{Tr}\left[\boldsymbol{\Theta}\mathbf{R}\mathbf{Q}_n(z_2)\,\mathbf{R}\right] \asymp \operatorname{Tr}\left[\boldsymbol{\Theta}\mathbf{R}\mathbf{Q}(z_2)\,\mathbf{R}\right],$$

and

$$\frac{1}{N}\operatorname{Tr}\left[\mathbf{R}\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_n(z_2)\right] \asymp \frac{1}{N}\operatorname{Tr}\left[\mathbf{R}\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\right], \tag{2.75}$$

where all four terms are bounded for all $M, N$, together with Lemma A.5 in order to express (2.74) asymptotically as

$$\begin{aligned}\operatorname{Tr}\left[\boldsymbol{\Theta}\mathbf{R}\mathbf{Q}(z_2)\,\mathbf{R}\right]\frac{1}{N}\operatorname{Tr}\left[\mathbf{R}\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\right]\frac{1}{N}\sum_{n=1}^{N}\hat{\omega}_n(z_2)&\\ \le \frac{|z_2|}{\operatorname{Im}\{z_2\}}\operatorname{Tr}\left[\boldsymbol{\Theta}\mathbf{R}\mathbf{Q}(z_2)\,\mathbf{R}\right]\frac{1}{N}\operatorname{Tr}\left[\mathbf{R}\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\right] &< +\infty.\end{aligned}$$

# Appendix 2.C Proof of (2.48)

By Lemma A.2, as well as the fact that $\omega_n(z)$ is absolutely bounded for all $M, N$, it suffices to prove that

$$\max_{1\leq m,n\leq N} \mathsf{E}\left[\left|\frac{1}{N^2}\mathbf{y}_m^H\mathbf{Q}_m(z_2)\,\mathbf{R\Theta R Q}_n(z_1)\mathbf{y}_n\mathbf{y}_n^H\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_m(z_2)\mathbf{y}_m\right|^p\right] \leq \frac{K}{N^{1+\delta}}, \tag{2.76}$$

for some constants $K$, $\delta > 0$ and $p > 1$ not depending on $N$. Using the Cauchy-Schwarz inequality in the linear space of $M \times M$ complex square matrices equipped with the weak norm, the expectation in the LHS of (2.76) can be written as

$$\mathsf{E}\left[\left|\frac{1}{N}\operatorname{Tr}\left[\mathbf{A}_n\mathbf{B}_n\right]\right|^p\right] \leq \mathsf{E}^{1/2}\left[\left(\frac{1}{N}\operatorname{Tr}\left[\mathbf{A}_n\mathbf{A}_n^H\right]\right)^p\right]\mathsf{E}^{1/2}\left[\left(\frac{1}{N}\operatorname{Tr}\left[\mathbf{B}_n\mathbf{B}_n^H\right]\right)^p\right], \tag{2.77}$$

where we have defined $\mathbf{A}_n = \mathbf{R\Theta R Q}_n(z_1)\mathbf{y}_n\mathbf{y}_n^H\mathbf{Q}_n(z_1)\mathbf{R}$ and $\mathbf{B}_n = \frac{1}{N}\mathbf{Q}_m(z_2)\mathbf{y}_m\mathbf{y}_m^H\mathbf{Q}_m(z_2)$. Regarding the first factor in (2.77), the Cauchy-Schwarz inequality can be used to write

$$\begin{aligned}\mathsf{E}&\left[\left(\mathbf{y}_n^H\mathbf{Q}_n^H(z_1)\mathbf{R\Theta R}^2\mathbf{\Theta R Q}_n(z_1)\mathbf{y}_n\right)^p\left(\frac{1}{N}\mathbf{y}_n^H\mathbf{Q}_n(z_1)\mathbf{R}^2\mathbf{Q}_n^H(z_1)\mathbf{y}_n\right)^p\right]\\ &\leq \mathsf{E}^{1/2}\left[\left|\mathbf{y}_n^H\mathbf{Q}_n^H(z_1)\mathbf{R\Theta R}^2\mathbf{\Theta R Q}_n(z_1)\mathbf{y}_n\right|^{2p}\right]\mathsf{E}^{1/2}\left[\left|\frac{1}{N}\mathbf{y}_n^H\mathbf{Q}_n(z_1)\mathbf{R}^2\mathbf{Q}_n^H(z_1)\mathbf{y}_n\right|^{2p}\right],\end{aligned}$$

that can be readily found to be bounded using Lemma A.4. Hence, according to the second factor in (2.77), it remains to show that

$$\max_{1\leq m\leq N}\frac{1}{N^{p/2}}\mathsf{E}\left[\left(\frac{1}{N}\mathbf{y}_m^H\mathbf{Q}_m(z_2)\,\mathbf{Q}_m^H(z_2)\mathbf{y}_m\right)^{2p}\right] \leq \frac{K'}{N^{1+\delta}},$$

but this is straightforward using Lemma A.4 for $p > 2$.

# Appendix 2.D Proof of (2.49)

According to Lemma A.2, using the matrix inversion lemma as in (2.42), the absolutely boundness of $\omega_n(z)$ (cf. Equation (2.69)) and the Jensen inequality, the proof of (2.49) reduces to show that

$$\max_{1\leq m,n\leq N}\mathsf{E}\left[\left|\frac{1}{N}\mathbf{y}_m^H\mathbf{Q}_m(z_2)\mathbf{R\Theta}\left(\mathbf{R}-\mathbf{y}_n\mathbf{y}_n^H\right)\mathbf{Q}_{mn}(z_1)\mathbf{y}_m\frac{1}{N}\mathbf{y}_m^H\mathbf{Q}_{mn}(z_1)\mathbf{R}\mathbf{Q}_m(z_2)\mathbf{y}_m\right|^p\right] \leq \frac{K}{N^{1+\delta}}, \tag{2.78}$$

for some constants $K$, $\delta > 0$ and $p > 1$ not depending on $N$. Using the Cauchy-Schwarz inequality, the expectation in (2.78) can be bounded above by

$$\mathsf{E}^{1/2}\left[\left|\frac{1}{N}\mathbf{y}_m^H\mathbf{Q}_m(z_2)\mathbf{R\Theta}\left(\mathbf{R}-\mathbf{y}_n\mathbf{y}_n^H\right)\mathbf{Q}_{mn}(z_1)\mathbf{y}_m\right|^{2p}\right]\mathsf{E}^{1/2}\left[\left|\frac{1}{N}\mathbf{y}_m^H\mathbf{Q}_{mn}(z_1)\mathbf{R}\mathbf{Q}_m(z_2)\mathbf{y}_m\right|^{2p}\right]. \tag{2.79}$$

The second factor of the RHS of (2.79) is clearly bounded by Lemma A.4. Regarding the first one, using Jensen's inequality we find that

$$\mathsf{E}^{1/2}\left[\left|\frac{1}{N}\mathbf{y}_m^H\mathbf{Q}_m(z_2)\mathbf{R}\boldsymbol{\Theta}\mathbf{R}\mathbf{Q}_{mn}(z_1)\mathbf{y}_m\right|^{2p}\right] \leq \frac{1}{N^p}K', \tag{2.80}$$

and, further,

$$\mathsf{E}^{1/2}\left[\left|\frac{1}{N}\mathbf{y}_m^H\mathbf{Q}_m(z_2)\mathbf{R}\boldsymbol{\Theta}\mathbf{y}_n\mathbf{y}_n^H\mathbf{Q}_{mn}(z_1)\mathbf{y}_m\right|^{2p}\right] \leq \frac{1}{N^p}K'' + \frac{1}{N^{2p}}K'''. \tag{2.81}$$

The inequality in (2.80) is straightforward, for the expectation $\mathsf{E}\left[\left|\mathbf{y}_m^H\mathbf{Q}_m(z_2)\mathbf{R}\mathbf{h}_2\mathbf{h}_1^H\mathbf{R}\mathbf{Q}_{mn}(z_1)\mathbf{y}_m\right|^{2p}\right]$ is bounded by Lemma A.4. To see (2.81), we apply the matrix inversion lemma on $\mathbf{Q}_m(z_2)$ and, using again the Jensen inequality, the bound in (2.69) and the Cauchy-Schwarz inequality, it is straigthforward to find that

$$\begin{aligned}
&\mathsf{E}\left[\left|\frac{1}{N}\mathbf{y}_m^H\mathbf{Q}_m(z_2)\mathbf{R}\boldsymbol{\Theta}\mathbf{y}_n\mathbf{y}_n^H\mathbf{Q}_{mn}(z_1)\mathbf{y}_m\right|^{2p}\right] \leq \\
&\qquad \leq K\left\{\frac{1}{N^{2p}}\mathsf{E}\left[\left|\mathbf{y}_m^H\mathbf{Q}_m(z_2)\mathbf{R}\boldsymbol{\Theta}\mathbf{y}_n\mathbf{y}_n^H\mathbf{Q}_{mn}(z_1)\mathbf{y}_m\right|^{2p}\right]\right. \\
&\qquad \left. +\frac{1}{N^{4p}}\mathsf{E}^{1/2}\left[\left|\mathbf{y}_n^H\mathbf{Q}_{mn}(z_2)\mathbf{R}\boldsymbol{\Theta}\mathbf{y}_n\right|^{4p}\right]\mathsf{E}^{1/2}\left[\left|\mathbf{y}_n^H\mathbf{Q}_{mn}(z_1)\mathbf{y}_m\mathbf{y}_m^H\mathbf{Q}_{mn}(z_2)\mathbf{y}_n\right|^{4p}\right]\right\},
\end{aligned}$$

whereby we readily identify, from Lemmas A.4 and A.5, that all three expectations are bounded. Hence, it remains to choose $p > 1$ to finally prove the result.

## Appendix 2.E Proof of (2.50)

In the analysis of the convergence of (2.50), two different cases regarding the double index can be identifed:

- Case $(m = n)$. In this case we have

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\omega}_n(z_1)\,\hat{\omega}_n(z_2)\,\frac{1}{N}\operatorname{Tr}\left[\boldsymbol{\Theta}\left(\mathbf{R}-\mathbf{y}_n\mathbf{y}_n^H\right)\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_n(z_2)\left(\mathbf{R}-\mathbf{y}_n\mathbf{y}_n^H\right)\right]. \tag{2.82}$$

  We expand the argument of the trace and observe that all resulting terms but the following asymptotic equivalent of (2.82) can be be neglected in the limiting regime, namely

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\omega}_n(z_1)\,\hat{\omega}_n(z_2)\,\frac{1}{N}\mathbf{y}_n^H\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_n(z_2)\mathbf{y}_n\mathbf{y}_n^H\boldsymbol{\Theta}\mathbf{y}_n. \tag{2.83}$$

  This can be checked by directly applying Lemma A.2 together with Lemmas A.4 and A.5 upond considering the bound in (2.69). On the other hand, using the previously proved fact that (cf. Equation (2.73))

$$\frac{1}{N}\sum_{n=1}^{N}\xi_n \mathbf{y}_n^H \mathbf{\Theta} \mathbf{y}_n \asymp 0,$$

along with (2.69), we can regard (2.83) in the asymptotic regime as

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\omega}_n\left(z_1\right)\hat{\omega}_n\left(z_2\right)\mathbf{y}_n^H \mathbf{\Theta} \mathbf{y}_n \frac{1}{N}\operatorname{Tr}\left[\mathbf{R}\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_n(z_2)\right]. \tag{2.84}$$

On the other hand, using (2.75) as well as

$$\frac{1}{N}\sum_{n=1}^{N}\mathbf{y}_n^H \mathbf{\Theta} \mathbf{y}_n \left\{\hat{\omega}_n\left(z_1\right)\hat{\omega}_n\left(z_2\right) - w\left(z_1\right)w\left(z_2\right)\right\} \asymp 0, \tag{2.85}$$

we can equivalently write (2.84) asymptotically as

$$w\left(z_1\right)w\left(z_2\right)\frac{1}{N}\operatorname{Tr}\left[\mathbf{R}\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\right]\frac{1}{N}\sum_{n=1}^{N}\mathbf{y}_n^H \mathbf{\Theta} \mathbf{y}_n. \tag{2.86}$$

To see (2.85), we define first $\tau_n = \hat{\omega}_n\left(z_1\right)\hat{\omega}_n\left(z_2\right) - w\left(z_1\right)w\left(z_2\right)$ and write the LHS of (2.85) as

$$\frac{1}{N}\sum_{n=1}^{N}\tau_n\left(\mathbf{y}_n^H \mathbf{\Theta} \mathbf{y}_n - \operatorname{Tr}\left[\mathbf{\Theta}\mathbf{R}\right]\right) + \operatorname{Tr}\left[\mathbf{\Theta}\mathbf{R}\right]\frac{1}{N}\sum_{n=1}^{N}\tau_n, \tag{2.87}$$

where, clearly, both $\tau_n$ and $\operatorname{Tr}\left[\mathbf{\Theta}\mathbf{R}\right]$ are bounded by assumption. The convergence of the first summation in (2.87) can be proved using Burkholder's inequality (cf. Lemma A.5). As for the second term, we write the sum as e.g.

$$\frac{1}{N}\sum_{n=1}^{N}\hat{\omega}_n\left(z_1\right)\left(\hat{\omega}_n\left(z_2\right) - w\left(z_2\right)\right) + \frac{1}{N}\sum_{n=1}^{N}\left(\hat{\omega}_n\left(z_1\right) - w\left(z_1\right)\right)w\left(z_2\right),$$

and eventually realize that both terms vanish asymptotically from (2.39). Furthermore, direct application of the SLLN allows us to finally write (2.86) as

$$w\left(z_1\right)w\left(z_2\right)\operatorname{Tr}\left[\mathbf{\Theta}\mathbf{R}\right]\frac{1}{N}\operatorname{Tr}\left[\mathbf{R}\mathbf{Q}(z_1)\mathbf{R}\mathbf{Q}(z_2)\right],$$

which is already the result we wanted to establish.

- Case $(m \neq n)$. From (2.82), we have to show for this remaining case that the summation converges to zero almost surely, i.e.

$$\frac{1}{N^2}\sum_{\substack{n,m=1\\ n\neq m}}^{N}\hat{\omega}_n\left(z_1\right)\hat{\omega}_n\left(z_2\right)\operatorname{Tr}\left[\mathbf{\Theta}\left(\mathbf{R} - \mathbf{y}_n\mathbf{y}_n^H\right)\mathbf{Q}_n(z_1)\mathbf{R}\mathbf{Q}_m(z_2)\left(\mathbf{R} - \mathbf{y}_m\mathbf{y}_m^H\right)\right] \asymp 0. \tag{2.88}$$

Expanding the expression inside the summation upon applying the matrix inversion lemma on $\mathbf{Q}_n(z_1)$ and $\mathbf{Q}_m(z_2)$ and considering (2.69), we can use again Lemmas A.2, A.4 and

A.5 in order to (asymptotically) neglect vanishing terms and finally replace (2.88) with the following asymptotic equivalent

$$\frac{1}{N}\sum_{n=1}^{N}\eta_n^{(1)}+\frac{1}{N}\sum_{n=1}^{N}\eta_n^{(2)}+\frac{1}{N}\sum_{n=1}^{N}\eta_n^{(3)}+\frac{1}{N}\sum_{m=1}^{N}\eta_m^{(4)}, \tag{2.89}$$

where we have defined

$$\eta_n^{(1)} = \frac{1}{N}\sum_{\substack{m=1\\m\neq n}}^{N}\operatorname{Tr}\left[\boldsymbol{\Theta}\left(\mathbf{R}-\mathbf{y}_n\mathbf{y}_n^H\right)\mathbf{Q}_{nm}(z_1)\mathbf{R}\mathbf{Q}_{nm}(z_2)\mathbf{R}\right], \tag{2.90}$$

$$\eta_n^{(2)} = \frac{1}{N}\sum_{\substack{m=1\\m\neq n}}^{N}\operatorname{Tr}\left[\boldsymbol{\Theta}\left(\mathbf{y}_n\mathbf{y}_n^H-\mathbf{R}\right)\mathbf{Q}_{nm}(z_1)\mathbf{R}\mathbf{Q}_{nm}(z_2)\mathbf{y}_m\mathbf{y}_m^H\right], \tag{2.91}$$

$$\eta_n^{(3)} = \frac{1}{N}\sum_{\substack{m=1\\m\neq n}}^{N}\operatorname{Tr}\left[\boldsymbol{\Theta}\left(\mathbf{R}-\mathbf{y}_n\mathbf{y}_n^H\right)\mathbf{Q}_{nm}(z_2)\mathbf{y}_m\mathbf{y}_m^H\right]\frac{1}{N}\mathbf{y}_m^H\mathbf{Q}_{nm}(z_2)\mathbf{R}\mathbf{Q}_{nm}(z_2)\mathbf{y}_m, \tag{2.92}$$

$$\eta_m^{(4)} = \frac{1}{N}\sum_{\substack{n=1\\n\neq m}}^{N}\operatorname{Tr}\left[\boldsymbol{\Theta}\mathbf{y}_n\mathbf{y}_n^H\mathbf{Q}_{nm}(z_1)\left(\mathbf{R}-\mathbf{y}_m\mathbf{y}_m^H\right)\right]\frac{1}{N}\mathbf{y}_n^H\mathbf{Q}_{nm}(z_1)\mathbf{R}\mathbf{Q}_{nm}(z_1)\mathbf{y}_n. \tag{2.93}$$

Now, observe that the sequences (2.90) to (2.93) satisfy the martingale difference sequence condition, namely

$$\mathsf{E}\left[\left|\eta_k^{(l)}\right|\right]<+\infty,\ \mathsf{E}\left[\eta_k^{(l)}\middle|\mathcal{F}_{k-1}^{(l)}\right]=0, \tag{2.94}$$

$l=1,\ldots,M$, where $\left\{\mathcal{F}_k^{(l)}\right\}$ is the increasing sequence of $\sigma$-fields generated by $\left\{\eta_k^{(l)}\right\}$, $k=1,\ldots,M$. The condition in (2.94) can be easily verified using the Cauchy-Schwarz and Jensen inequalities together with Lemmas A.4 and A.5. Consequently, we are allowed to use the Burkholder inequality as applied throughout the proof of Theorem 2.2 in order to show the almost sure convergence to zero of the normalized sums in (2.89), or equivalently (2.88).

## Appendix 2.F Proof of (2.54)

The proof follows essentially the same arguments as in the proof of [Bai07, Theorem 1]. From Section 1.1 and the dominated convergence theorem, it follows that

$$\mathsf{E}\left[\mathbf{v}_k^H\mathbf{Q}(z_1)\mathbf{u}_k\right]\to\mathbf{v}_k^H\mathbf{P}(z_1)\mathbf{u}_k.$$

Using the previous fact, we just need to equivalently prove that

$$\mathsf{E}\left[\left|\mathbf{v}_k^H\mathbf{Q}(z_1)\mathbf{u}_k-\mathsf{E}\left[\mathbf{v}_k^H\mathbf{Q}(z_1)\mathbf{u}_k\right]\right|^p\right]\leq\frac{C}{N^{1+\delta}}.$$

Now, let $\mathcal{F}_j$ be defined as the $\sigma$-field generated by the random sequence $\{\mathbf{u}_1, \ldots, \mathbf{u}_j\}$ and $\mathsf{E}_j[\cdot]$ the conditional expectation given the $\sigma$-field $\mathcal{F}_j$, namely $\mathsf{E}_j[\cdot] = \mathsf{E}[\cdot\,|\mathcal{F}_j]$. Then, considering the two extreme cases, namely

$$\begin{aligned}\mathsf{E}_N\left[\mathbf{v}_k^H\mathbf{Q}(z_1)\mathbf{u}_k\right] &= \mathbf{v}_k^H\mathbf{Q}(z_1)\mathbf{u}_k,\\ \mathsf{E}_0\left[\mathbf{v}_k^H\mathbf{Q}(z_1)\mathbf{u}_k\right] &= \mathsf{E}\left[\mathbf{v}_k^H\mathbf{Q}(z_1)\mathbf{u}_k\right],\end{aligned}$$

regarding both trivial $\sigma$-fields, namely the one generated by all random elements and the empty set, respectively, as well as the following identity

$$\mathsf{E}_j\left[\mathbf{v}_k^H\mathbf{Q}_j(z_1)\mathbf{u}_k\right] - \mathsf{E}_{j-1}\left[\mathbf{v}_k^H\mathbf{Q}_j(z_1)\mathbf{u}_k\right] = 0,$$

we can write

$$\begin{aligned}\mathbf{v}_k^H\mathbf{Q}(z_1)\mathbf{u}_k - \mathsf{E}\left[\mathbf{v}_k^H\mathbf{P}(z_1)\mathbf{u}_k\right] &= \mathsf{E}_N\left[\mathbf{v}_k^H\mathbf{Q}(z_1)\mathbf{u}_k\right] - \mathsf{E}_0\left[\mathbf{v}_k^H\mathbf{Q}(z_1)\mathbf{u}_k\right]\\ &= \sum_{j=1}^{N}\mathsf{E}_j\left[\mathbf{v}_k^H\mathbf{Q}(z_1)\mathbf{u}_k\right] - \mathsf{E}_{j-1}\left[\mathbf{v}_k^H\mathbf{Q}(z_1)\mathbf{u}_k\right]\\ &= \sum_{j=1}^{N}\mathsf{E}_j\left[\mathbf{v}_k^H\left(\mathbf{Q}(z_1)-\mathbf{Q}_j(z_1)\right)\mathbf{u}_k\right] - \mathsf{E}_{j-1}\left[\mathbf{v}_k^H\left(\mathbf{Q}(z_1)-\mathbf{Q}_j(z_1)\right)\mathbf{u}_k\right]\\ &= -\sum_{j=1}^{N}\left\{\mathsf{E}_j - \mathsf{E}_{j-1}\right\}\left[\omega_j(z_1)\frac{1}{N}\mathbf{y}_j^H\mathbf{Q}_j(z_1)\mathbf{u}_k\mathbf{v}_k^H\mathbf{Q}_j(z_1)\mathbf{y}_j\right]\\ &= -\sum_{j=1}^{N}\mathsf{E}_j\left[\beta_j\omega_j(z_1)\right] - \left\{\mathsf{E}_j - \mathsf{E}_{j-1}\right\}\left[\gamma_j\omega_j(z_1)\frac{\frac{1}{N}\mathbf{y}_j^H\mathbf{Q}_j(z_1)\mathbf{u}_k\mathbf{v}_k^H\mathbf{Q}_j(z_1)\mathbf{y}_j}{1+\frac{1}{N}\operatorname{Tr}\left[\mathbf{R}\mathbf{Q}_j(z_1)\right]}\right],\end{aligned}$$

where we have defined $\beta_j = \frac{1}{N}\mathbf{y}_j^H\mathbf{Q}_j(z_1)\mathbf{u}_k\mathbf{v}_k^H\mathbf{Q}_j(z_1)\mathbf{y}_j - \frac{1}{N}\mathbf{v}_k^H\mathbf{Q}_j(z_1)\mathbf{R}\mathbf{Q}_j(z_1)\mathbf{u}_k$ and $\gamma_j = \frac{1}{N}\mathbf{y}_j^H\mathbf{Q}_j(z_1)\mathbf{y}_j + \frac{1}{N}\operatorname{Tr}\left[\mathbf{R}\mathbf{Q}_j(z_1)\right]$ Note further that the random series coefficients form a martingale difference sequence, since, clearly

$$\mathsf{E}_{j-1}\left[\mathsf{E}_j\left[\mathbf{v}_k^H\left(\mathbf{Q}(z_1)-\mathbf{Q}_j(z_1)\right)\mathbf{u}_k\right]\right] = \mathsf{E}_{j-1}\left[\mathbf{v}_k^H\left(\mathbf{Q}(z_1)-\mathbf{Q}_j(z_1)\right)\mathbf{u}_k\right].$$

Hence, we can use the Burkholder inequality to write

$$
\begin{aligned}
&\mathsf{E}\left[\left|\mathbf{v}_k^H\mathbf{Q}(z_1)\mathbf{u}_k-\mathsf{E}\left[\mathbf{v}_k^H\mathbf{P}(z_1)\mathbf{u}_k\right]\right|^p\right]\leq\\
&\quad\leq K_p\left\{\mathsf{E}\left[\left[\sum_{j=1}^N\mathsf{E}_{j-1}\left[\left|\{\mathsf{E}_j-\mathsf{E}_{j-1}\}\left[\omega_j(z_1)\frac{1}{N}\mathbf{y}_j^H\mathbf{Q}_j(z_1)\mathbf{u}_k\mathbf{v}_k^H\mathbf{Q}_j(z_1)\mathbf{y}_j\right]\right|^2\right]\right]^{p/2}\right]\right.\\
&\qquad\left.+\mathsf{E}\left[\sum_{j=1}^N\left|\{\mathsf{E}_j-\mathsf{E}_{j-1}\}\left[\omega_j(z_1)\frac{1}{N}\mathbf{y}_j^H\mathbf{Q}_j(z_1)\mathbf{u}_k\mathbf{v}_k^H\mathbf{Q}_j(z_1)\mathbf{y}_j\right]\right|^p\right]\right\}\\
&\leq\mathsf{E}\left[\left(\sum_{j=1}^N\left\{K_p'\mathsf{E}_{j-1}\left[\left|\mathsf{E}_j\left[\beta_j\right]\right|^2\right]++K_p''\mathsf{E}_{j-1}\left[\left|\{\mathsf{E}_j-\mathsf{E}_{j-1}\}\left[\gamma_j\frac{1}{N}\mathbf{y}_j^H\mathbf{Q}_j(z_1)\mathbf{u}_k\mathbf{v}_k^H\mathbf{Q}_j(z_1)\mathbf{y}_j\right]\right|^2\right]\right\}\right)^{p/2}\right]\\
&\qquad+K_p'''\sum_{j=1}^N\mathsf{E}\left[\left|\frac{1}{N}\mathbf{y}_j^H\mathbf{Q}_j(z_1)\mathbf{u}_k\mathbf{v}_k^H\mathbf{Q}_j(z_1)\mathbf{y}_j\right|^p\right]\\
&\quad\leq\mathsf{E}\left[\left(\sum_{j=1}^N\left\{K_p'\mathsf{E}_{j-1}\left[\left|\beta_j\right|^2\right]++K_p''\mathsf{E}_{j-1}\left[\left|\gamma_j\frac{1}{N}\mathbf{y}_j^H\mathbf{Q}_j(z_1)\mathbf{u}_k\mathbf{v}_k^H\mathbf{Q}_j(z_1)\mathbf{y}_j\right|^2\right]\right\}\right)^{p/2}\right]\\
&\qquad+K_p'''\sum_{j=1}^N\frac{1}{N^p}\mathsf{E}\left[\left|\mathbf{y}_j^H\mathbf{Q}_j(z_1)\mathbf{u}_k\mathbf{v}_k^H\mathbf{Q}_j(z_1)\mathbf{y}_j\right|^p\right]\\
&\quad\leq\frac{1}{N^{p/2}}\left[\sum_{j=1}^N\left\{K_p'\mathsf{E}_{j-1}\left[\left|\beta_j\right|^2\right]++K_p''\mathsf{E}_{j-1}\left[\left|\gamma_j\mathbf{y}_j^H\mathbf{Q}_j(z_1)\mathbf{u}_k\mathbf{v}_k^H\mathbf{Q}_j(z_1)\mathbf{y}_j\right|^2\right]\right\}\right]^{p/2}\\
&\qquad+\frac{1}{N^p}K_p'''\sum_{j=1}^N\mathsf{E}\left[\left|\mathbf{y}_j^H\mathbf{Q}_j(z_1)\mathbf{u}_k\mathbf{v}_k^H\mathbf{Q}_j(z_1)\mathbf{y}_j\right|^p\right]\leq\frac{K'}{N^{p/2}}+\frac{K''}{N^{p-1}},
\end{aligned}
$$

where we have repeatedly used the Jensen inequality as well as (2.69) and

$$
\left|\frac{1}{1+\frac{1}{N}\operatorname{Tr}\left[\mathbf{R}\mathbf{Q}_j(z_1)\right]}\right|\leq\frac{|z|}{|\operatorname{Im}\{z\}|}<+\infty,
$$

together with the fact, from Lemma A.4, that all expectations in the last inequality are bounded. Thus, the result is proved by choosing $p>2$.

## Appendix 2.G Proof of Proposition 2.3

In order to prove (2.29), note first that the two factors of $\bar{\eta}(z_1,z_2)$ in (2.27) can be written as

$$
\begin{aligned}
&\mathbf{h}^H\left(w(z_1)\mathbf{R}-z_1\mathbf{I}_M\right)^{-1}\mathbf{R}\left(w(z_2)\mathbf{R}-z_2\mathbf{I}_M\right)^{-1}\mathbf{h}\\
&=\sum_{m=1}^M\frac{\lambda_m\left|\mathbf{h}^H\mathbf{e}_m\right|^2}{\left(w(z_1)\lambda_m-z_1\right)\left(w(z_2)\lambda_m-z_2\right)}=\frac{z_1\mathbf{h}^H\left(w(z_1)\mathbf{R}-z_1\mathbf{I}_M\right)^{-1}\mathbf{h}-z_2\mathbf{h}^H\left(w(z_2)\mathbf{R}-z_2\mathbf{I}_M\right)^{-1}\mathbf{h}}{z_1w(z_2)-z_2w(z_1)},
\end{aligned}
$$

and

$$\frac{f(z_1)-f(z_2)}{z_1-z_2}=\frac{1}{z_1-z_2}\frac{z_1 w(z_2)-z_2 w(z_1)}{w(z_1)\,w(z_2)}.$$

Now, using the asymptotic equivalent of $w(z)$ in (2.39), namely $w(z)\asymp\hat{w}(z)=1-c-cz\hat{b}(z)$ (c.f. Section 1.1), as well as $\mathbf{h}^H(w(z)\mathbf{R}-z\mathbf{I}_M)^{-1}\mathbf{h}\asymp\mathbf{h}^H\left(\hat{\mathbf{R}}-z\mathbf{I}_M\right)^{-1}\mathbf{h}$, we find that

$$\bar{\eta}(z_1,z_2)\asymp\frac{1}{\hat{w}(z_1)\,\hat{w}(z_2)}\mathbf{h}^H\left(\hat{\mathbf{R}}-z_1\mathbf{I}_M\right)^{-1}\hat{\mathbf{R}}\left(\hat{\mathbf{R}}-z_2\mathbf{I}_M\right)^{-1}\mathbf{h}.$$

Thus, we have to solve for the integral

$$\frac{1}{(2\pi\,\mathrm{j})^2}\oint_\Gamma\oint_\Gamma z_1^i z_2^j\frac{h(z_1,z_2)}{\left(1-\frac{c}{M}\sum_{m=1}^{M}\frac{\hat{\lambda}_m}{\hat{\lambda}_m-z_1}\right)\left(1-\frac{c}{M}\sum_{m=1}^{M}\frac{\hat{\lambda}_m}{\hat{\lambda}_m-z_2}\right)}dz_1dz_2, \tag{2.95}$$

where we have defined $h(z_1,z_2)=\mathbf{h}^H\left(\hat{\mathbf{R}}-z_1\mathbf{I}_M\right)^{-1}\hat{\mathbf{R}}\left(\hat{\mathbf{R}}-z_2\mathbf{I}_M\right)^{-1}\mathbf{h}$ and also used that

$$\hat{w}(z)=1-\frac{c}{M}\operatorname{Tr}\left[\hat{\mathbf{R}}\left(\hat{\mathbf{R}}-z\mathbf{I}_M\right)^{-1}\right].$$

To that effect, we may first concentrate on

$$\frac{1}{2\pi\,\mathrm{j}}\oint_\Gamma\frac{P(z_1)}{Q(z_1)}dz_1, \tag{2.96}$$

where we have defined numerator and denominator as $P(z_1)=z_1^i h(z_1,z_2)$ and $Q(z_1)=\hat{w}(z_1)$, respectively. In this case, the integrand has $M$ simple poles at $\mu_k$, $k=1,\ldots,M$, namely solutions to the following equation in $\mu_p$

$$\frac{1}{M}\sum_{m=1}^{M}\frac{\hat{\lambda}_m}{\hat{\lambda}_m-\mu_p}=\frac{1}{c}. \tag{2.97}$$

Note that the solutions of (2.97) are such that $0\le\mu_1<\hat{\lambda}_1<\mu_2<\hat{\lambda}_2<\cdots<\mu_M<\hat{\lambda}_M$, with $\mu_1=0$ if and only if $c>1$. Moreover, using the characterization of the (asymptotic) support of the eigenvalue density of the SCM in [Mes06a] (see also outline in Section 1.1), all $M$ poles can be seen to be located inside the eigenvalue support of $\hat{\mathbf{R}}$ as $M,N\to\infty$ , and, correspondingly, to belong to the region of integration. Thus, the integral may be obtained as the sum of the residues at each one of all these poles. Now, since $P(\mu_k)\neq 0$, $Q(\mu_k)=0$ and $Q'(\mu_k)\neq 0$, we may obtain the residue at each pole as (see e.g. [Mar99a])

$$\operatorname{Res}\left(\frac{P(z_1)}{Q(z_1)},\mu_k\right)=\frac{P(\mu_k)}{Q'(\mu_k)}=-\frac{\mu_k^i h(\mu_k,z_2)}{\frac{c}{M}\operatorname{Tr}\left[\hat{\mathbf{R}}\left(\hat{\mathbf{R}}-\mu_k\mathbf{I}_M\right)^{-2}\right]},$$

and, hence, solve for (2.96) as

$$\frac{1}{2\pi\,\mathrm{j}}\oint_\Gamma\frac{P(z_1)}{Q(z_1)}\,dz_1=-\sum_{k=1}^{M}\mu_k^i\frac{h(\mu_k,z_2)}{\frac{c}{M}\operatorname{Tr}\left[\hat{\mathbf{R}}\left(\hat{\mathbf{R}}-\mu_k\mathbf{I}_M\right)^{-2}\right]}.$$

Further, we continue

$$\frac{1}{2\pi \mathrm{j}} \oint_{\Gamma} \frac{R(z_2)}{S(z_2)} dz_2,$$

upon defining

$$R(z_2) = -z_2^j \sum_{k=1}^{M} \mu_k^i \frac{h(\mu_k, z_2)}{\frac{c}{M} \operatorname{Tr}\left[\hat{\mathbf{R}}\left(\hat{\mathbf{R}} - \mu_k \mathbf{I}_M\right)^{-2}\right]},$$

and $S(z_2) = \hat{w}(z_2)$. Similarly as before, we have

$$\operatorname{Res}\left(\frac{R(z_2)}{S(z_2)}, \mu_l\right) = \frac{R(\mu_l)}{S'(\mu_l)} = \mu_l^j \sum_{k=1}^{M} \mu_k^i \frac{h(\mu_k, \mu_l)}{\frac{c}{M} \operatorname{Tr}\left[\hat{\mathbf{R}}\left(\hat{\mathbf{R}} - \mu_k \mathbf{I}_M\right)^{-2}\right] \frac{c}{M} \operatorname{Tr}\left[\hat{\mathbf{R}}\left(\hat{\mathbf{R}} - \mu_l \mathbf{I}_M\right)^{-2}\right]},$$

and conclude with

$$\mathbf{h}^H \hat{\mathbf{R}}^i \mathbf{R} \hat{\mathbf{R}}^j \mathbf{h} = \sum_{l=1}^{M} \sum_{k=1}^{M} \mu_l^j \mu_k^i \frac{h(\mu_k, \mu_l)}{\frac{c}{M} \operatorname{Tr}\left[\hat{\mathbf{R}}\left(\hat{\mathbf{R}} - \mu_k \mathbf{I}_M\right)^{-2}\right] \frac{c}{M} \operatorname{Tr}\left[\hat{\mathbf{R}}\left(\hat{\mathbf{R}} - \mu_l \mathbf{I}_M\right)^{-2}\right]},$$

as we wanted to prove. Alternatively, the integral in (2.95) can be directly obtained via the Cauchy's integral formula for two complex variables (see e.g. [Hor73]).

# Chapter 3

# Covariance Eigenspectrum Inference and the Problem of Signal Power Estimation

## 3.1 Summary

The fundamental problem of source power estimation in sensor array processing is addressed in this chapter. The emphasis is on the empirical power estimation problem under practical conditions characterized by a finite sample-size and a relatively large array dimension. In such scenarios, classical filter implementations based on the observed array samples usually suffer from a considerable performance degradation. Essentially, the loss in estimation accuracy is due to the fact that the sample estimate only represents a reliable approximation of the true covariance matrix for an increasing number of array samples of strictly fixed dimension. In order to tackle this fundamental limitation, an improved source power estimator is proposed that builds upon an extension of the Capon method delivering remarkably accurate approximations provided a precise knowledge of the noise variance is available. In particular, based on Stieltjes transform methods from random matrix theory, a power estimator is proposed that is consistent for arbitrarily large arrays. In the unknown noise power level case, an estimator of the minimum eigenvalue of the array observation covariance matrix is proposed that builds upon well-known power methods and can be appropriety analyzed with the Stieltjes transform of the sample covariance spectrum. Our power estimator is shown via numerical simulations to generalize and outperform the classical implementation.

## 3.2 Introduction

### 3.2.1 Statistical Inference of Signal Power Level

The statistical estimation of the signal power of a received source impinging on an antenna array from a certain given direction is a fundamental problem in sensor array signal processing [Tre02]. Indeed, the received power level of an intended source is required for the implementation of the minimum variance as well as the minimum mean-square error filtering solutions for the estimation of its signal waveform. On the other hand, the analogous problem of estimating the amplitude of a number of received users is also of special interest in the field of wireless communications, namely, for the computation of the signal-to-interference-plus-noise ratio (SINR) required for power control algorithms, as well as for the implementation of multiuser detectors in code-division multiple access (CDMA) systems, either in forward or reverse link transmissions.

Maximum likelihood techniques have been traditionally proposed in the literature as optimal solutions being only hindered by their associated computational complexity and their sensibility to an inaccurate knowledge of the signal signatures in practice. In array processing applications, the Capon power estimator has been widely applied to the case of uncorrelated signal transmissions in order to obtain a suboptimal yet low-complex minimum variance power estimate.

Under the assumption of temporarily and spatially white noise, a power estimation solution was proposed in [McC02] that, based on subspace-fitting concepts, improves on the minimum variance source power estimator under the assumption of knowledge of the array covariance matrix and the noise power level. In practice, neither the second-order statistics of the received array observations nor the noise variance are available. More importantly, exact approximations of these two quantities by a consistent estimator are theoretically only affordable for an infinite number of observed samples. However, the practical implementation of the power estimators introduced above must be necessary accomplished under the constraint of a limited number of noisy array observations, and so may the estimator performance suffer from a severe degradation even rendering power estimates of an unacceptable quality.

In order to enhance the estimator performance under finite sample-size situations, we will proceed by following a two-level refinement of the traditional construction of the previous subspace estimator. More specifically, we rely on the Taylor expansion of the source of interest (SOI) power approximant in terms of the estimates of the noise variance and a set of vector-valued functions of the negative powers of the covariance matrix. In particular, the previous series expansion provides us with an extra number of filtering degrees of freedom to be possibly exploited in order to alleviate the effects of an imprecise input data (as e.g. an erroneous noise variance or an inaccurate sample estimate of the covariance matrix). Only the received array samples and the spatial signature vector associated with the intended user are considered to be available. Under this assumption, relying on the whiteness assumption on the noise process, the noise power is estimated as the minimum eigenvalue of the sample estimate of the observation covariance matrix, namely the sample covariance matrix (SCM). Indeed, for Gaussian observations, the latter is the maximum likelihood (ML) estimator of the true minimum eigenvalue [And03, Mui82].

Furthermore, in order to improve the estimation performance, the filter expansion is approximated via a generalized consistent estimator that is consistent for arbitrarily high dimensional array observations, or, equivalently, for a limited number of samples per filtering degree of freedom (i.e., number of array sensors). Consequently, the proposed SOI power estimator generalizes conventional constructions appropriately approximating the true power level only for an increasingly large number of samples of fixed dimension. In particular, an improved estimation performance is to be expected in realistic setups since, as in practice, the observation size and dimension are assumed to be comparable in magnitude. For our purposes, we resort to the theory of the spectral analysis of large dimensional random matrices, or random matrix theory (RMT). Specifically, we build upon the asymptotic characterization based on Stieltjes transform methods of the eigenvalue spectrum of SCM-type matrices in the double-limiting regime defined as both the number samples and the observation dimension increase without bound at the same rate. Also based on RMT tools, a similar implementation of the previous source power estimator has been proposed in [Mes07].

Next, we address the problem of inferring the eigenvalue spectrum of the array observation covariance matrix is reviewed and, particularly, the consistent estimation of the minimum eigenvalue in the previous doubly-asymptotic regime.

### 3.2.2 Eigenspectrum Estimation in Signal Processing Applications

The solution of a large number of estimation problems addressed in statistical signal processing relies on the second-order statistics of a set of multidimensional observations [Sch91]. Indeed, the covariance matrix of the received signals plays a fundamental role in most sensor array signal processing applications. In many situations, the eigenspectrum of this matrix, especially the extremal part, i.e., maximum and/or minimum eigenmodes and their associated eigensubspaces, is required (for instance, in eigen-filtering problems and principal component analysis). The algebraic symmetric eigenvalue problem is since long a thoroughly reviewed topic in the mathematical literature of numerical multivariate analysis [Par80, Wil88]. However, as mentioned above, since the true covariance matrix of the array observations is unknown, a collection of received signals must be used to compute a sample estimate, and so is the eigenspectrum of the theoretical covariance matrix in practice necessarily inferred from the spectrum of the sample covariance matrix.

In many practical situations, a sufficiently large number of samples is not a tenable assumption under realistic operation conditions, being hardly justified by a sample-size that is usually not much (or even) larger than the dimension of the array observations. Consequently, the eigenvalue estimates obtained via the eigendecomposition of the sample covariance matrix are often a rather imprecise approximation of the true covariance spectrum. Moreover, this loss in estimation performance is clearly more apparent in high dimensional, relatively low sample support scenarios.

For our purposes of estimating the SOI power level under the alternative more plausible assumption of a sample size and dimension being comparable in magnitude, we consider an estimator of the minimum eigenvalue of the theoretical covariance matrix that, unlike the traditional estimator based on the direct eigendecomposition of the SCM, is consistent in the doubly asymptotic regime introduced above. In particular, we build on classical power methods, which are particularly well posed for a convenient application of known Stieltjes transform based results on the asymptotic convergence of resolvents of SCM-type matrices. Also based on results from RMT, in [Eve00] a Bayesian estimation framework is provided for the inference of the eigenvalues of covariance matrices from limited sample data, that is based on results from RMT. Related contributions from the statistics community that are also based on the spectral analysis of large-dimensional random matrices are [Rao07, Kar07]. In particular, our approach is similar to that presented in [Mes06a], but allows for less restrictive application requirements.

The rest of the chapter is organized as follows. Section 3.3 presents the class of power methods in which the generalized consistent noise power estimator is based. In Section 3.4, the problem of source power estimation is presented. Section 3.5 reviews the performance of classical SCM-based implementations in the more suitable doubly asymptotic regime considered in this work. An improved consistent estimation is proposed in Section 3.6 and numerically evaluated in Section 3.7. After the final conclusions in Section 3.8, the main derivations are provided in the appendices.

## 3.3 A class of power methods

The family of power methods essentially develop the fundamental mechanism driving a matrix to a diagonal form by detecting a certain class of invariant subspaces. Basically, the power method is based on the observation that if we multiply a given vector by a certain diagonalizable matrix, then each eigenvector component in the vector is multiplied by the corresponding eigenvalue. This fact constitutes the basis of an iterative procedure delivering the dominant eigenpair of a particular matrix. Since the covariance matrix in most signal processing applications is defined as a Hermitian (strictly) positive definite matrix, we will here focus on the class of power methods for the symmetric eigenvalue problem [Par80, Gol96]. Let $\mathbf{R} \in \mathbb{C}^{M \times M}$ be a covariance matrix and let $\lambda_m(\mathbf{R})$, $\mathbf{q}_m(\mathbf{R})$, $m = 1, \ldots, M$, denote, respectively, the $m$th eigenvalue and $m$th eigenvector of $\mathbf{R}$, such that $\lambda_1(\mathbf{R}) > \ldots > \lambda_M(\mathbf{R})$. We further consider an arbitrary vector $\mathbf{v} \in \mathsf{C}^M$ such that, choosing the eigenvectors of $\mathbf{R}$ as a complete set of orthonormal vectors, we can write

$$\mathbf{v} = \sum_{m=1}^{M} \gamma_m \mathbf{q}_m(\mathbf{R}),$$

where $\gamma_m \neq 0$, $m = 1, \ldots, M$[1]. According to the power method, the sequence $\mathbf{R}\mathbf{v}, \mathbf{R}(\mathbf{R}\mathbf{v}), \mathbf{R}\left(\mathbf{R}^2\mathbf{v}\right), \ldots$, converges to a vector pointing in the direction of the principal eigenvector of $\mathbf{R}$. The iteration vectors $\mathbf{w}_k = \mathbf{R}^{k-1}\mathbf{v}$ are conventionally scaled in order to prevent over- or underflow. Then, normalizing by the vector Euclidean norm, it turns out that

$$\lambda_1(\mathbf{R}) = \lim_{k\to\infty} \frac{\mathbf{w}_k^H \mathbf{R} \mathbf{w}_k}{\mathbf{w}_k^H \mathbf{w}_k} \tag{3.1}$$

$$= \lim_{k\to\infty} \frac{\mathbf{v}^H \mathbf{R}^{2k-1} \mathbf{v}}{\mathbf{v}^H \mathbf{R}^{2k-2} \mathbf{v}}. \tag{3.2}$$

The convergence of the limit in (3.2) to the principal eigenvalue can be readily seen to depend upon the ratio $\lambda_2(\mathbf{R})/\lambda_1(\mathbf{R})$. Equivalently, the same idea can be applied to the matrix $\mathbf{R}^{-1}$ in order to approximate the minimum eigenvalue and its associated eigenvector. The previous convergence rate can be increased applying a scalar shift as $(\mathbf{R} - \alpha\mathbf{I}_M)$, such that the ratio between the first two dominant eigenvalues becomes smaller. The shifted version of the power method can also be used to approximate other than the principal eigenvalue with some limitations. Alternatively, a modification consisting in the joint application of these two variants is mostly employed as a plausible method to improve the speed of convergence and approximate interior eigenmodes without limitations. The iteration is given by the RHS of (3.1), where now $\mathbf{w}_k = (\mathbf{R} - \alpha\mathbf{I}_M)^{-k+1}\mathbf{v}$ and the approximant converges to the eigenvalue closest to $\alpha$. The convergence can be straightforwardly shown to become quadratic and can even be improved on to cubic with the introduction of an updated shift (Rayleigh quotient iteration).

In order to estimate the extreme eigenvalues and their associated eigen-subspaces, we use the following function of the eigenspectrum of $\mathbf{R}$, namely

$$\lambda_\alpha(\mathbf{R}) = \frac{\mathbf{v}^H (\mathbf{R} - \alpha\mathbf{I}_M)^{-k+1} \mathbf{R} (\mathbf{R} - \alpha\mathbf{I}_M)^{-k+1} \mathbf{v}}{\mathbf{v}^H (\mathbf{R} - \alpha\mathbf{I}_M)^{-2k+2} \mathbf{v}}, \tag{3.3}$$

which corresponds to the previous Rayleigh quotient iteration with fixed shift $\alpha$, chosen as an initial candidate arbitrarily close to the approximated eigenmode. Thus, the eigen-subspace associated with the searched eigenvalue is spanned by the approximate eigenvector $\mathbf{w}_k$. For the sake of application purposes, observe that

$$\lambda_\alpha(\mathbf{R}) = \alpha + \frac{\mathbf{v}^H (\mathbf{R} - \alpha\mathbf{I}_M)^{-2k+3} \mathbf{v}}{\mathbf{v}^H (\mathbf{R} - \alpha\mathbf{I}_M)^{-2k+2} \mathbf{v}} \tag{3.4}$$

[1] This condition can be relaxed to only the coefficient associated with the intended eigenvector being non-zero.

In particular, consider the approximation of the minimum eigenvalue of the covariance matrix (or, equivalently, the noise variance). Clearly, since the noise variance is strictly positive (indeed, the matrix $\mathbf{R}$ is Hermitian positive definite), in order to approximate $\lambda_M(\mathbf{R})$, we use $\alpha = 0$ and define from (3.4)

$$\sigma^2_{\mathsf{ISPM}}(k) = \frac{\mathbf{v}^H \mathbf{R}^{-2k+3} \mathbf{v}}{\mathbf{v}^H \mathbf{R}^{-2k+2} \mathbf{v}}, \tag{3.5}$$

for a positive integer $k$, where $\mathbf{v}$ is an arbitrary randomly generated vector. Considerably accurate approximations can be obtained using up to the fourth negative power of the covariance matrix.

## 3.4 Enhanced Power Estimation under Sample-Size Constraints

Consider a collection of $N$ multivariate observations $\left\{\mathbf{y}(n) \in \mathsf{C}^M\right\}$ obtained by sampling across an antenna array with $M$ sensors, namely, $\{y_m(n), n = 1, \ldots, N, m = 1, \ldots, M\}$, such that $\mathbf{y}(n) = \left[ y_1(n) \cdots y_M(n) \right]^T$. A number of $K$ different sources are supposed to impinge on the antenna array from different directions. Under the assumption of narrowband signals and linear antenna elements, the array observation $\mathbf{y}(n) = \left[ y_1(n) \cdots y_M(n) \right]^T \in \mathsf{C}^M$ can be additively decomposed as

$$\mathbf{y}(n) = x(n)\,\mathbf{s} + \mathbf{n}(n)\,, \tag{3.6}$$

where $x(n) \in \mathsf{C}$ models the signal waveform (or fading channel coefficient) associated with a given signal of interest at the $n$th discrete-time instant and $\mathbf{s} \in \mathsf{C}^M$ is its spatial signature vector (also steering vector or array transfer vector); furthermore, $\mathbf{n}(n) \in \mathsf{C}^M$ is the additive contribution of the interfering sources and background noise, which can be additively decomposed as $\mathbf{n}(n) = \sum_{k=1}^{K-1} x_k(n)\,\mathbf{s}_k + \boldsymbol{\upsilon}(n)$, where, for $k = 1, \ldots, K-1$, $x_k(n) \in \mathsf{C}$ and $\mathbf{s}_k \in \mathsf{C}^M$ are, respectively, the interfering signal processes and associated steering signatures, and $\boldsymbol{\upsilon}(n) \in \mathsf{C}^M$ is the system noise and out-of-system interference. Conventionally, the signals and the noise are assumed to be independent and jointly distributed wide-sense stationary random processes, with SOI power and noise covariance given, respectively, by $\mathsf{E}\left[x^*(n)\,x(n)\right] = \sigma_x^2 \delta_{m,n}$ and $\mathsf{E}\left[\mathbf{n}(m)\,\mathbf{n}^H(n)\right] = \mathbf{R}_{\mathsf{n}} \delta_{m,n}$. Note that this model analogously encompasses a broad range of system configurations described by the general vector channel model in signal processing and wireless communications.

In this work, we focus on the problem of estimating the signal waveform of the intended source and, specifically, on the statistical approximation of the SOI power using optimal spatial filtering techniques. In particular, the Capon beamformer is obtained from the following linearly constrained quadratic optimization problem, namely, [Sto05]

$$\mathbf{w}_{\mathsf{CAPON}} = \arg \min_{\mathbf{w} \in \mathsf{C}^M} \mathbf{w}^H \mathbf{R} \mathbf{w} \;\; \text{subject to}\, \mathbf{w}^H \mathbf{s} = 1, \tag{3.7}$$

where $\mathbf{R}$ is the theoretical covariance matrix of the array observation, which under the previous statistical assumptions is given by

$$\mathbf{R} = \sigma_x^2 \mathbf{s}\mathbf{s}^H + \mathbf{R}_{\mathsf{n}}. \tag{3.8}$$

The solution to (5.2) can be straigthforwardly found as

$$\mathbf{w}_{\mathsf{CAPON}} = \frac{\mathbf{R}^{-1}\mathbf{s}}{\mathbf{s}^H \mathbf{R}^{-1}\mathbf{s}}. \tag{3.9}$$

Indeed, the filtering solution in (3.9) is known to maximize the SINR, defined for a filter $\mathbf{w} \in \mathsf{C}^M$ as

$$\mathsf{SINR}(\mathbf{w}) = \frac{\sigma_x^2 \left|\mathbf{w}^H \mathbf{s}\right|^2}{\mathbf{w}^H \mathbf{R}_{\mathsf{n}} \mathbf{w}} = \left( \frac{\mathbf{w}^H \mathbf{R} \mathbf{w}}{\sigma_x^2 \left|\mathbf{w}^H \mathbf{s}\right|^2} - 1 \right)^{-1}. \tag{3.10}$$

On the other hand, knowledge of the spatial signature vectors corresponding to the sources in the scenario can be conveniently exploited by classical ML methods in order to estimate the power of the SOI (see e.g. [Ott93]). Here, as stated above, we consider methods only relying on the knowledge of the desired source. In this case, from the signal model above, the SOI power can be approximated by $\mathsf{E}\left[\left|\hat{x}(n)\right|^2\right] = \mathbf{w}^H \mathbf{R} \mathbf{w}$. Hence, with some abuse of notation, the Capon SOI power estimate is defined as

$$\sigma^2_{\mathsf{CAPON}} = \frac{1}{\mathbf{s}^H \mathbf{R}^{-1}\mathbf{s}}. \tag{3.11}$$

The rationale behind this procedure is that a natural (indirect) solution for the SOI power approximant must be possibly obtained by minimizing the power of the interference-plus-noise received contribution in (5.2) while keeping the intended signal unchanged, such that the SINR is maximized.

Different direct interpretations of the solution in (3.11) have been reported in the literature. Based on the formulation in [Mar83] of the classical Capon beamforming problem, the optimum power approximant is directly obtained in [Sto03] by recasting the problem into the following covariance-fitting form, namely

$$\sigma^2_{\mathsf{CAPON}} = \max_{\sigma} \sigma \ \text{ subject to } \mathbf{R} - \sigma \mathbf{s}\mathbf{s}^H > 0, \tag{3.12}$$

from which the optimum solution is derived as the argument of the the largest possible scaled version of SOI covariance that can be a part of $\mathbf{R}$ under the constraint of preserving the positive definite signal structure associated with the residual covariance matrix.

In [Lag08], the quantity scaling the (rank-one) SOI signature covariance matrix is identified with the Lagrange multiplier associated with the equality constraint in (5.2). In fact, the Lagrange multiplier associated with the optimum linear transformation is given by the solution in (3.11) to the power estimation problem. More interestingly, this interpretation provides a direct connection between the optimum SOI power estimate (along with the associated signature vector) and the eigenspectra of $\mathbf{R}$. Indeed, $\sigma^2_{\mathsf{CAPON}}$ can be equivalently found as the maximum generalized eigenvalue of the matrix pencil $\left\{\mathbf{R}, \mathbf{s}\mathbf{s}^H\right\}$, being the associated generalized eigenvector proportional to $\mathbf{w}_{\mathsf{SINR}} = \mathbf{R}^{-1}\mathbf{s}$. Hence, similar to (3.12), the power estimation problem can be reformulated as

$$\sigma^2_{\mathsf{CAPON}} = \max_{\lambda} \lambda \;\; \text{subject to} \, \lambda_M \left(\mathbf{R} - \lambda \mathbf{s}\mathbf{s}^H\right) = 0, \tag{3.13}$$

where $\lambda_M \left(\mathbf{R} - \lambda \mathbf{s}\mathbf{s}^H\right)$ denotes the minimum eigenvalue of the matrix $\mathbf{R} - \lambda \mathbf{s}\mathbf{s}^H$.

Interestingly enough, the problem formulation in (3.13) turns out to be of significant importance, since it suggests a straightforward refinement leading to an improvement of the minimum variance power estimation method. Assume the noise covariance matrix is $\mathsf{E}\left[\boldsymbol{v}(m)\,\boldsymbol{v}^H(n)\right] = \sigma_v^2 \mathbf{I}_M \delta_{m,n}$. Then, an enhanced power estimate can be obtained by setting the minimum eigenvalue of the resulting spectrum substruction to the value of the noise variance [Lag08], i.e.

$$\sigma^2_{\mathsf{SSMUSIC}} = \max_{\lambda} \lambda \;\; \text{subject to} \, \lambda_M \left(\mathbf{R} - \lambda \mathbf{s}\mathbf{s}^H\right) = \sigma_v^2. \tag{3.14}$$

The solution to (3.14) accepts a closed-form expression, which is given by

$$\sigma^2_{\mathsf{SSMUSIC}} = \frac{1}{\mathbf{s}^H \left(\mathbf{R} - \sigma_v^2 \mathbf{I}_M\right)^{\#} \mathbf{s}}. \tag{3.15}$$

This SOI power estimate was also derived in [McC02] by taking advantage of the signal-plus-noise structure of the input covariance matrix and its subspace decomposition. Moreover, note that the connection to the classical Capon power estimate can be readily found by expanding the previous solution in its Taylor series about zero as

$$\sigma^2_{\mathsf{SSMUSIC}}(D) = \frac{1}{\sum_{d=0}^{D} \left(\sigma_v^2\right)^d \;\; \mathbf{s}^H \mathbf{R}^{-(d+1)} \mathbf{s}}, \tag{3.16}$$

such that

$$\sigma^2_{\mathsf{SSMUSIC}} = \lim_{D \to \infty} \sigma^2_{\mathsf{SSMUSIC}}(D).$$

Indeed, the Capon power estimate is the approximation of the improved solution given by the first-term of the expansion.

In view of the lack of exact knowledge of the noise power level, note that a further approximation can be obtained as

$$\sigma^2_{\mathsf{SSMUSIC}}(D, k) = \frac{1}{\sum_{d=0}^{D} \left(\sigma^2_{\mathsf{ISPM}}(k)\right)^d \;\; \mathbf{s}^H \mathbf{R}^{-(d+1)} \mathbf{s}}, \tag{3.17}$$

where $\sigma^2_{\mathsf{ISPM}}$ is the noise variance approximant given in (3.5) for a given $k$.

Unfortunately, usually neither the true covariance matrix nor the noise variance are accurately known in practice, so that their values are necessarily to be inferred from the received array observations. Consequently, in a practical setup, the best approximation of the true $\sigma^2_{\mathsf{SSMUSIC}}$ in (4.10) in terms of the expansion in (3.17) is given for a finite (often small) order $D$.

Conventional implementations of statistical inference methods based on the second-order statistics of the array observations usually rely on the direct substitution of the true covariance matrix for the SCM, defined by

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{y}(n) \mathbf{y}^H(n). \tag{3.18}$$

Classically, such an approach has been traditionally reported in the literature as optimal. Indeed, as discussed above, such estimators can be readily shown to be consistent in the classical sense, i.e., they converge stochastically to the true quantity as the sample-size tends to infinity whereas all other dimensions in the signal model remain constant. However, contrary to this conventionally implicit assumption, practical operation conditions are characterized by a sample size and dimension being comparable in magnitude. In these situations, the performance of the previous methods may unavoidably suffer from a considerable degradation. In the next section, we provide a characterization of the performance of sample estimates in the regime defined by a comparatively large sample size and dimension.

## 3.5 Asymptotic Performance Analysis of Sample Estimators

In this section, we provide an asymptotic characterization of the traditional SCM-based implementation of the power level approximant in (3.17) in a limiting regime defined by not only the number of samples ($N$) but also the observation dimension ($M$) going to infinity at a constant rate. To that effect, we resort to the theory of the spectral analysis of large-dimensional random matrices or RMT. Specifically, we build upon results involving the Stieltjes transform of spectral probability measures, a fundamental building block allowing us to characterize the asymptotic eigenspectrum of the SCM in terms of the limiting spectral distribution of the theoretical covariance matrix. For our purposes, not only the asymptotic spectrum but also the limiting behavior of the associated eigensubspaces are of interest. In particular, we obtain the asymptotic limit of the quantities $\mathbf{s}^H \hat{\mathbf{R}}^{-k} \mathbf{s}$, $k = 1, 2, \ldots$, describing the sample estimate version of (3.17), as both the number of samples $N$ and the observation dimension $M$ increase without bound with $M/N \to c < +\infty$.

For instance, an approximation in the previous double-asymptotic regime of the output SINR of a diagonally-loaded minimum variance distorsionless response (MVDR) beamformer is afforded in [Mes06c] that allows for an improved estimation of the diagonal loading parameter. Indeed, note that a meaningful extension to the more general problem in practice of SINR estimation under an unknown power level of the desired source can be readily obtained by additionally providing an estimate of the source power.

From the statistical characterization of the signal model in (5.1), observe that we can statistically model the observed samples as $\mathbf{y}(n) = \mathbf{R}^{1/2}\mathbf{u}(n)$, where $\mathbf{u}(n) \in \mathbb{C}^M$, $n = 1, \ldots, N$, is a collection of i.i.d. random vectors, whose entries have zero mean real and imaginary parts with variance $1/2$ and bounded higher moments. Therefore, the SCM in (5.10) can be modeled as

$$\hat{\mathbf{R}} = \frac{1}{N}\mathbf{R}^{1/2}\mathbf{U}\mathbf{U}^H\mathbf{R}^{1/2}, \tag{3.19}$$

where the matrix $\mathbf{U} \in \mathbb{C}^{M\times N}$ is constructed using as its columns the vectors $\mathbf{u}(n)$, $n = 1, \ldots, N$. Furthermore, the following assumptions of purely technical interest will be used in our derivations:

(**As1**) The vector $\mathbf{s}$ has uniformly bounded Euclidean norm for all $M$.

(**As2**) The matrix $\mathbf{R}$ has uniformly bounded spectral radius for all $M$.

Without loss of generality, we consider normalized versions of the steering vectors with Euclidean norm equal to 1. In particular, note that the previous consideration allows (**As1**) and (**As2**) to be trivially fulfilled for finite signals and noise variance. In the sequel, given two quantities $a, b$, $a \asymp b$ will denote the fact that both quantities are asymptotic equivalents, i.e., $|a - b| \overset{a.s.}{\to} 0$, with $a.s.$ denoting almost sure convergence.

Define the $k$th eigenvalue moment of the SCM and the theoretical covarince matrix as $\hat{m}_{(-k)} \triangleq \frac{1}{M}\operatorname{Tr}\left[\hat{\mathbf{R}}^{-k}\right]$ and $m_{(-k)} \triangleq \frac{1}{M}\operatorname{Tr}\left[\mathbf{R}^{-k}\right]$, respectively.

**Proposition 3.1** *Let $\mathbf{R}$ be a $M\times M$ Hermitian non-negative definite matrix, whose eigenvalues are uniformly bounded for all $M$, and define $\hat{\mathbf{R}} = \mathbf{R}^{1/2}\mathbf{X}\mathbf{X}^H\mathbf{R}^{1/2}$, with $\mathbf{R}^{1/2}$ denoting any Hermitian square-root of the matrix $\mathbf{R}$ and $\mathbf{X}$ a $M \times N$ complex random matrix, such that the real and imaginary parts of the entries of $N^{-1/2}\mathbf{X}$ are i.i.d. random variables with mean zero, variance $1/2$ and bounded moments. Moreover, consider two nonrandom vectors $\mathbf{a}, \mathbf{b} \in \mathbb{C}^M$ with uniformly bounded Euclidean norm for all $M$. Then, as $M, N \to c < +\infty$,*

$$\mathbf{a}^H\hat{\mathbf{R}}^{-k}\mathbf{b} \asymp \sum_{l=1}^{k}\mu_T(l,k)\ \mathbf{a}^H\mathbf{R}^{-l}\mathbf{b}, \tag{3.20}$$

*and*

$$\frac{1}{M}\operatorname{Tr}\left[\hat{\mathbf{R}}^{-k}\right] \asymp \sum_{l=1}^{k}\mu_T(l,k)\ \frac{1}{M}\operatorname{Tr}\left[\mathbf{R}^{-l}\right]. \tag{3.21}$$

*where $\mu_T(l,k)$ can be recursively obtained in terms of the eigenvalues and eigenvectors of $\mathbf{R}$ as*

$$\mu_T(l,k) = \eta_T(l,k)\sum_{i=1}^{k}\zeta_S(i,l), \tag{3.22}$$

*with*

$$\zeta_S(l,k) = \sum_{\sigma}(-1)^{|\sigma|+1}\frac{\eta_S\left(b_1^+, b_1^-\right)\eta_S\left(b_2^+, b_2^-\right)\cdots}{\eta_S\left(b_1^-, b_1^-\right)\eta_S\left(b_1^+, b_1^+\right)\eta_S\left(b_2^-, b_2^-\right)\mu_t\left(b_2^+, b_2^+\right)\cdots},$$

*where the sum is over all partitions* $\sigma$ *of the set* $\{l, \ldots, k-1\}$ *in contiguous intervals such that* $\sigma$ , $\{(b_1^-, b_1^+), (b_2^-, b_2^+) \ldots, (b_r^-, b_r^+)\}$*, for an* $r \leq k-1$ *and such that* $b_i^+ = b_{i+1}^-$ *for each* $i = 1, \ldots, r$*;* $\eta_S(l,k) = -k!\hat{m}_{(l-k)}$*, if* $l \neq k$*, and* $\eta_S(l,k) = k!(1-c) \equiv k!\beta$*, if* $l = k$*, and*

$$\eta_T(l,k) = l! \sum_{t=(l_1,\ldots,l_k)\in T(l,k)} \xi_P(t)\ \phi_1^{l_1} \cdots \phi_k^{l_k},$$

*where* $T(l,k)$ *is the set of partitions of* $k \in \mathsf{N}$ *in* $l$ *parts,*

$$\xi_P(t) = \frac{l!}{l_1!l_2! \cdots l_k! 1!^{l_1} 2!^{l_2} \cdots k!^{l_k}},$$

$$\phi_q = q \sum_{n=1}^{q-1} \frac{n!}{\beta^{n+1}} \sum_{t\in T(n,q-1)} (-1)^{n+n_1+\ldots+n_{q-1}} \binom{n}{n_1, n_2, \ldots, n_{q-1}} \hat{m}_{(-1)}^{n_1} \cdots \hat{m}_{(-q+1)}^{n_{q-1}}.$$

**Proof.** See Appendix 3.9. ■

Hence, in order to obtain the asymptotic limit of the traditional SCM-based construction of the signal power estimate $\sigma^2_{\mathsf{SSMUSIC}}(D,k)$ in (3.17), it is enough to replace each $\mathbf{a}^H\hat{\mathbf{R}}^{-k}\mathbf{b}$ , $k = 1, 2, \ldots$, by the RHS of (3.20). Clearly, we have that

**Remark 3.1** *The convensional SCM-based estimator of the signal power approximant* $\sigma^2_{\mathsf{SSMUSIC}}(D,k)$ *is not consistent for arbitrarily large-dimensional array observations.*

In order to alleviate the negative effects of a limited sample-support and comparably large array observation dimension, in the next section we introduce a class of generalized consistent estimators of the previous power approximant that are strongly consistent for arbitrarily large arrays (or, equivalently, for a limited number of observations per degree-of-freedom). Before presenting the generalized consistent estimator, we briefly discuss an important special case of the result in Proposition 3.1.

### 3.5.1 Asymptotic Moments of the Inverse Wishart Distribution

From the next general result (3.21) in Proposition 3.1 regarding the asymptotic eigenvalues moments of inverse Wishart matrices with arbitrary correlation follows, the following expression follows for the negative moments of the Marçenko-Pastur distribution, namely,

$$M_{(-k)}^{MP} = \sum_{l=1}^{k} \mu_S(l,k). \tag{3.23}$$

The characterization of the limiting moments in (3.23) can also be obtained from the inverse moments of the complex central Wishart distribution (see [Gra03]) by letting the matrix dimension and the number of degrees of freedom go to infinity with a fixed aspect ratio (see also [Tul04, Section 2.3.3])[2]. As indicated in [Mai00], the analytical characterization of the moments

[2]Note that $M^{-1}\hat{\mathsf{R}} \sim \mathcal{W}_M(N, \mathsf{R})$.

of SCM-type matrices is very often of interest in the study of the statistical properties of parameter estimators in statistical signal processing. For instance, the first and second order moments of the SCM were obtained in [Xia05] with the aim of evaluating the large-system output SINR of the recursive least-square filter.

## 3.6 Extremal Eigenspectrum Inference and Consistent Power Estimation

In this section, we present an generalized consistent estimator of the SOI power estimate $\sigma^2_{\mathsf{SSMUSIC}}(D,k)$ in (3.17) based on a SCM constructed from arbitrarily high-dimensional array observations. To that effect, note that it is enough to consider the estimation of $\mathbf{s}^H\mathbf{R}^{-k}\mathbf{s}$, $k = 1, 2, \ldots$. In particular, based on RMT results on the asymptotic spectrum of the SCM, a function of the negative moments of $\hat{\mathbf{R}}$ is obtained that converges to $\mathbf{s}^H\mathbf{R}^{-k}\mathbf{s}$ as $M, N \to \infty$, with $M/N \to c < +\infty$. We will refer to these estimators as $M, N$-consistent as a generalization of traditional $N$-consistent estimators.

In this context, we make use of the Stieltjes transform from RMT that allows us to characterize the asymptotic distribution of the eigenvalues of $\hat{\mathbf{R}}$ in terms of the limiting eigenvalue distribution of $\mathbf{R}$.

**Proposition 3.2** *Consider the assumptions and definitions in Proposition 3.1. Then, as $M, N \to c < +\infty$,*

$$\mathbf{a}^H\mathbf{R}^{-k}\mathbf{b} \asymp \sum_{l=1}^{k} \mu_S(l,k)\ \mathbf{a}^H\hat{\mathbf{R}}^{-l}\mathbf{b}, \tag{3.24}$$

*and*

$$\frac{1}{M}\operatorname{Tr}\left[\hat{\mathbf{R}}^{-k}\right] \asymp \sum_{l=1}^{k} \mu_S(l,k)\ \frac{1}{M}\operatorname{Tr}\left[\mathbf{R}^{-l}\right], \tag{3.25}$$

*where*

$$\mu_S(l,k) = \eta_S(l,k)\sum_{i=1}^{l}\zeta_T(i,k).$$

**Proof.** See Appendix 3.9. ■

As an example, the coefficients defining the estimators in (3.24) and (3.25), for $k = 1, 2, 3$, are given in Table 4.1.

Clearly, the RHS of (3.24) and (3.25) are strongly consistent estimators of the respective LHS. Moreover, a construction of the SOI power estimate $\sigma^2_{\mathsf{SSMUSIC}}(D,k)$ obtained by replacing $\mathbf{s}^H\mathbf{R}^{-k}\mathbf{s}$ with the $M, N$-consistent estimator in (3.24) can be readily shown to be consistent even for an arbitrarily high dimensional array observation. In particular, note that the proposed

Table 3.1: Example estimator coefficients: $k = 1, 2, 3, 4$.

| | | | |
|---|---|---|---|
| $\mu_S(1,1) = \beta$ | – | – | – |
| $\mu_S(1,2) = -\beta\hat{m}_{(-1)}$ | $\mu_S(2,2) = \beta^2$ | – | – |
| $\mu_S(1,3) = \left(\hat{m}^2_{(-1)} - \beta\hat{m}_{(-2)}\right)\beta$ | $\mu_S(2,3) = 2\hat{m}_{(-1)}\beta^2$ | $\mu_S(3,3) = \beta^3$ | – |
| $\mu_S(1,4) = \left(\hat{m}^2_{(-1)} + 3\beta\hat{m}_{(-2)} - \beta^2\hat{m}_{(-3)}\right)\beta$ | $\mu_S(2,4) = \left(3\hat{m}^2_{(-1)} - 2\beta\hat{m}_{(-2)}\right)\beta^2$ | $\mu_S(3,4) = -3\hat{m}_{(-1)}\beta^3$ | $\mu_S(4,4) = \beta^4$ |

implementation does not make any assumption on spectrum separation regarding the asymptotic eigenvalue density of the SCM (see further [Mes06a, Mes07]).

## 3.7 Simulation results

In this section, we numerically evaluate the performance of the proposed SOI power level estimator estimator and compare the results with those obtained by directly replacing the true theoretical covariance matrix and the noise variance with the SCM and its minimum eigenvalue, respectively. Throughout this section, the previous methods will be referred to as *proposed* and *conventional*. We assume an array observation covariance matrix taking the form $\mathbf{R} = \mathbf{SPS}^H + \sigma_v^2\mathbf{I}_M$, where $\mathbf{P}$ is a diagonal matrix containing the power level associated with each source and $\sigma_v^2$ is the noise variance.

Let us first consider the numerical evaluation of the power method based minimum eigenvalue (resp. noise variance) estimator. In Figure 3.1, the histograms obtained from the conventional and proposed eigenvalue estimates over 100 runs are depicted. We have assumed a number of $N = 50$ sample observations are collected with an array of $M = 10$ sensors, and describing an scenario with $K = 3$ sources, one of them received 10dB over the noise level, whereas the rest are received 5dB above the level of the first signal. The noise variance is $\sigma_v^2 = 0.1$. From the empirical probability density function, the variance of the proposed estimator is observed to be larger than that of the conventional estimator. However, the latter is shown to be clearly biased whereas the proposed one can be seen to approximate on average the eigenvalue to be estimated.

Regarding the eigenspace associated with the estimated eigenvalue, in order to uniquely assess the goodness of the eigenvector estimate, we evaluate the following distance function defined in terms of an eigenprojection matrix onto the spectrum of $\mathbf{R}$, namely

$$O(n) = \mathbf{v}^H (\mathbf{R} - \alpha\mathbf{I}_M)^{-k+1} \mathbf{E}_s\mathbf{E}_s^H (\mathbf{R} - \alpha\mathbf{I}_M)^{-k+1} \mathbf{v}, \tag{3.26}$$

for $k = 3$, where the columns of the matrix $\mathbf{E}_s \in \mathbb{C}^{M\times K}$ span the signal subspace of $\mathbf{R}$. Note that the signal subspace is orthogonal to the noise subspace embedding the true eigenvector to be estimated. Hence, a smaller magnitude of the measure in (3.26) should indicate a better estimation performance. The results are shown in Figure 3.2, where a superior performance of

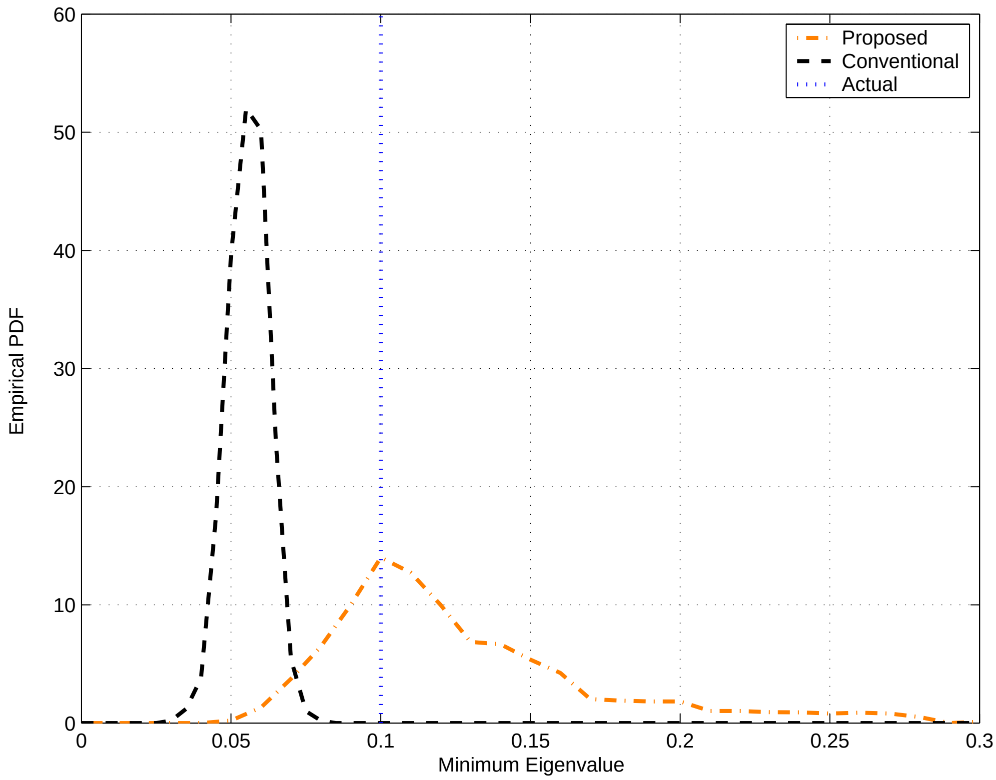


Figure 3.1: Conventional and proposed estimates of the smallest eigenvalue ($M = 10$, $N = 50$ and $K = 3$).

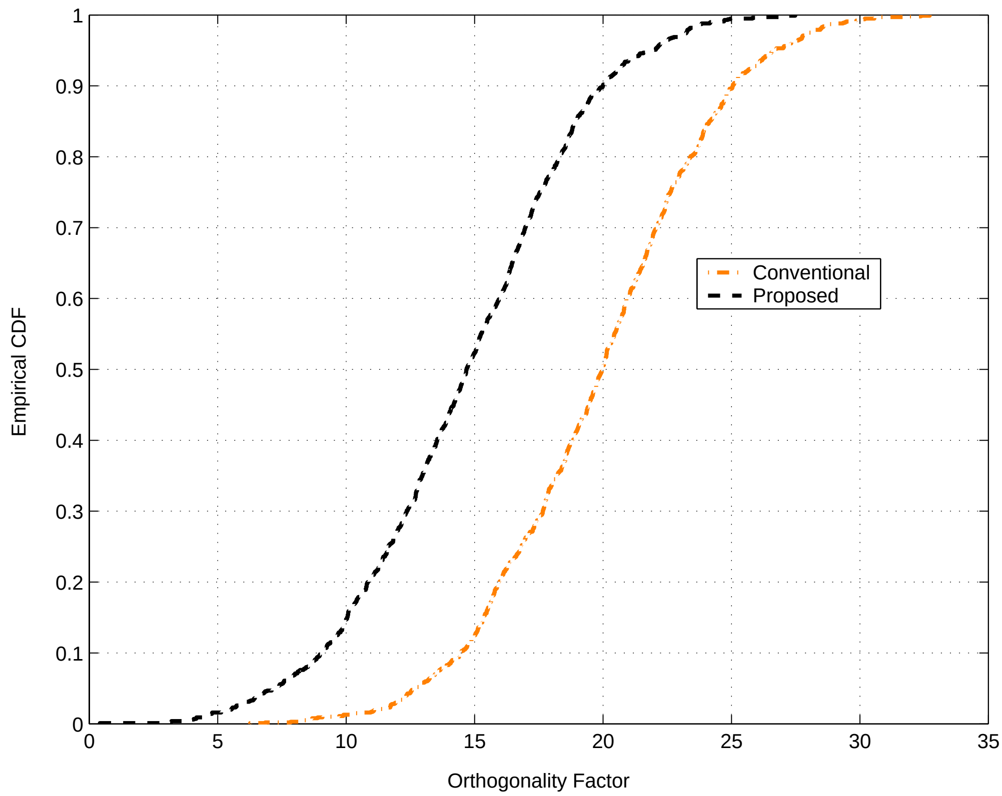


Figure 3.2: Orthogonality factor for conventional and proposed eigenvector estimates

the proposed estimator can be appreciated.

Let us now consider the approximation of (3.17). In Figure 3.3, performance curves are depicted for both implementations of the theoretical estimator in (3.16), along with the direct implementation of (4.10) in terms of the SCM and its minimum eigenvalue (referred to as *traditional* method). We have assumed that the observation dimension is equal to $M = 8$ and that $K = 2$ signals are being received, both with a power level 10dB over the noise level $\sigma_v^2 = 1$. An expansion order of $D = 1$ for the implementation of (3.16) and value of $k = 3$ for the estimation of the noise variance have been used. Averaged results over 1000 realizations are shown.

## 3.8 Conclusions

We have addressed the fundamental problem of source power estimation in sensor array signal processing. Traditional estimators are based on the second-order statistics of the array observations (as e.g. Capon beamformer) and the background noise power (e.g. ML estimator). Specifically, we have focused on the empirical estimation problem in practical scenarios charac-

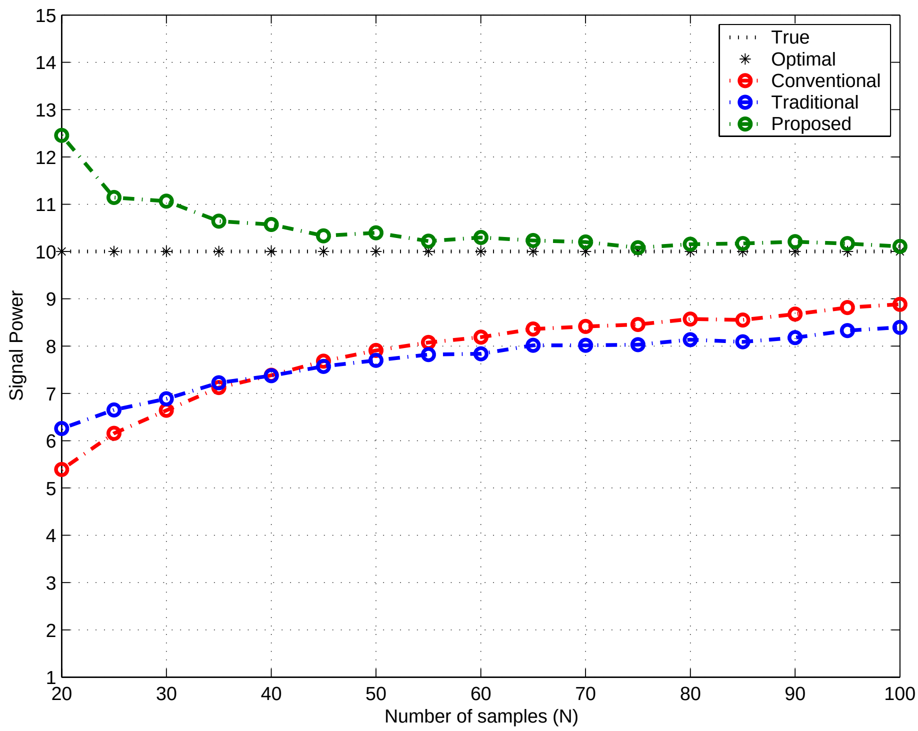


Figure 3.3: Average SOI power estimates obtained with the proposed, conventional and traditional methods. True SOI power level and theoretical improved estimates are shown for reference.

terized by a finite sample-size and a relatively large array observation dimension. Under such conditions, classical filter implementations based on the direct substitution of the true unknown covariance matrix by the SCM may incur a severe performance degradation, as they only prove to be consistent for an increasing number of samples of strictly fixed dimension. In particular, we have built upon an extension of the Capon power estimator that allows for a considerably improved source power estimate provided a precise knowledge of the noise variance is available. In order to avoid the performance degradation due to inaccuracies in the approximation of the observation covariance matrix and the noise power level, we consider an expansion of the optimal solution is terms of their estimates. In particular, the previous quantities are consistently approximated in an asymptotic regime defined by a limited number of samples per array dimension. As a noise power estimator, an approximant of the minimum eigenvalue of the covariance matrix based on the shifted inverse power method has been considered. The motivation to this choice is that such an approximant can be appropriately analyzed and consistently estimated using Stieltjes transform methods from random matrix theory. By assuming that the array dimension can be comparable in magnitude to the sample-size, unlike the estimator directly based on the eigendecomposition of the SCM, a significantly improved estimation accuracy of the proposed power estimator can be obtained in practical finite sample-support scenarios.

# Appendix 3.A Proof of Propositions 3.1 and 3.2

## 3.9 Proof of Propositions 3.1 and 3.2

The following lemma will be useful in proving Propositions 3.1 and 3.2.

**Lemma 3.1** *Let* $\mathbf{R}$ *be a* $M \times M$ *Hermitian non-negative definite matrix, whose eigenvalues are uniformly bounded for all* $M$*, and define* $\hat{\mathbf{R}} = \mathbf{R}^{1/2}\mathbf{X}\mathbf{X}^H\mathbf{R}^{1/2}$*, with* $\mathbf{R}^{1/2}$ *denoting any Hermitian square-root of the matrix* $\mathbf{R}$ *and* $\mathbf{X}$ *a* $M \times N$ *complex random matrix, such that the real and imaginary parts of the entries of* $N^{-1/2}\mathbf{X}$ *are i.i.d. random variables with mean zero, variance* $1/2$ *and bounded moments. Moreover, consider two nonrandom vectors* $\mathbf{a}, \mathbf{b} \in \mathbb{C}^M$ *with uniformly bounded Euclidean norm for all* $M$*. Then, as* $M, N \to c < +\infty$*,*

$$\sum_{l=0}^{k} \eta_T(l,k)\ \mathbf{a}^H \mathbf{R}^{-(l+1)} \mathbf{b} \asymp \sum_{l=0}^{k} \eta_S(l,k)\ \mathbf{a}^H \hat{\mathbf{R}}^{-(l+1)} \mathbf{b}, \tag{3.27}$$

*where* $\eta_S(l,k) = -k!\hat{m}_{(l-k)}$*, if* $l \neq k$*, and* $\eta_S(l,k) = k!(1-c)$*, if* $l = k$*, and*

$$\eta_T(l,k) = l! \sum_{t=(l_1,\dots,l_k)\in T(l,k)} \xi_P(t)\ \phi_1^{l_1}\cdots\phi_k^{l_k},$$

*where* $T(l,k)$ *is the set of partitions of* $k \in \mathbb{N}$ *in* $l$ *parts,*

$$\xi_P(t) = \frac{l!}{l_1! l_2! \cdots l_k! 1!^{l_1} 2!^{l_2} \cdots k!^{l_k}},$$

*and*

$$\phi_q = q \sum_{n=1}^{q-1} \frac{n!}{\beta^{n+1}} \sum_{t=(l_1,\dots,l_{q-1})\in T(n,q-1)} (-1)^{n+n_1+\dots+n_{q-1}} \binom{n}{n_1, n_2, \dots, n_{q-1}} \hat{m}_{(-1)}^{n_1}\cdots\hat{m}_{(-q+1)}^{n_{q-1}}.$$

**Proof.** Recall from Section 1.1 the following results regarding the asymptotic behavior of the eigenvalues and eigenvectors of SCM-type matrices as $M, N \to c < +\infty$, for all $z \in \mathbb{C}^+$, namely,

$$\frac{1}{M}\operatorname{Tr}\left[\left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1}\right] \asymp \frac{1}{M}\operatorname{Tr}\left[\left(w(z)\mathbf{R} - z\mathbf{I}_M\right)^{-1}\right],$$

where $w(z) = 1 - c - czm(z)$ and $m = m(z)$ is the unique solution in the set $\{m \in \mathbb{C} : -(1-c)/z + cm \in \mathbb{C}^+\}$ to the following equation, namely,

$$m = \frac{1}{M}\sum_{m=1}^{M} \frac{\lambda_m(\mathbf{R})}{\lambda_m(\mathbf{R})(1-c-czm) - z}.$$

Furthermore, for two $M$ dimensional deterministic vectors $\mathbf{a}, \mathbf{b}$ with uniformly bounded Euclidean norm for all $M$,

$$\mathbf{a}^H\left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1}\mathbf{b} \asymp \mathbf{a}^H\left(w(z)\mathbf{R} - z\mathbf{I}_M\right)^{-1}\mathbf{b}. \tag{3.28}$$

Equivalently, if we consider $f(z) = z/w(z)$, we have

$$w(z) \asymp \hat{w}(z), \tag{3.29}$$

$$f(z) \asymp \hat{f}(z), \tag{3.30}$$

where

$$\hat{w}(z) = 1 - c - cz\frac{1}{M}\operatorname{Tr}\left[\left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1}\right] \tag{3.31}$$

$$= 1 - \frac{c}{M}\sum_{m=1}^{M}\frac{\lambda_m\left(\hat{\mathbf{R}}\right)}{\lambda_m\left(\hat{\mathbf{R}}\right) - z}, \tag{3.32}$$

and

$$\hat{f}(z) = \frac{z}{\hat{w}(z)}. \tag{3.33}$$

By analytic continuation, it can be proved that the convergence results above hold for $z = x \in \mathsf{R}$ outside the limiting eigenvalue support of $\hat{\mathbf{R}}$. Then, in order to prove the lemma, using

$$\mathbf{a}^H\left(\mathbf{R} - f(z)\mathbf{I}_M\right)^{-1}\mathbf{b} \asymp \hat{w}(z)\,\mathbf{a}^H\left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1}\mathbf{b}, \tag{3.34}$$

we just need to consider

$$\left.\frac{\partial^k}{\partial x^k}\mathbf{a}^H\left(\mathbf{R} - f(x)\mathbf{I}_M\right)^{-1}\mathbf{b}\right|_{x=0} \asymp \left.\frac{\partial^k}{\partial x^k}\hat{w}(x)\,\mathbf{a}^H\left(\hat{\mathbf{R}} - x\mathbf{I}_M\right)^{-1}\mathbf{b}\right|_{x=0}. \tag{3.35}$$

On the one hand, we have

$$\frac{\partial^k}{\partial x^k}\left\{\hat{w}(x)\,\mathbf{a}^H\left(\hat{\mathbf{R}} - x\mathbf{I}_M\right)^{-1}\mathbf{b}\right\} = \sum_{l=0}^{k}\binom{k}{l}\frac{\partial^l}{\partial x^l}\left\{\hat{w}(x)\right\}\frac{\partial^{k-l}}{\partial x^{k-l}}\left\{\mathbf{a}^H\left(\hat{\mathbf{R}} - x\mathbf{I}_M\right)^{-1}\mathbf{b}\right\}.$$

Then, using ($\beta$ , $1-c$)

$$\left.\frac{\partial^l}{\partial x^l}\left\{\hat{w}(x)\right\}\right|_{x=0} = \begin{cases}\beta, & l = 0\\ -l!\ \hat{m}_{(-l)}, & l > 0,\end{cases}$$

and

$$\left.\frac{\partial^{k-l}}{\partial x^{k-l}}\left\{\mathbf{a}^H\left(\hat{\mathbf{R}} - x\mathbf{I}_M\right)^{-1}\mathbf{b}\right\}\right|_{x=0} = (k-l)!\ \mathbf{a}^H\hat{\mathbf{R}}^{-(k-l+1)}\mathbf{b}, \quad q \geq 0,$$

it follows that

$$\left.\frac{\partial^k}{\partial x^k}\left\{\hat{w}(x)\,\mathbf{a}^H\left(\hat{\mathbf{R}} - x\mathbf{I}_M\right)^{-1}\mathbf{b}\right\}\right|_{x=0} = \sum_{l=0}^{k}\eta_S(l,k)\ \mathbf{a}^H\hat{\mathbf{R}}^{-(l+1)}\mathbf{b}, \tag{3.36}$$

with $\eta_S(l,k) = -k!\hat{m}_{(l-k)}$, $l \neq k$, and $\eta_S(l,k) = k!\,(1-c)$, $l = k$.

On the other hand, regarding the LHS of (3.35), we can write the $k$th derivative of the composite function $g(f(x)) = \mathbf{a}^H\left(\mathbf{R} - f(x)\mathbf{I}_M\right)^{-1}\mathbf{b}$ as

$$\frac{\partial^k}{\partial x^k}\left\{\mathbf{a}^H\left(\mathbf{R}-f(x)\,\mathbf{I}_M\right)^{-1}\mathbf{b}\right\}=\sum_{l=1}^{k}\frac{\partial^l}{\partial x^l}\left\{g(f)\right\}\sum_{t=(l_1,\ldots,l_k)\in T(l,k)}\xi_P(t)\prod_{i=1}^{k}\left(f^{(i)}(x)\right)^{l_i}. \quad (3.37)$$

Now, using

$$\left.\frac{\partial^l}{\partial x^l}\left\{g\left(f(x)\right)\right\}\right|_{x=0}=l!\;\mathbf{a}^H\mathbf{R}^{-(l+1)}\mathbf{b},\quad l\geq 0,$$

we can rewrite (3.37) as

$$\left.\frac{\partial^k}{\partial x^k}\left\{\mathbf{a}^H\left(\mathbf{R}-f(x)\,\mathbf{I}_M\right)^{-1}\mathbf{b}\right\}\right|_{x=0}=\sum_{l=1}^{k}\eta_T(l,k)\;\mathbf{a}^H\mathbf{R}^{-(l+1)}\mathbf{b}, \quad (3.38)$$

where we have defined

$$\eta_T(l,k)=l!\sum_{t=(l_1,\ldots,l_k)\in T(l,k)}\xi_P(t)\;\phi_1^{l_1}\cdots\phi_k^{l_k}, \quad (3.39)$$

and

$$\phi_q=\left.\frac{\partial^q}{\partial x^q}\left\{f(x)\right\}\right|_{x=0}.$$

Regarding $\phi_q$, from (5.17), we may write

$$\phi_q=\left.\frac{\partial^q}{\partial x^q}\left\{\frac{x}{\hat{w}(x)}\right\}\right|_{x=0}=q\left.\frac{\partial^{q-1}}{\partial x^{q-1}}\left\{\frac{1}{\hat{w}(x)}\right\}\right|_{x=0},$$

and, further,

$$\left.\frac{\partial^{q-1}}{\partial x^{q-1}}\left\{\frac{1}{\hat{w}(x)}\right\}\right|_{x=0}=\sum_{n=1}^{q-1}\frac{n!}{\beta^{n+1}}\sum_{t=(n_1,\ldots,n_k)\in T(n,q-1)}\kappa_t\binom{n}{n_1,n_2,\ldots,n_{q-1}}\hat{m}_{(-1)}^{n_1}\cdots\hat{m}_{(-q+1)}^{n_{q-1}},$$

with $\kappa_t=(-1)^{n+n_1+\ldots+n_{q-1}}$, and where we have used the fact that

$$\left.\frac{\partial^n}{\partial\hat{w}^n}\left\{\frac{1}{\hat{w}}\right\}\right|_{x=0}=\frac{n!\,(-1)^n}{\beta^{n+1}}.$$

Finally, the result follows from (3.36) and (3.38). ■

Now, observe that the expression in (3.27) can be recurrently inverted in order to obtain either the asymptotic limit in Proposition 3.1 and the $M,N$-consistent estimator in Proposition 3.2. We elaborate on the generalized consistent estimator and note that an equivalent derivation follows for asymptotic performance analysis.

Let us define

$$\gamma_S(k)=\sum_{l=0}^{k}\eta_S(l,k)\;\mathbf{a}^H\hat{\mathbf{R}}^{-l}\mathbf{b}, \quad (3.40)$$

$$
\begin{aligned}
\mathbf{a}^H\mathbf{R}^{-2}\mathbf{b} =& \frac{1}{\eta_T(1,1)}\gamma_S(1),\\
\mathbf{a}^H\mathbf{R}^{-3}\mathbf{b} =& -\frac{\eta_T(1,2)}{\eta_T(2,2)\,\eta_T(1,1)}\gamma_S(1) + \frac{1}{\eta_T(2,2)}\gamma_S(2),\\
\mathbf{a}^H\mathbf{R}^{-4}\mathbf{b} =& -\left(\frac{\eta_T(1,3)}{\eta_T(3,3)\,\eta_T(1,1)} - \frac{\eta_T(1,2)\,\eta_T(2,3)}{\eta_T(3,3)\,\eta_T(2,2)\,\eta_T(1,1)}\right)\gamma_S(1) - \frac{\eta_T(2,3)}{\eta_T(3,3)\,\eta_T(2,2)}\gamma_S(2)\\
&+\frac{1}{\eta_T(3,3)}\gamma_S(3),\\
\mathbf{a}^H\mathbf{R}^{-5}\mathbf{b} =& \left(\frac{\eta_T(1,3)\,\eta_T(3,4)}{\eta_T(4,4)\,\eta_T(3,3)\,\eta_T(1,1)} + \frac{\eta_T(1,2)\,\eta_T(2,4)}{\eta_T(4,4)\,\eta_T(2,2)\,\eta_T(1,1)} - \frac{\eta_T(1,2)\,\eta_T(2,3)\,\eta_T(3,4)}{\eta_T(4,4)\,\eta_T(3,3)\,\eta_T(2,2)\,\eta_T(1,1)}\right.\\
&\left.-\frac{\eta_T(1,4)}{\eta_T(4,4)\,\eta_T(1,1)}\right)\gamma_S(1) - \left(\frac{\eta_T(2,4)}{\eta_T(4,4)\,\eta_T(2,2)} - \frac{\eta_T(2,3)\,\eta_T(3,4)}{\eta_T(4,4)\,\eta_T(3,3)\,\eta_T(2,2)}\right)\gamma_S(2)\\
&-\frac{\eta_T(3,4)}{\eta_T(4,4)\,\eta_T(3,3)}\gamma_S(3) + \frac{1}{\eta_T(4,4)}\gamma_S(4),
\end{aligned}
$$

where the pattern is obvious, namely

$$
\mathbf{a}^H\mathbf{R}^{-k}\mathbf{b} = \sum_{l=1}^{k-1}\zeta_T(l,k)\,\gamma_S(l),
$$

with

$$
\zeta_t(n,l) = \sum_{\sigma}(-1)^{|\sigma|+1}\frac{\eta_T\left(b_1^+,b_1^-\right)\eta_T\left(b_2^+,b_2^-\right)\cdots}{\eta_T\left(b_1^-,b_1^-\right)\eta_T\left(b_1^+,b_1^+\right)\eta_T\left(b_2^-,b_2^-\right)\eta_T\left(b_2^+,b_2^+\right)\cdots},
$$

where the sum is over all partitions $\sigma$of the set $\{n,\ldots,l-1\}$ in contiguous intervals such that $\sigma$ , $\left\{\left(b_1^-,b_1^+\right),\left(b_2^-,b_2^+\right)\ldots,\left(b_r^-,b_r^+\right)\right\}$, for an $r \leq l-1$ and such that $b_i^+ = b_{i+1}^-$ for each $i = 1,\ldots,r$. Finally, using (3.40), we obtain the expression of $\mu_S(l,k)$ in the proposition.

# Chapter 4

# Estimation on Low-Rank Krylov Subspaces of Arbitrary Dimension

## 4.1 Summary

In this chapter, the problem of Krylov subspace estimation using a limited number of received data samples is addressed. The focus is on signal processing applications where the Krylov subspace is defined from the unknown second-order statistics of the observed samples and the signature vector associated with the desired parameter. In particular, the consistency of traditionally optimal estimators is revised and analytically characterized under a more meaningful asymptotic regime, where not only the number of samples but also the observation dimension grow without bound at the same rate. Furthermore, an improved construction of Krylov subspace methods is proposed that is based on the generalized consistent estimation of a set of vector-valued quadratic functions of the covariance matrix powers. To that effect, results on the estimation of spectral covariance functions are borrowed from random matrix theory in order to approximate the previous quantities depending upon not only the spectrum of the covariance matrix but also the associated eigenspace. As a result, a new class of estimators is derived that generalizes conventional filter implementations by proving to be consistent for observations of arbitrarily high dimension. The proposed estimators are shown to outperform traditional constructions via the numerical evaluation of two fundamental problems in sensor array processing, namely the problem of estimating the power of an intended source and the estimation of the principal eigenspace and dominant eigenmodes of a structured covariance matrix.

## 4.2 I Introduction

Linear estimation over Krylov subspace expansions [Saa96, Vor03b] has found a wide variety of applications in different areas of statistical signal processing. In particular, iterative linear approximations of the minimum variance unbiased estimator (MVUE) and the minimum mean-square error (MMSE) estimator on low-dimensional subspaces are extensively applied to inference problems in the fields of communications, such as in channel estimation, equalization and symbol detection, and sensor array signal processing, as in adaptive beamforming and passive radar/sonar. In the literature, Krylov subspace expansions for optimum filtering are well-known to be dual to the problem of iterative search for quadratic minimization [Die07]. More specifically, the Krylov subspace defined by the observation covariance matrix and the signature vector associated with the signal of interest (SOI) is identified in [Hon01] as the span describing the expansion of the orthogonal multistage Wiener filter (MSWF) introduced in [Gol98] as, essentially, a concatenation of filtering stages based on the generalized sidelobe canceller (GSC) [Sch91]. In [Wei02], and more generally in [Sch03a], the equivalence between the previous iterative subspace Wiener filter and the conjugate gradient method (CGM) (see e.g. [Lue84, Section 8.3]) is ascertained. Indeed, the CGM is known to yield numerically stable solutions of symmetric positive linear systems (as well as unconstrained quadratic optimization problems) through the recursive construction of an orthonormal basis for the associated Krylov subspace. Furthermore, the previously defined Krylov subspace turns out to represent the optimum linear projection space in the sense of minimizing the squared error norm of the filter approximation [Lue84, Gol96].

In practice, implementations of the optimum (full-rank) estimator based on the above subspace filtering solutions are usually considered due to their robustness against the two major problems related to the direct conventional realization: the computational complexity associated with the matrix inversion operation and the sample-support requirements for the estimation of the unknown covariance matrix. Interestingly enough, rank-reduction based on Krylov subspaces enables a decrease of the approximation subspace dimension without performance loss. Indeed, contrary to dimension reduction based on the eigendecomposition of the covariance matrix, Krylov subspace methods can be shown to achieve optimum (full-rank) performance for a number of iterations or rank lower than the dimension of the signal subspace. In [Ge04], this fact is exemplified in the context of arbitrarily loaded code-division multiple access (CDMA) systems employing Gold spreading codes, as well as in sensor array signal processing applications, where the number of sources within a beamwidth is identified as the relevant measure of convergence rate characterizing angle-dependent data dimensionality reduction.

Classical examples of reduced-rank linear estimation on Krylov subspaces have been reported in the literature for sensor array signal processing applications, particularly for adaptive beamforming and passive radar/sonar in, for instance, [Gue00, San03, Ge06, San07] and the much earlier works [Erm93, Erm94]; moreover, contributions to the theory of adaptive filtering, equalization and interference cancellation can be found in [Bur02, Xia05] and, respectively, [Hon06, Mou07, Dum07] and [Hon02]; applications to the vector channel model underlying the problem of CDMA multiuser detection and the design of MIMO linear transceivers have been proposed in e.g. [Lou03, Li04b, Tri05, Cot05]. Finally, other signal processing applications of Krylov subspace methods in the engineering literature include the estimation of the principal eigenspace of covariance matrices with signal-plus-noise structure [Xu94], as well as the estimation of the error variance of the Bayesian minimum mean-square error (MMSE) estimate [Sch00] and the covariance low-rank approximation problem in oceanographic remote sensing [Sch03b].

As mentioned above, the practical realization of the optimum Krylov subspace estimator relies on the sample estimate of the (unknown) theoretical covariance matrix. For a reasonably large number of data samples, the sample covariance matrix (SCM) approximates the true covariance with plausibly low estimation error. However, the performance loss incurred by a covariance approximation based on a particularly limited number of samples is quickly aggravated as the observation dimension becomes higher, since the number of required observations is considerably increased. Equivalently, a severe degradation of the filter performance can be expected in scenarios characterized by a low number of filtering degrees of freedom when the sample-support is comparably small. Essentially, the previous observations are due to the fact that the convergence rate of classical consistent estimators based on the SCM is significantly reduced in situations where the sample size and dimension are comparable in magnitude.

In this chapter, we provide a characterization of the performance of Krylov subspace inferential methods under a bounded number of samples per degree-of-freedom. Furthermore, a class of reduced-rank estimators is proposed that generalize traditional implementations by proving to be consistent for observations of arbitrarily high dimension. This fact is shown via numerical simulations to translate into an improved performance in practical finite sample-size scenarios. In particular, we evaluate the proposed estimators for the sensor array processing applications of minimum variance estimation of the SOI power and the identification of the principal eigenspace of the array covariance matrix.

The chapter is organized as follows. In Section 4.3, the definition of Krylov subspace is introduced and the two subspace estimation problems addressed in the chapter are described. In Section 4.4, an asymptotic performance characterization of Krylov subspace methods under a comparably large sample size and dimension is provided. Section 4.5 presents the proposed generalized consistent estimator, which is numerically evaluated in Section 4.6 in the context of the applications of source power approximation and estimation of the covariance principal eigenspaces. After a short discussion and the concluding remarks in Section 4.7, the theoretical framework for the derivations in the chapter is developed in the appendices.

## 4.3 Linear estimation on Krylov subspaces

In this section, we review the linear signal model underlying a typical array processing application and introduce the associated Krylov subspace of interest for estimation purposes. Two fundamental applications are then presented, namely the problem of linearly estimating the power and signal waveform of an intended source via the subspace equivalent representation of classical minimum variance methods and the problem of inferring the principal eigenpair of the covariance matrix of the array observations.

Consider a collection of multivariate observations $\left\{\mathbf{y}(n) \in \mathbb{C}^M\right\}$ obtained by sampling across an antenna array with $M$ sensors, namely, $\{y_m(n), n = 1, \ldots, N, m = 1, \ldots, M\}$, such that $\mathbf{y}(n) = \left[y_1(n) \cdots y_M(n)\right]^T$. The received signals can be most generally modeled as

$$\mathbf{y}(n) = x(n)\,\mathbf{s} + \mathbf{n}(n), \tag{4.1}$$

where $x(n) \in \mathbb{C}$ is the signal waveform transmitted by the source of interest at the discrete-time instant $n$, $\mathbf{s} \in \mathbb{C}^M$ is the associated steering signature vector and $\mathbf{n}(n) \in \mathbb{C}^M$ is the additive contribution of the interference and system noise. Conventionally, the signal and interference-plus-noise components are assumed to be independent and jointly distributed wide-sense stationary random processes, with signal power and noise covariance given, respectively, by $\mathsf{E}\left[x^*(n)\,x(m)\right] = \sigma_x^2\delta_{m,n}$ and $\mathsf{E}\left[\mathbf{n}(m)\,\mathbf{n}^H(n)\right] = \mathbf{R}_{\mathsf{n}}\delta_{m,n}$, where $\delta_{l,m}$ is the Kronecker delta function. Note that this model analogously encompasses a broad range of system configurations described by the general vector channel model in signal processing and wireless communications. Of particular interest is the special case in which the vector $\mathbf{n}(n)$ can be decomposed into a linear interference contribution and an additive temporally and spatially white noise process as $\mathbf{n}(n) = \sum_{k=1}^{K-1} x_k(n)\,\mathbf{s}_k + \mathbf{v}(n)$, where $x_k(n) \in \mathbb{C}$ and $\mathbf{s}_k \in \mathbb{C}^M$ are, respectively, the information process and effective signature associated with the $k$th interfering signals, and $\mathbf{v}(n) \in \mathbb{C}^M$ is a circularly symmetric complex Gaussian noise vector, with mean zero and variance $\mathsf{E}\left[\mathbf{v}(m)\,\mathbf{v}^H(n)\right] = \sigma_n^2\mathbf{I}_M\delta_{m,n}$. Specifically, under the previous statistical assumptions, the covariance matrix of the array observations takes on the following common structure, namely,

$$\mathbf{R} = \mathsf{E}\left\{(\mathbf{y}-\mathsf{E}\{\mathbf{y}\})(\mathbf{y}-\mathsf{E}\{\mathbf{y}\})^H\right\} = \sigma_x^2\mathbf{s}\mathbf{s}^H + \mathbf{S}_I\mathbf{P}_I\mathbf{S}_I^H + \sigma_n^2\mathbf{I}_M \equiv \mathbf{H}\mathbf{H}^H + \sigma_n^2\mathbf{I}_M, \tag{4.2}$$

where $\mathbf{S}_I = \left[\mathbf{s}_1 \cdots \mathbf{s}_{K-1}\right] \in \mathbb{C}^{M\times K-1}$ and $\{\mathbf{P}_I\}_{i,j} = \mathsf{E}\left[x_j^*(n)\,x_i(n)\right]$. The covariance matrix $\mathbf{R}$ in (4.2) admits a spectral decomposition into mutually orthogonal signal and noise subspaces given by

$$\mathbf{R} = \mathbf{U}_s\mathbf{\Lambda}_s\mathbf{U}_s^H + \sigma_n^2\mathbf{U}_n\mathbf{U}_n^H, \tag{4.3}$$

where $\mathbf{\Lambda}_s \in \mathbb{C}^{K\times K}$ is a diagonal matrix containing the signal eigenspectrum, the range space of $\mathbf{U}_s \in \mathbb{C}^{M\times K}$ is the associated signal subspace and the columns of $\mathbf{U}_n \in \mathbb{C}^{M\times(M-K)}$ span the noise subspace. Indeed, the structure defined by the signal subspace and its orthogonal complement has been efficiently exploited in eigendecomposition methods in spectral analysis and array processing [Tre02, Sto05].

Consider now the following (full-rank) matrix, namely,

$$\mathbf{S}_D = \left[ \mathbf{s} \, \mathbf{R}\mathbf{s} \cdots \, \mathbf{R}^{D-1}\mathbf{s} \right]. \tag{4.4}$$

The columns of the so-called *Krylov matrix* $\mathbf{S}_D$ span a corresponding *Krylov subspace* of rank $D$ (denoted in the sequel by $\mathcal{K}_D(\mathbf{R}, \mathbf{s})$), that will be of special assistance in deriving efficient representations of the solution to the two fundamental array processing problems in the sequel. Note that the following holds in a Krylov subspace expansion, namely,

$$\mathcal{K}_D(\mathbf{R}, \mathbf{s}) \subseteq \mathcal{K}_{D+1}(\mathbf{R}, \mathbf{s}). \tag{4.5}$$

Moreover, since the columns of $\mathbf{S}_D$ are linearly independent and the signature vector $\mathbf{s}$ is orthogonal to the noise subspace, it is clear from (4.5) that $\mathcal{K}_D(\mathbf{R}, \mathbf{s})$ spans the signal subspace for $D \leq K$. In the literature of numerical methods, Krylov subspaces have served as building blocks for subspace iterative algorithms solving large symmetric eigenvalue problems [Saa96, Vor03b], with the aim of approximating Hermitian matrices of particularly high dimension through the identification of the dominant eigensubspaces. Rather than in the approximation problem, the array covariance in (4.2) is exactly expressed as a low-rank matrix plus a shift (hereafter, we assume $K < M$). Thus, this structure can be appropriately exploited by Krylov methods (cf. Section 4.3.1). Alternatively, Krylov methods generalize the standard power iteration by extracting the *best approximation* (in the Euclidean norm) of a certain eigenpair from a Krylov subspace of given rank. This procedure is known as Rayleigh-Ritz projection [Par80], and can in fact be applied to many situations of practical interest, where only some extreme eigenvalues (often at one end of the spectrum) are required (cf. Section 4.3.2).

Before presenting the two applications in this section, the following simply verified invariance properties of Krylov subspaces are in order [Par80]:

(**IP1**) Scaling: $\mathcal{K}_D(\alpha\mathbf{R}, \beta\mathbf{s}) = \mathcal{K}_D(\mathbf{R}, \mathbf{s}), \ \alpha, \beta \in \mathsf{C}, \ \alpha \neq 0, \beta \neq 0.$

(**IP2**) Translation: $\mathcal{K}_D(\mathbf{R} - \mu\mathbf{I}_M, \mathbf{s}) = \mathcal{K}_D(\mathbf{R}, \mathbf{s}), \ \mu \in \mathsf{C}.$

(**IP3**) Similarity: $\mathcal{K}_D\left(\mathbf{Q}^H\mathbf{R}\mathbf{Q}, \mathbf{Q}^H\mathbf{s}\right) = \mathbf{Q}^H\mathcal{K}_D(\mathbf{R}, \mathbf{s}), \mathbf{Q}^H\mathbf{Q} = \mathbf{I}_M.$

### 4.3.1 Minimum variance source power estimation

Consider the problem of estimating the signal waveform of a given source of interest via a linear transformation of the received observations, i.e., $\hat{x}(n) = \mathbf{w}^H\mathbf{y}(n)$. The optimum minimium variance distortionless response[1] (MVDR) filter can be obtained by solving for the following

[1] In signal processing applications, minimum variance unbiased estimators are usually implemented using the linear additive structure of the underlying signal model as minimum variance distortionless response (MVDR) or,

linearly-constrained quadratic optimization problem, namely,

$$\mathbf{w}_{\mathsf{MVDR}} = \arg\min_{\mathbf{w}\in\mathsf{C}^M} \mathbf{w}^H\mathbf{R}\mathbf{w} \text{ subject to } \mathbf{w}^H\mathbf{s} = 1. \quad (4.6)$$

The solution to (4.6) can be easily obtained applying the method of Lagrange multipliers as

$$\mathbf{w}_{\mathsf{MVDR}} = \frac{\mathbf{R}^{-1}\mathbf{s}}{\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}}. \quad (4.7)$$

The optimum MVDR filter is clearly confined to a subspace of dimension $K < M$. Indeed, using (4.3), we have that

$$\mathbf{w}_{\mathsf{MVDR}} = \frac{\mathbf{U}_s\boldsymbol{\Lambda}_s^{-1}\mathbf{U}_s^H\mathbf{s}}{\mathbf{s}^H\mathbf{U}_s\boldsymbol{\Lambda}_s^{-1}\mathbf{U}_s^H\mathbf{s}},$$

Furthermore, from the above definition of source power, an estimate of the SOI power can be obtained as $\mathsf{E}\left[|\hat{x}(n)|^2\right]$. Indeed, the so-called Capon source power estimate is equivalently found as

$$\sigma^2_{\mathsf{CAPON}} = \min_{\mathbf{w}\in\mathsf{C}^M} \mathbf{w}^H\mathbf{R}\mathbf{w} \text{ subject to } \mathbf{w}^H\mathbf{s} = 1, \quad (4.8)$$

which is straightforwardly seen to return a filter solution identical to $\mathbf{w}_{\mathsf{MVDR}}$ and a SOI power approximant given by

$$\sigma^2_{\mathsf{CAPON}} = \frac{1}{\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}}. \quad (4.9)$$

The rationale behind this procedure is that a natural (indirect) solution for the SOI power estimate must be possibly obtained by minimizing the power of the received interference-plus-noise contribution in (5.2) while keeping the intended signal unchanged.

An improved SOI power estimate was alternatively obtained in [McC02] by taking advantage of the signal-plus-noise structure of the array covariance matrix in (4.2) and its subspace decomposition, namely,

$$\sigma^2_{\mathsf{SSMUSIC}} = \frac{1}{\mathbf{s}^H\left(\mathbf{R} - \sigma_n^2\mathbf{I}_M\right)^{\#}\mathbf{s}}. \quad (4.10)$$

We now turn our attention to an alternative representation of the solutions (5.5) and (4.10). As noted above, interestingly enough, the optimum filtering operation can be equivalently formulated as being performed on a projected instance of the received observation onto the range space of the Krylov matrix $\mathbf{S}_D$ in (4.4), with $D \leq K$. Indeed, it can be shown that the filter in (4.7) is confined to a subspace of dimension equal to the number of distinct signal eigenvalues of $\mathbf{R}$. To see this, we shortly recall here two useful definitions from linear algebra[2].

**Definition 2** *Let $\mathbf{R} \in \mathsf{C}^{M\times M}$ denote a Hermitian matrix with eigenvalues $\lambda_1(\mathbf{R}) \leq \lambda_2(\mathbf{R}) \leq \cdots \leq \lambda_M(\mathbf{R})$. The minimum polynomial of $\mathbf{R}$ is the unique monic polynomial $m(\lambda)$ of minimal degree such that $m(\mathbf{R}) = \mathbf{0}_{M\times M}$.*

more generally, linearly constrained minimum variance (LCMV) filters [Sch91, Tre02].

[2] The arguments that follow hold when $\mathsf{R}$ is replaced by $\mathsf{R}_n$ in the definition of $\mathsf{S}_D$, as it can be straighforwardly checked by using the matrix inversion lemma.

The Cayley-Hamilton theorem guarantees that the degree of $m(\lambda)$ is bounded by $M$. We remark that the definition of minimum polynomial applies more generally to square non-diagonalizable matrices, even though we have here restricted ourselves only to matrices admiting a spectral decomposition. In this latter case, the next necessary and sufficient condition holds, namely [Mey00, Chapter 7.11]

$$m(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_L),$$

where $\{\lambda_1, \ldots, \lambda_L\}$, $L \leq M$, is the set of distinct eigenvalues of $\mathbf{R}$.

**Definition 3** *Consider also a vector* $\mathbf{s} \in \mathsf{C}^M$. *Then, the minimum polynomial of* $\mathbf{R}$ *with respect to* $\mathbf{s}$ *is defined to be the monic polynomial* $v(\lambda)$ *of minimal degree such that* $v(\mathbf{R})\mathbf{s} = \mathbf{0}_M$.

It is apparent that $v(\lambda)$ is precisely the right algebraic object describing the properties of Krylov subspace methods. Indeed, according to the definition of the observation covariance matrix in (4.2), note that we may use the spectral factorization theorem in order to state

$$\mathbf{R}^{-1}\mathbf{s} = \sum_{l=1}^{L-1} \lambda_l^{-1}\mathbf{P}_l\mathbf{s},$$

where the matrix $\mathbf{P}_l \in \mathsf{C}^{M\times M}$ is the spectral projector onto the eigenspace associated with the $l$th distinct eigenvalue (observe that the projection space associated with $\mathbf{P}_L$ is exactly the noise subspace, so that we therefore have $\mathbf{P}_L\mathbf{s} = \mathbf{0}_M$). In particular, note that $L-1$ is the degree of the minimum polynomial $v(\lambda)$ related to $\mathcal{K}_D(\mathbf{R}, \mathbf{s})$.

Alternatively, a class of reduced-rank filters can be obtained from the previous projective formulation of the optimal filter on $\mathcal{K}_D(\mathbf{R}, \mathbf{s})$. In particular, for $D < K$, we have $\mathbf{w}_D = \mathbf{S}_D\boldsymbol{\omega}_D$, where

$$\boldsymbol{\omega}_{\mathsf{MVDR}}(D) = \arg\min_{\boldsymbol{\omega}_D \in \mathsf{C}^M} \boldsymbol{\omega}_D^H\mathbf{S}_D^H\mathbf{R}\mathbf{S}_D\boldsymbol{\omega}_D \ \text{subject to}\, \boldsymbol{\omega}_D^H\mathbf{S}_D^H\mathbf{s} = 1, \tag{4.11}$$

with the optimum coefficients linearly describing the reduced-rank MVDR filter on $\mathcal{K}_D(\mathbf{R}, \mathbf{s})$ being given by

$$\boldsymbol{\omega}_{\mathsf{MVDR}}(D) = \frac{\left(\mathbf{S}_D^H\mathbf{R}\mathbf{S}_D\right)^{-1}\mathbf{S}_D^H\mathbf{s}}{\mathbf{s}^H\mathbf{S}_D\left(\mathbf{S}_D^H\mathbf{R}\mathbf{S}_D\right)^{-1}\mathbf{S}_D^H\mathbf{s}}. \tag{4.12}$$

Additionally, the best approximation of the SOI power estimate in (5.5) on $\mathcal{K}_D(\mathbf{R}, \mathbf{s})$ is given by

$$\sigma^2_{\mathsf{CAPON}}(D) = \frac{1}{\mathbf{s}^H\mathbf{S}_D\left(\mathbf{S}_D^H\mathbf{R}\mathbf{S}_D\right)^{-1}\mathbf{S}_D^H\mathbf{s}}. \tag{4.13}$$

In fact, as mentioned above, the reduced-rank MVDR filter defined by (4.12) is the solution extracted from a subspace of rank $D$ that is optimal in the sense of minimizing the mean-square error norm for each increasing dimension starting with the matched filter basis vector and terminating at the (full-rank) MVDR filter in (4.7). The statement follows by a subspace expansion argument after noting the fact in (4.5) and searching for the best approximation in the least-squares norm for a given rank (see [Lue84, Chapter 8][Gol96, Section 9.1.1] for further details). A construction of the reduced-rank SOI power estimate in (4.13) based on a conjugate gradient recursion was proposed in [San03] (see also [Sch02]). Note that the rank $D$ should be large enough to include the information in the observed sample on the SOI parameter, yet small enough to maintain good detection performance in a practical setting where the estimator is constructed from a collection of sample observations (see discussion in Section 4.5). From the discussion above, the reduced-rank version of the filter in (4.7) is given by

$$\mathbf{w}_{\mathsf{MVDR}}\left(D\right) = \frac{\mathbf{S}_D\left(\mathbf{S}_D^H\mathbf{R}\mathbf{S}_D\right)^{-1}\mathbf{S}_D^H\mathbf{s}}{\mathbf{s}^H\mathbf{S}_D\left(\mathbf{S}_D^H\mathbf{R}\mathbf{S}_D\right)^{-1}\mathbf{S}_D^H\mathbf{s}}.$$

Furthermore, the full-rank filter is given by $\mathbf{w}_{\mathsf{MVDR}} = \mathbf{w}_{\mathsf{MVDR}}\left(D\right)$, with $D = L-1$, namely the number of distinct signal eigenvalues.

Moreover, from the shift-invariance property of Krylov subspaces (cf. **IP2**), an equivalent Krylov subspace representation of the improved power estimate in (4.10) can be similarly obtained on $\mathcal{K}_D\left(\mathbf{R},\mathbf{s}\right)$ as

$$\sigma^2_{\mathsf{SSMUSIC}}\left(D\right) = \frac{1}{\mathbf{s}^H\mathbf{S}_D\left(\mathbf{S}_D^H\left(\mathbf{R}-\sigma_n^2\mathbf{I}_M\right)\mathbf{S}\right)^{\#}\mathbf{S}^H\mathbf{s}}. \tag{4.14}$$

Observe that, since the degree of the minimum polynomial of $\mathbf{R}$ wrt. $\mathbf{s}$ is invariant to the shift in $\mathbf{R}-\sigma_n^2\mathbf{I}_M$, it is clear that $\mathcal{K}_D\left(\mathbf{R}-\sigma_n^2\mathbf{I}_M,\mathbf{s}\right) = \mathcal{K}_D\left(\mathbf{H}\mathbf{H}^H,\mathbf{s}\right) = \mathcal{K}_D\left(\mathbf{R},\mathbf{s}\right)$ and the exact Krylov subspace representation is confined to a space of the same dimension $D = L-1$ (see also discussion in [Xu90, Lemma 2]). As before, for $D < K$ we have a Krylov-subspace reduced-rank version of $\sigma^2_{\mathsf{SSMUSIC}}$. Note that the previous equivalent representations of the optimum SOI power estimators avoid both the inversion as well as the eigendecomposition of the covariance matrix, possibly leading to a considerable reduction of the computational complexity.

### 4.3.2 Estimation of the principal eigenspace of the array observations

Many applications in sensor array signal processing rely essentially on the estimation of the principal eigenspectrum of the covariance matrix of the array observations. In fact, rathen than a complete knowledge of the signal and noise eigensubspaces, it is often the case that only some eigenvalues (and the associated eigenspaces) at one of the extremes of the spectrum are of interest. As pointed out above, Krylov subspaces can be utilized to extract arbitrarily good approximations of the principal eigenpair of a structured Hermitian matrix as the one in (4.2). In particular, let $\tilde{\mathbf{S}}_D$ be an *orthonormal* matrix defined as

$$\tilde{\mathbf{S}}_D = \begin{bmatrix}\mathbf{s}_1\,\mathbf{s}_2\cdots\,\mathbf{s}_D\end{bmatrix},$$

where the columns are obtained recursively as

$$\mathbf{s}_{i+1} = \frac{\prod_{k=1}^{i}\left(\mathbf{I}_M - \mathbf{s}_k\mathbf{s}_k^H\right)\mathbf{R}\mathbf{s}_i}{\left\|\prod_{k=1}^{i}\left(\mathbf{I}_M - \mathbf{s}_k\mathbf{s}_k^H\right)\mathbf{R}\mathbf{s}_i\right\|} = \frac{\sum_{k=1}^{i}\left(\mathbf{I}_M - \mathbf{s}_k\mathbf{s}_k^H\right)\mathbf{R}\mathbf{s}_i}{\left\|\sum_{k=1}^{i}\left(\mathbf{I}_M - \mathbf{s}_k\mathbf{s}_k^H\right)\mathbf{R}\mathbf{s}_i\right\|}, \quad i = 1, 2, \ldots \tag{4.15}$$

with $\mathbf{s}_1 = \frac{\mathsf{s}}{\|\mathsf{s}\|}$. Clearly, a simple subspace expansion argument can be used as before in order to show that the columns of $\tilde{\mathbf{S}}_D$ span the subspace $\mathcal{K}_D(\mathbf{R}, \mathbf{s})$. Then, we have the following lemma[3].

**Lemma 4.1** *[Par80] Let $D = K$ and consider the matrix $\mathbf{M} = \tilde{\mathbf{S}}_D^H\mathbf{R}\tilde{\mathbf{S}}_D \in \mathsf{C}^{K\times K}$, with eigenvalues $\lambda_1(\mathbf{M}) \leq \lambda_2(\mathbf{M}) \leq \cdots \leq \lambda_K(\mathbf{M})$, and associated eigenvectors $\{\mathbf{q}_k(\mathbf{M})\}$, $k = 1, \ldots, K$. Moreover, consider the matrix $\mathbf{Z} = \left[\mathbf{z}_1\,\mathbf{z}_2\cdots\mathbf{z}_K\right] = \tilde{\mathbf{S}}_D\mathbf{Q} \in \mathsf{C}^{M\times K}$, with $\mathbf{Q} = \left[\mathbf{q}_1(\mathbf{M})\,\mathbf{q}_2(\mathbf{M})\cdots\mathbf{q}_K(\mathbf{M})\right]$. Then,*

$$\lambda_k(\mathbf{M}) = \lambda_{M-K+k}(\mathbf{R}), \tag{4.16}$$

$$\mathbf{z}_k \equiv \mathbf{q}_{M-K+k}(\mathbf{R}). \tag{4.17}$$

Essentially, the invariance property of Krylov subspaces to unitary transformations (cf. **IP3**) can be used to prove the statement of the lemma as follows. Specifically, since the columns of $\tilde{\mathbf{S}}_D$ are known from the discussion above to span the signal subspace, we clearly have $\mathbf{M} = \tilde{\mathbf{S}}_D^H\mathbf{R}\tilde{\mathbf{S}}_D = \tilde{\mathbf{S}}_D^H\mathbf{U}_s\mathbf{\Lambda}_s\mathbf{U}_s^H\tilde{\mathbf{S}}_D$. Indeed, the previous operation represents a similarity transformation of the $K$-dimensional signal subspace, so that the signal eigenvalues remain invariant in (4.16). Thus, regarding (4.17), the matrix $\tilde{\mathbf{S}}_D\tilde{\mathbf{S}}_D^H$ defines an orthogonal projector that represents a change of basis on the same linear space. In fact, $\mathbf{M}$ is exactly the symmetric Hessenberg matrix returned by the Lanczos tridiagonalization algorithm when used in order to provide an orthonormal basis for $\mathcal{K}_D(\mathbf{R}, \mathbf{s})$ (see e.g. [Gol96]). Lemma 4.1 provides a method to extract the signal eigenvalues and eigenvectors of the observation covariance matrix. In the literature of numerical analysis, the eigenvalues of $\mathbf{M}$ and the columns of $\mathbf{Z}$ are regarded as the Ritz eigenvalues and Ritz eigenvectors, respectively, delivered namely by the Ritz-Rayleigh procedure [Par80], which can be seen as a rank-revealing procedure providing the best approximation to the true eigenpairs whenever $D < K$. Consequently, we can use the Ritz-Rayleigh procedure as a low-complexity method to estimate the principal eigenspace of the array covariance without matrix inversion or eigendecomposition by solely using its positive powers. Moreover, since $\mathbf{R}$ and $-\mathbf{R}$ generate the same subspaces (cf. **IP1**), the left part of the spectrum can be equally well approximated. Accordingly, the previous method can be equivalently used to extract the noise subspace associated with the array covariance matrix, defined as the orthogonal complement of columns space of $\mathbf{Z}$.

In practice, the theoretical covariance matrix defining the Krylov subspace that describes the previous estimation problems is most often not available. Consider for instance the problem

[3] For the sake of notational convenience, we assume without loss of generality that all signal eigenvalues of the covariance matrix have single multiplicity.

of SOI power estimation. In particular, we may write $\sigma^2_{\mathsf{CAPON}}(D) = \left(\mathbf{v}^H\mathbf{B}^{-1}\mathbf{v}\right)^{-1}$, where we have defined $\mathbf{B} = \mathbf{S}_D^H\mathbf{R}\mathbf{S}_D$ and $\mathbf{v} = \mathbf{S}_D^H\mathbf{s}$. On the other hand, note that $[\mathbf{B}]_{i,j} = \mathbf{s}^H\mathbf{R}^{i+j-2}\mathbf{s}$ and $[\mathbf{v}]_i = \mathbf{s}^H\mathbf{R}^{i-1}\mathbf{s}$, for $i,j = 1,\ldots,D$. Indeed, both the SOI power estimate in (4.14) as well as the solution to the principal eigenspace approximation problem in Lemma 4.1 can be equivalently expressed in terms of the vector-valued quadratic forms $\mathbf{s}^H\mathbf{R}^k\mathbf{s}$, $k = 0,\ldots,D-1$. Hence, the problem of implementing in practice the Krylov subspace methods introduced above reduces to the problem of estimating the previous key quantities using a SCM computed from a collection of observed data samples as

$$\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{y}(n)\,\mathbf{y}^H(n)\,. \tag{4.18}$$

Traditionally, implementations of Krylov subspace methods that are based on the direct substitution of the true covariance matrix for the SCM has been unquestionably regarded in the literature as fairly optimal. Indeed, such estimators can be readily shown to be consistent in the classical sense, i.e., they converge stochastically to the true quantity as the sample-size tends to infinity whereas all other dimensions in the signal model remain constant. However, contrary to this conventionally implicit assumption, practical operation conditions are characterized by a sample size and dimension being comparable in magnitude. In these situations, the performance of the previous methods may unavoidably suffer from a considerable degradation. In the next section, we provide a characterization of the performance of sample Krylov-subspace estimates in the low sample-size, relatively high-dimensional observation regime.

## 4.4 Asymptotic Performance Analysis of Sample Estimates

In this section, we derive an asymptotic expression of the estimators in Section 4.3 when constructed using the available SCM under the realistic assumption of an array sample observation of comparably large size and dimension. To that effect, we resort to the theory of the spectral analysis of large-dimensional random matrices, or random matrix theory (RMT). Specifically, we build upon results involving the Stieltjes transform of spectral probability measures. This fundamental building block allows us to characterize the asymptotic eigenspectrum of the SCM in terms of the limiting spectral distribution of the theoretical covariance matrix as all dimensions of the random matrix model increase without bound at a constant rate. For our purposes, not only the asymptotic spectrum but also the limiting behavior of the associated eigensubspaces are of interest. In particular, we obtain the asymptotic limit of the quantities $\mathbf{s}^H\hat{\mathbf{R}}^k\mathbf{s}$, $k = 1,\ldots,D$, describing the sample Krylov-subspace estimates, as both the number of samples $N$ and the observation dimension $M$ increase without bound with $M/N \to c < +\infty$.

From the statistical characterization of the signal model in (5.1), observe that we can statistically model the observed samples as $\mathbf{y}(n) = \mathbf{R}^{1/2}\mathbf{u}(n)$, where $\mathbf{u}(n) \in \mathbb{C}^M$, $n = 1,\ldots,N$, is a collection of i.i.d. random vectors, whose entries have zero mean real and imaginary parts with variance 1/2 and bounded higher moments. Therefore, the SCM in (5.10) can be modeled as

$$\hat{\mathbf{R}} = \frac{1}{N}\mathbf{R}^{1/2}\mathbf{U}\mathbf{U}^{H}\mathbf{R}^{1/2}, \tag{4.19}$$

where the matrix $\mathbf{U} \in \mathbb{C}^{M\times N}$ is constructed using as its columns the vectors $\mathbf{u}(n), n = 1, \ldots, N$. Furthermore, the following assumptions of purely technical interest will be used in our derivations:

(**As1**) The vector $\mathbf{s}$ has uniformly bounded Euclidean norm for all $M$.

(**As2**) The matrix $\mathbf{R}$ has uniformly bounded spectral radius for all $M$.

In the sequel, given two quantities $a, b$, $a \asymp b$ will denote the fact that both quantities are asymptotic equivalents, i.e., $|a-b| \overset{a.s.}{\to} 0$, with $a.s.$ denoting almost sure convergence. Define the $k$th eigenvalue moment of the SCM and the theoretical covarince matrix as $\hat{m}_k \triangleq \frac{1}{N}\operatorname{Tr}\left[\hat{\mathbf{R}}^k\right]$ and $m_k \triangleq \frac{1}{N}\operatorname{Tr}\left[\mathbf{R}^k\right]$, respectively. The following result regarding the asymptotic convergence of the eigenvalue moments of SCM-type matrices will be of interest.

**Lemma 4.2** *(Asymptotic moments of sample covariance matrices with outer correlations) Let $\hat{\mathbf{R}}$ be defined as in (5.11). Then, as $M, N \to \infty$, with $M/N \to c < +\infty$,*

$$\hat{m}_k \asymp \frac{1}{k+1}\sum_{l=1}^{k}\binom{k+1}{l}\sum_{\substack{i_1+\cdots+i_l=k \\ i_1,\ldots,i_l\in\{1,\ldots,k\}}} m_{i_1}\cdots m_{i_l}. \tag{4.20}$$

**Proof.** See Appendix A. ■

The asymptotic eigenvalue moments of SCM-type matrices are also independently studied in [Li01] using the combinatorics of the lattice of non-crossing partitions (cf. Appendix B) in the context of linear multiuser detection, as well as in [Xia05], where the limiting problem is formulated in terms of a combinatorial coloring problem and the results are used to analyze the transient behavior of adaptive least square filters. In appendix A, we provide a more direct and standard derivation solely based on the properties of the Stieltjes transform. Furthermore, we have the following non-trivial generalization involving also the eigensubspaces of the SCM.

**Proposition 4.1** *With the previous definitions and under the assumptions above, as $M, N \to \infty$, with $M/N \to c < +\infty$,*

$$\mathbf{s}^H\hat{\mathbf{R}}^k\mathbf{s} = \sum_{l=1}^{k}\eta(l,k)\;\mathbf{s}^H\mathbf{R}^l\mathbf{s}, \tag{4.21}$$

*where*

$$\eta(l,k) = \sum_{t\in T(l,k)}\binom{l}{l_1, l_2, \ldots, l_k}\hat{m}_1^{l_2}\hat{m}_2^{l_3}\cdots\hat{m}_{k-1}^{l_k},$$

*and $\hat{m}_k$, $k = 1, \ldots, D$, is given in Lemma 4.2 in terms of the eigenvalue spectrum of* $\mathbf{R}$.

**Proof.** See Appendix B. ■

Hence, on the basis of the previous result, we may claim the following

**Corollary 4.1** *Traditional SCM-based estimators of Krylov subspace solutions are not consistent for arbitrarily large-dimensional array observations.*

In fact, the accurate approximation of optimal Krylov subspace solutions based on observation data records of finite size represents the major problem in practical implementations. As mentioned above, under unknown second-order statistics, reduced-rank solutions may indeed outperform the optimum (full-rank) filter. In particular, two considerations happen to prove especially relevant regarding the improvement in estimation quality achieved by a Krylov projection method acting on a subspace of particularly reduced dimension. On the one hand, the selection of a higher rank determines the need to estimating functions of powers of the SCM of higher order, namely having associated an increased variance. On the other hand, a certain degree of robustness against the non-stationarity of the observed process can be expected from low-rank filtering solutions, as the underlying reduced-dimensional Krylov subspace parameterized by the spatial parameters of the SOI may not change so rapidly over the processing interval time [Ge06]. Moreover, adaptive beamforming methods based on Krylov subspaces of reduced-dimension are also shown in [Ge06] to benefit from a further level of robustness to an inaccurate model-order selection.

In order to alleviate the effects of a limited sample-support and comparably large array observation dimension, in the next section we introduce a class of generalized consistent estimators that are strongly consistent for arbitrarily large arrays (or, equivalently, for a limited number of observations per degree-of-freedom).

## 4.5 Consistent estimation under sample-size limitations

In this section, we provide an improved estimator of the key quantities $\mathbf{s}^H\mathbf{R}^k\mathbf{s}$ describing the Krylov subspace methods presented in Section 4.3 that is consistent for a limited number of observations per array element. Moreover, the proposed estimators are consistent under more general conditions than traditional ones, in the sense that they converge to the original spectral function of the true covariance matrix as not only the number of samples $N$ but also the observation dimension $M$ go to infinity at a constant rate, i.e., as $M, N \to \infty$, with $M/N \to c < +\infty$. We will refer to these estimators as $M, N$-consistent as a generalization of traditional $N$-consistent estimators. In particular, we build upon results from Girko's theory of general statistical analysis (GSA) or G-analysis [Gir98], namely providing us with a set of estimators of certain functions of the spectrum of $\mathbf{R}$ in terms of solely the eigenvalues of $\hat{\mathbf{R}}$. The following proposition provides the building blocks of the proposed generalized consistent Krylov subspace estimator.

**Proposition 4.2** *With the previous definitions and under the assumptions above, as $M, N \to \infty$, with $M/N \to c < +\infty$,*

$$\frac{1}{M}\operatorname{Tr}\left[\mathbf{R}^k\right] \asymp \sum_{l=1}^{k} \mu_S(l,k)\,\frac{1}{M}\operatorname{Tr}\left[\hat{\mathbf{R}}^l\right], \tag{4.22}$$

*and*

$$\mathbf{s}^H\mathbf{R}^k\mathbf{s} \asymp \sum_{l=1}^{k} \mu_S(l,k)\,\mathbf{s}^H\hat{\mathbf{R}}^l\mathbf{s}, \tag{4.23}$$

*with*

$$\mu_S(l,k) = (-1)^{k+l}\,\frac{l!}{k!}\sum_{t=(l_1,\ldots,l_k)\in T(l,k)} \xi_P(t)\;\theta_1^{l_1}\theta_2^{l_2}\cdots\theta_k^{l_k},$$

*where $T(l,k)$ is the set of partitions of $k \in \mathsf{N}$ in $l$ parts, and*

$$\xi_P(t) = \frac{l!}{l_1!l_2!\cdots l_k!1!^{l_1}2!^{l_2}\cdots k!^{l_k}},$$

*Furthermore, the quantities $\theta_1, \theta_2, \ldots$, are defined as*

$$\theta_{k+1} = \sum_{l=1}^{k}(-1)^l\,l!\sum_{t\in T(l,k)}\xi_P(t)\;\psi_1^{l_1}\psi_2^{l_2}\cdots\psi_k^{l_k},\quad k=1,2,\ldots,$$

*with $\theta_1 = 1$ and*

$$\psi_k = \sum_{l=1}^{k}(-1)^n\,\frac{(k+1)!}{(k-l+1)!}\sum_{t\in T(l,k)}\binom{l}{l_1,l_2,\ldots,l_l}\;\hat{m}_1^{l_1}\hat{m}_2^{l_2}\cdots\hat{m}_l^{l_l}.$$

**Proof.** See Appendix C. ■

**Corollary 4.2** *For any finite $k = 1, 2, \ldots$, the RHS of (4.23) is a strongly consistent estimator of the scalar quantity $\mathbf{s}^H\mathbf{R}^k\mathbf{s}$.*

As an example, the coefficients defining the estimators in (4.22) and (4.23), for $k = 1, 2, 3$, are given in Table 4.1.

A recurrent formula for the computation of (4.22) and (4.23) is also given in Appendix C. Hence, by replacing in the Krylov subspace solutions described in Section 4.3 the key quantities $\mathbf{s}^H\mathbf{R}^k\mathbf{s}$, $k = 1, 2, \ldots, D$, with the RHS of (4.23), an estimator is obtained that is consistent even for a finite number of samples per filtering degree-of-freedom. Consequently, the proposed estimators generalize the conventional notion of estimation consistency by allowing for arbitrarily high-dimensional observations without approximation performance degradation. Finally, note

Table 4.1: Example estimator coefficients: $k = 1, 2, 3, 4$.

| | | | |
|---|---|---|---|
| $\mu_S(1,1) = 1$ | – | – | – |
| $\mu_S(1,2) = -\hat{m}_1$ | $\mu_S(2,2) = 1$ | – | – |
| $\mu_S(1,3) = 2\hat{m}_1^2 - \hat{m}_2$ | $\mu_S(2,3) = -2\hat{m}_1$ | $\mu_S(3,3) = 1$ | – |
| $\mu_S(1,4) = 5\hat{m}_1^3 + 5\hat{m}_1\hat{m}_2 - \hat{m}_3$ | $\mu_S(2,4) = 5\hat{m}_1^2 - 2\hat{m}_2$ | $\mu_S(3,4) = -3\hat{m}_1$ | $\mu_S(4,4) = 1$ |

that the class of $M, N$-consistent eigenvalue moment estimators in (4.22) is of interest by itself for the signal processing community, since they can be used to improve the performance of reduced-rank MVDR/MMSE filtering schemes in systems with random signatures (see references in [Tul04, Section 3.1.6] and also [Rub06]) as well as for the purpose of eigenspectrum estimation[4].

In the next section, we evaluate the performance of the estimator in (4.23) via numerical simulations in a typical array processing scenario. In particular, the performance of the two applications of Krylov subspace methods in Section 4.3 are evaluated under both proposed and conventional implementations.

## 4.6 Numerical evaluations

In order to numerically evaluate the proposed generalized consistent estimators, we consider a typical array processing scenario described by the observation covariance matrix structure in (4.2). Throughout the simulations, we will denote by *proposed* the implementations relying on our generalized $M, N$-consistent estimator, whereas those based on the $N$-consistent direct substitution of the true covariance matrix for the SCM will be regarded as *conventional*. We assume two sources impinging on a uniform linear array with $M = 30$ sensor elements separated half a wavelength apart. The angles of arrival are 10 and 20 degrees, respectively, and both sources are received with the same power equal to 10dB above the noise floor. First, an improved performance in the estimation of the moments $\mathbf{s}^H\mathbf{R}^k\mathbf{s}$, $k = 3, \ldots, 5$, is demonstrated in Figure 4.1, where an averaged squared error several orders of magnitude smaller for the proposed estimator can be appreciated, especially for low sample-supports.

Figure 4.2 illustrates the performance of both the conventional and proposed implementations of the SOI power estimate in (4.14) (cf. Section 4.3.1). An additional source is assumed to impinge on the array from an angle of 15 degrees and equal power. The full-rank solutions are shown versus different sample sizes, along with the traditional Capon SOI power estimate involving the inversion of the SCM as in (5.5). Again, an improved convergence associated with

[4] Note that the spectrum of a Hermitian positive matrix can also be recovered in terms of (traces of) its powers from the application of Newton's identities [Mey00] to the characteristic polynomial of the matrix.

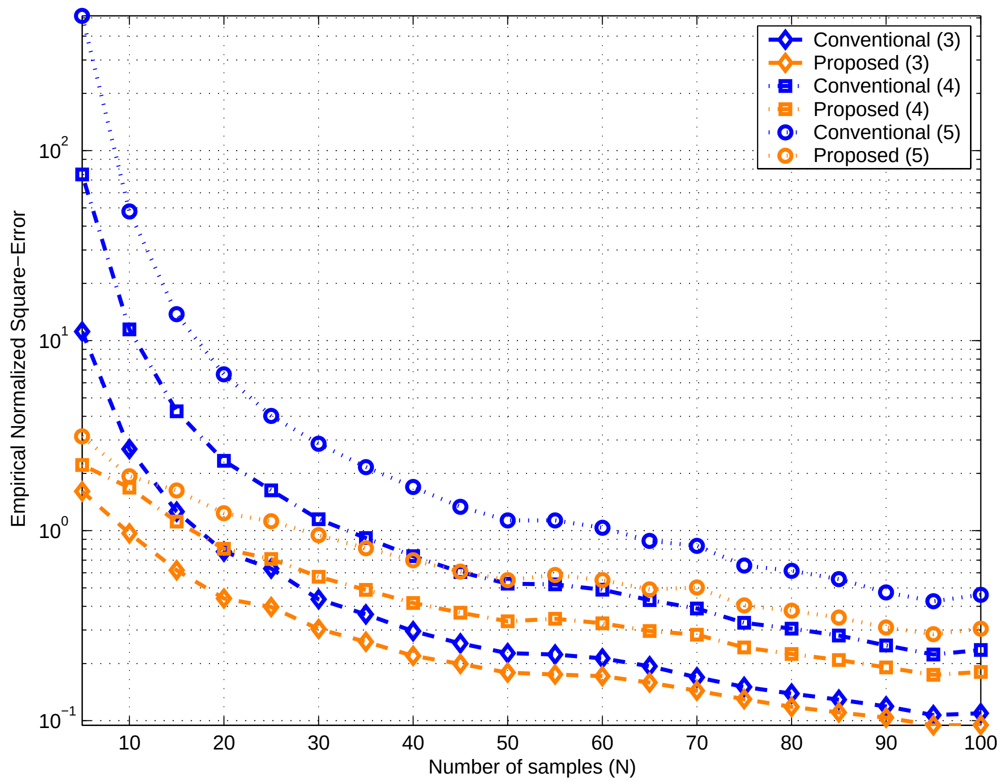


Figure 4.1: Averaged square-error of the estimation of $\mathbf{s}^H\mathbf{R}^k\mathbf{s}$, $k = 3, 4, 5$, (normalized by true moment value) versus number of samples. $K = 2$, $M = 30$.

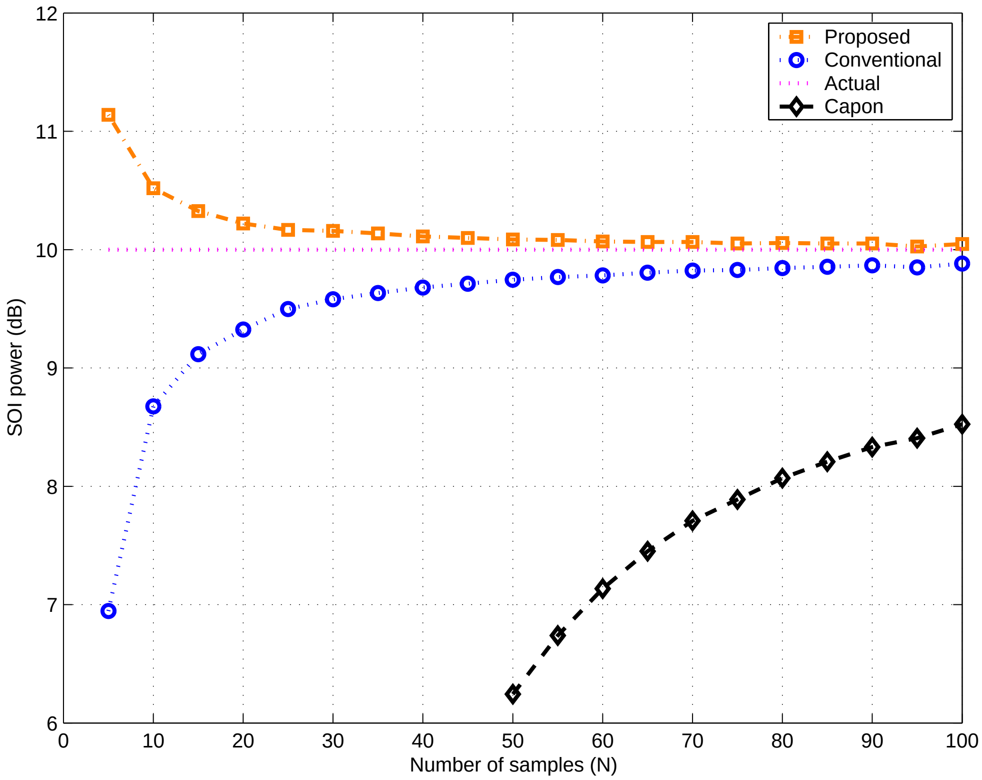


Figure 4.2: Averaged performance of proposed and conventional implementations of SOI power estimate $\sigma^2_{\mathsf{SSMUSIC}}$ versus number of samples. $K = 3$, $M = 30$.

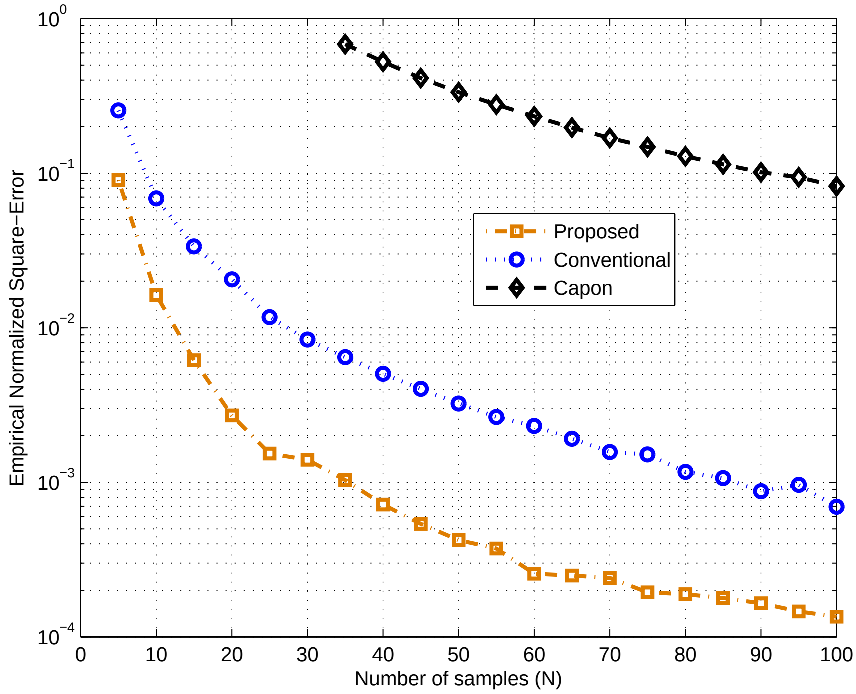


Figure 4.3: Averaged square-error in the estimation of $\sigma^2_{\mathsf{SSMUSIC}}$ (normalized by actual SOI power) versus number of samples. $K = 3$, $M = 30$.

the proposed estimator can be observed. Additionally, the averaged empirical square-norm of the error is shown in Figure 4.3 normalized by the true source power.

Regarding the problem of estimating the covariance principal eigenspace from a limited number of data samples (cf. Section 4.3.2), the histogram obtained from 5000 realizations of both proposed and conventional estimators of the largest eigenvalue of $\mathbf{R}$ obtained from the Rayleigh-Ritz procedure is shown in Figure 4.4. A third estimate obtained from the direct computation of the maximum eigenvalue of $\hat{\mathbf{R}}$ (denoted as *traditional*) is also depicted for the purpose of comparison. A sample-size of $N = M = 30$ is considered. Observe that the conventional and traditional eigenvalue estimators are increasingly biased, although they present a smaller variance. Finally, the averaged empirical square-norm of the estimation error versus an increasing number of samples is illustrated in Figure 4.5 normalized by the actual value of the dominant eigenmode.

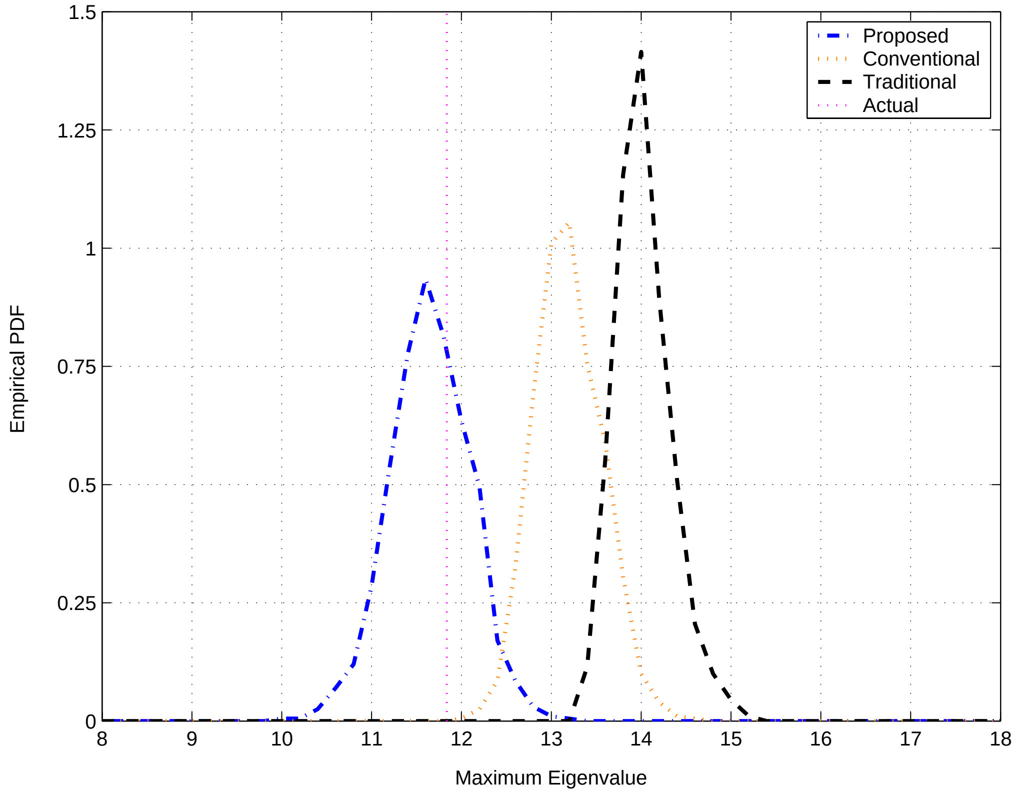


Figure 4.4: Histogram of maximum eigenvalue estimate obtained with different estimators. $K = 2$, $M = 3$.

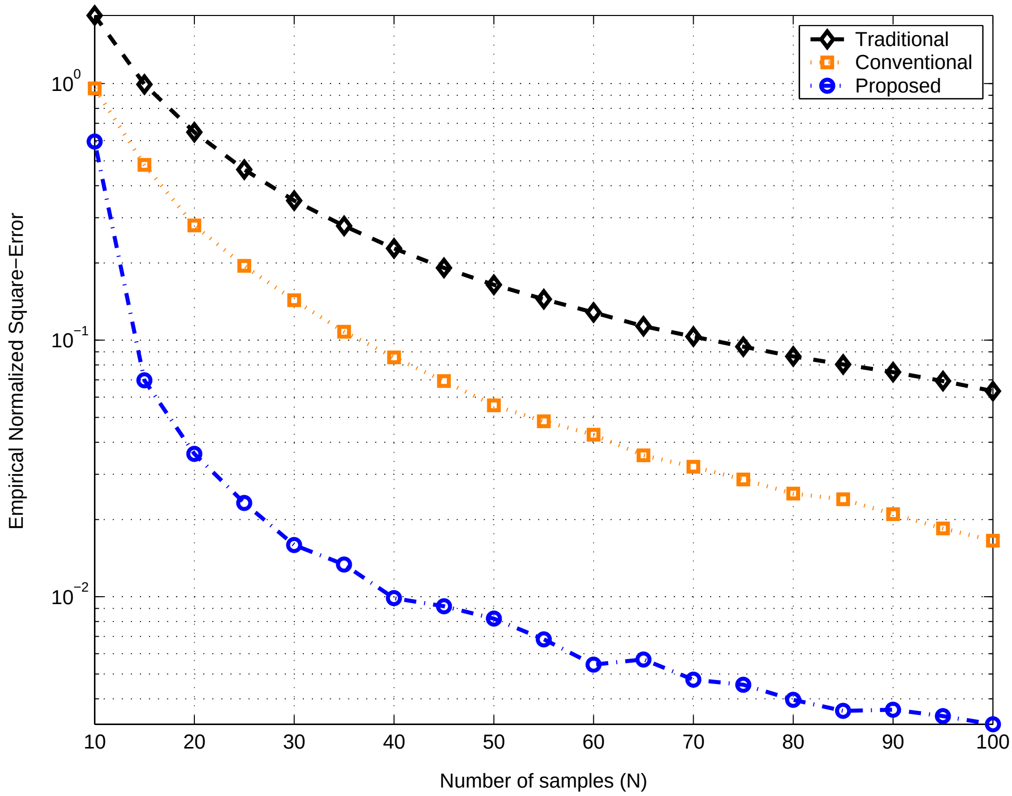


Figure 4.5: Averaged square-error of the maximum eigenvalue estimates (normalized by actual SOI power) versus number of samples. $K = 2$, $M = 30$.

## 4.7 Conclusion

In many parameter estimation problems in statistical signal processing, subspace methods based on the Krylov space defined by the observation covariance matrix and the effective SOI signature vector are known to allow for a reduction of the computational complexity as well as for providing a certain degree of robustness against finite sample-size constraints. In this chapter, we have addressed the statistical inference problem of estimating Krylov subspace solutions based on the unknown second-order statistics of a collection of received observations. The classical asymptotic regime under which consistency of traditionally optimal estimators is guaranteed does not often match realistic filtering operation conditions, namely characterized by a finite sample-size, and a relatively large observation dimension. Based on a more meaningful asymptotic regime, where not only the number of samples but also the observation dimension grow without bound at the same rate, we have shown using RMT that a significantly biased behavior of the conventionally implemented estimators is to be expected in practice. Building upon results from Girko's GSA on the asymptotic limit of certain spectral functions of the SCM in terms of the theoretical covariance, a correction has been afforded by finding a class of generalized consistent estimators of the key quantities $\mathbf{s}^H\mathbf{R}^k\mathbf{s}$. We have also provided a closed-form expression as well as a recurrent formula for the $M, N$-consistent estimators of the eigenvalue moments of the covariance matrix. An improved performance has been demonstrated via the numerical simulation of two typical array processing applications, namely the problem of SOI power estimation and the estimation of the principal eigenspace and dominant eigenmodes of a structured observation covariance matrix. Finally, by not requiring matrix inversion or eigendecomposition, the proposed estimators present a moderate computational complexity, similar to that of the theoretical solution and essentially due to the calculation of some matrix powers of low degree.

## Appendix 4.A Asymptotic eigenvalue moments of SCM-type matrices with outer correlations

We first recall a result regarding the compositional inverse operation of formal power series (see e.g. [Sta97]).

**Theorem 4.1** *(Lagrange inversion formula) Consider the formal power series $f(z) = \sum_{n\geq 1} f_n z^n$, with $f_1 \neq 0$, and let $g(z) = \sum_{n\geq 1} g_n z^n$ be its inverse for composition, i.e., $f(g(z)) = g(f(z)) = z$. Then,*

$$g_n = [z^n]\{g(z)\} = \left[z^{-1}\right]\left\{\frac{1}{nf(z)^n}\right\}, \tag{4.24}$$

*where $\left[z^{-l}\right]$ denotes the operator extracting the coefficient of $z^{-l}$ in a series expansion.*

**Corollary 4.3** *Let $f(z)$ be defined by $f(z) = z\phi(f(z))$, with $\phi(0) \neq 0$. Then,*

$$f_n = \left[z^{-n}\right]\{f(z)\} = \frac{1}{n}\left[z^{n-1}\right]\{\phi(z)^n\}. \tag{4.25}$$

A fundamental result in RMT (see [Sil95a]) establishes the weak convergence of the empirical eigenvalue distribution function of $\hat{\mathbf{R}}$ towards a limiting nonrandom distribution function with compactly supported density, as $M/N \to \infty$, $M/N \to c < +\infty$. The convergence is given in terms of the Stieltjes transform, defined for a probability distribution function $F(\lambda)$ on $\mathbb{C}$ as

$$S_F(z) = \int \frac{1}{\lambda - z} dF(\lambda), \tag{4.26}$$

and being analytic on $\mathbb{C} \setminus \operatorname{supp}(dF(\lambda))$, where $\operatorname{supp}(\cdot)$ denotes the density (compact) support. For our purposes, it will be of interest to define the matrix $\mathbf{B} = \boldsymbol{\Xi}^H \mathbf{R} \boldsymbol{\Xi} \in \mathbb{C}^{N\times N}$, where $\boldsymbol{\Xi} = \sqrt{N}\mathbf{U} \in \mathbb{C}^{M\times N}$. In particular, it is proved in [Sil95c] that the Stieltjes transform of the empirical distribution function of the eigenvalues of $\mathbf{B}$ can be obtained as the solution of the following functional equation, namely,

$$S = -\left(z - c\int \frac{\lambda dH(\lambda)}{1 + \lambda S}\right)^{-1}, \tag{4.27}$$

in the sense that for every $z \in \mathbb{C}^+$, $S$ is the unique solution in $\mathbb{C}^+$ to (4.27), and where $H(\lambda)$ is the (nonrandom) limiting empirical distribution function of the eigenvalues of $\mathbf{R}$. Note that the spectra of $\hat{\mathbf{R}}$ and $\mathbf{B}$ differ by $|M - N|$ zero eigenvalues. Thus, if $F(\lambda)$ and $G(\lambda)$ are the limiting spectral distribution functions of $\hat{\mathbf{R}}$ and $\mathbf{B}$, respectively, it can be stated that $G(\lambda) = (1 - c)\mathcal{I}_{[0,\infty)} + cF(\lambda)$, where $\mathcal{I}_\Omega$ denotes the indicator function over the set $\Omega$. As it is well-known, the Stieltjes transform can be regarded as an eigenvalue-moment generating function (see e.g. [Tul04, Section 2.2.1]). Indeed, since

$$\frac{1}{z - \lambda} = \frac{1}{z}\sum_{k=0}^{\infty}\left(\frac{\lambda}{z}\right)^k,$$

by the linearity of the integral, we get

$$S_G(z) = -\sum_{k=0}^{\infty} \frac{1}{z^{k+1}} \int \lambda^k dG(\lambda),$$

which is the Laurent series expansion of the integrand (or Taylor series about $\infty$), and where $S_G(z)$ is the Stieltjes transform of the probability distribution function $G(\lambda)$. Note that due to the compactness assumption on the eigenvalue density, the power series expansion of the Stieltjes transform is guaranteed to be holomorphic at infinity. Thus, using the fact that

$$\hat{m}_k = \frac{1}{N} \operatorname{Tr}\left[\hat{\mathbf{R}}^k\right] = \frac{1}{N} \operatorname{Tr}\left[\left(\mathbf{R}^{1/2}\mathbf{\Xi}\mathbf{\Xi}^H\mathbf{R}^{1/2}\right)^k\right] = \frac{1}{N} \operatorname{Tr}\left[\mathbf{B}^k\right], \tag{4.28}$$

i.e., $\hat{m}_k \asymp \int \lambda^k dG(\lambda)$, with some abuse of notation, we have

$$S_G(z) = -\frac{1}{z} - \sum_{k=1}^{\infty} \frac{\hat{m}_k}{z^{k+1}}. \tag{4.29}$$

Then, from the characterization of the Stieltjes transform of $G(\lambda)$ in (4.29) as an eigenvalue-moment generating function, and its limiting solution in (4.27) in terms of $H(\lambda)$, we can obtain an asymptotic expression for the moments $\hat{m}_k$ as a function of the spectrum of the true covariance matrix. To that effect, let us define

$$\eta(z) = -S_G\left(z^{-1}\right) = \sum_{k=0}^{\infty} \hat{m}_k z^{k+1} = \sum_{k=1}^{\infty} \hat{m}_{k-1} z^k = \sum_{k=1}^{\infty} u_k z^k, \tag{4.30}$$

where we have defined $u_k = \hat{m}_{k-1}$. Using the definition of $\eta(z)$, we observe that

$$z = \frac{\eta(z)}{1 + c\int \frac{\lambda\eta(z)}{1-\lambda\eta(z)} dH(\lambda)} = \frac{\eta(z)}{\phi(\eta(z))},$$

where

$$\phi(x) = 1 + c\int \frac{\lambda x}{1-\lambda x} dH(\lambda).$$

Furthermore, note that, since $\phi(0) \neq 0$, we may apply the Lagrange inversion formula in Corollary 4.3 to obtain the coefficients $u_k$ in (4.30) as

$$u_k = \frac{1}{k}\left[x^{k-1}\right]\phi(x)^k. \tag{4.31}$$

Hence, we get

$$\begin{aligned}
\phi(x) &= 1 - c + c\int \frac{1}{1-\lambda x} dH(\lambda) \\
&= 1 - c + c\int \sum_{k=0}^{\infty} (\lambda x)^i \, dH(\lambda) \\
&= 1 - c + \sum_{k=0}^{\infty} m_i x^i,
\end{aligned}$$

where we have used the fact that $\int dH(\lambda) = \frac{1}{M}\operatorname{Tr}\left[\mathbf{R}^0\right] = 1$. Since $\phi(x)$ consists of positive powers only, we can express (4.31) as

$$u_k = \frac{1}{k}\left[\frac{\partial}{\partial x^{k-1}}\left\{\phi^k(x)\right\}\right]_{x=0}, \tag{4.32}$$

with $u_1 = \phi(0) = 1$.

In order to obtain derivatives of higher order, we can use the binomial formula to write

$$\begin{aligned}
\left(1 + c\int \frac{\lambda x}{1-\lambda x} dH(\lambda)\right)^k &= \sum_{j=0}^{k}\binom{k}{j}\left(\int \frac{\lambda x}{1-\lambda x} dH(\lambda)\right)^j \\
&= 1 + \sum_{j=1}^{k}\binom{k}{j}\left(\int \frac{\lambda x}{1-\lambda x} dH(\lambda)\right)^j \\
&= 1 + \sum_{j=1}^{k}\binom{k}{j}\left(\int \frac{1}{1-\lambda x} dH(\lambda) - 1\right)^j \\
&= 1 + \sum_{j=1}^{k}\binom{k}{j}\left(\sum_{i\geq 0} x^i m_i\right)^j \\
&= 1 + \sum_{j=1}^{k}\binom{k}{j}\sum_{i_1\geq 1}\cdots\sum_{i_j\geq 1} x^{i_1+\cdots+i_j} m_{i_1}\cdots m_{i_j}
\end{aligned}$$

According to (4.32), we are particularly interested in the term $x^{k-1}$, which is the term such that $i_1 + \cdots + i_k = k-1$. Consequently, we can write

$$u_k = \frac{1}{k}\sum_{j=1}^{k}\binom{k}{j}\sum_{\substack{i_1+\cdots+i_j=k-1 \\ i_1,\ldots,i_j\in\{1,\ldots,k\}}} m_{i_1}\cdots m_{i_j}.$$

Note that only the first $k-1$ terms of the binomial summation contribute effectively, i.e.,

$$u_k = \frac{1}{k}\sum_{j=1}^{k-1}\binom{k}{j}\sum_{\substack{i_1+\cdots+i_j=k-1 \\ i_1,\ldots,i_j\in\{1,\ldots,k\}}} m_{i_1}\cdots m_{i_j}.$$

Hence, we finally find that

$$\begin{aligned}\hat{m}_k &= u_{k+1}\\ &= \frac{1}{k+1}\sum_{j=1}^{k}\binom{k+1}{j}\sum_{\substack{i_1+\cdots+i_j=k\\ i_1,\ldots,i_j\in\{1,\ldots,k\}}} m_{i_1}\cdots m_{i_j}\\ &= \frac{1}{k}\sum_{j=1}^{k}\binom{k}{j}\frac{k}{k+1-j}\sum_{\substack{i_1+\cdots+i_j=k\\ i_1,\ldots,i_j\in\{1,\ldots,k\}}} m_{i_1}\cdots m_{i_j}\\ &= \sum_{j=1}^{k}\binom{k}{j}\frac{1}{k+1-j}\sum_{\substack{i_1+\cdots+i_j=k\\ i_1,\ldots,i_j\in\{1,\ldots,k\}}} m_{i_1}\cdots m_{i_j}.\end{aligned}$$

### Recursive formula

Define further

$$\theta_k(j) = \sum_{\substack{i_1+\cdots+i_j=k\\ i_1,\ldots,i_j\in\{1,\ldots,k\}}} m_{i_1}\cdots m_{i_j}, \tag{4.33}$$

such that

$$\hat{m}_k = \sum_{j=1}^{k}\binom{k}{j}\frac{1}{k+1-j}\theta_k(j).$$

In order to derive a recursive formula for $\theta_k(j)$, we will use the relation between the moments of a probability measure and its Boolean cumulants [Spe95]. As classical cumulants linearize the convolution of probability measures, Boolean cumulants linearize the Boolean convolution as defined in [Spe95]. Let $m_k$ be the moments of a probability measure with distribution function $Q(\lambda)$ and Stieltjes transform $S_Q(z)$. For such a probability measure, the Boolean cumulants are defined as the coefficients of the following series, namely,

$$B_Q(z) = -\frac{1}{S_Q(z)} = z - \sum_{k=1}^{\infty} b_k z^{1-k}.$$

As for the classical and non-crossing cumulants, the moments can be written as a polynomial in terms of the Boolean cumulants. In particular, the coefficient for $b_1^{l_1}b_2^{l_2}\cdots b_k^{l_k}$ in $m_k$ is equal to the multinomial coefficient. Indeed, the combinatorial framework described in Appendices B and C applies also in the case of the Boolean cumulants, being the underlying enumerative structure governed by the lattice of interval partitions. Moreover, this relation can be inverted as

$$b_k = \sum_{j=1}^{k}(-1)^{j+1}\theta_k(j). \tag{4.34}$$

(A straightforward inversion is possible on the lattice of interval partitions, since it is isomorphic to the Boolean lattice of all subsets with Möbius function equal to a sign [Cam94].) More interestingly, the following recurrence holds

$$m_k = \sum_{i=1}^{k} b_i m_{k-1}, \tag{4.35}$$

where $b_1 = m_1$ and $b_0 = 0$. From (4.34) and (4.35), the next recursive formula can be found

$$\begin{aligned} \theta_k(j) &= \sum_{i=1}^{k-j+1} m_i \theta_{k-i}(j-1) \\ \theta_k(1) &= m_k. \end{aligned} \tag{4.36}$$

**Corollary 4.4** *(Eigenvalue moments of Marchenko-Pastur law) As a special case, when $\mathbf{R} = \mathbf{I}_M$ the classical formula for the moments of the Marçenko-Pastur distribution is obtained. Indeed, using*

$$m_k = c \int \lambda^k dH(\lambda) = c, \; k = 0, 1, 2, \ldots, \tag{4.37}$$

*we have*

$$\theta_k(j) = \sum_{\substack{i_1+\cdots+i_j=k \\ i_1,\ldots,i_j \in \{1,\ldots,k\}}} m_{i_1} \cdots m_{i_j} = \binom{k-1}{j-1} c^j, \tag{4.38}$$

*which is justified as follows. Each term in the sum above can be characterized by a string of the form*

$$\underbrace{11\cdots1}_{i_1}0\underbrace{11\cdots1}_{i_2}01\cdots10\underbrace{11\cdots1}_{i_n}, \tag{4.39}$$

*whre the number of ones is $n$ and the number of zeros is $k-1$. Thus, the number of terms that contribute effectively to the sum will be given by the number of possible sortings of the $k-1$ zeros into the $n-1$ positions available. In conclusion,*

$$\hat{m}_k = \frac{1}{k} \sum_{j=1}^{k} \binom{k}{j} c^j \frac{k}{k+1-j} \binom{k-1}{j-1} = \frac{1}{k} \sum_{j=1}^{k} \binom{k}{j} \binom{k}{j-1} c^j, \tag{4.40}$$

*which is the classical formula for the kth moment of the Marchenko-Pastur distribution [Ora97].*

### Checking with combinatorial moment-cumulant formula

In the following, we establish the connection between the formula (4.20) in Proposition 4.2 and the asymptotic expression obtained in the combinatorial framework presented in Appendix B (as alternatively derived in [Li01]).

To that effect, it is enough to evaluate the polynomial in the non-crossing cumulants (or equivalently the eigenvalue moments of $\mathbf{R}$) as indeterminates. In particular, using (4.53), we can write

$$\hat{m}_k = \sum_{l=1}^{k} N_{dis}(l,k) \sum_{\substack{i_1+\cdots+i_l=k \\ 1\le i_1\le\cdots\le i_l\le k}} m_{i_1} m_{i_2} \cdots m_{i_k}, \tag{4.41}$$

where the second summation enumerates the blocks $\{i_1, \ldots, i_k\}$ constituting the (non-crossing) partition. (Note that $i_i = 0$ makes no sense in this setup as it represents an empty block that has no meaning in the definition of the partition type, whereas $i_i \le n$ necessarily, since no block in the partition may certainly have cardinality larger than $n$.) In (4.41), the index set $\{i_1, \ldots, i_k\}$ is equipped with a partial order relation in order to avoid taking into account combinations of the indexes that are equivalent by permutation. If this condition is relaxed, we have

$$\hat{m}_k = \sum_{l=1}^{k} \frac{N_{dis}(l,k)}{k!} \sum_{i_1+\cdots+i_l=k} m_{i_1} m_{i_2} \cdots m_{i_k}.$$

Now, looking at the coefficients of the summation over $k$, observe that

$$\frac{N_{dis}(l,k)}{k!} = \frac{k!}{(k-l+1)!l!} = \frac{1}{k+1}\frac{(k+1)\,k!}{(k+1-l)!l!} = \frac{1}{k+1}\binom{k+1}{l} = \binom{k}{l}\frac{1}{k+1-l}.$$

This is exactly the coefficient of $\theta_k(j)$ in (4.33), since all partitions of equal type by permutation are being considered in the enumeration of the $k$ partitions.

## Appendix 4.B Proposition 4.1

We first introduce some useful concepts regarding the combinatorics of set partitions that will be of notational convenience throughout the appendix.

**Definition 4** *Let $l, k \in \mathsf{N}$, such that $l \le k$. We define $T(l,k)$ as the set of $k$-tuples $(l_1, \ldots, l_k)$ satisfying $1l_1 + 2l_2 + \ldots + kl_k = k$ with $l = l_1 + \ldots + l_k$.*

Additionally, for any $t = (l_1, \ldots, l_k) \in T(l,k)$, we define

$$\xi_P(t) = \frac{l!}{l_1!l_2!\cdots l_k!1!^{l_1}2!^{l_2}\cdots k!^{l_k}}, \quad \xi_{NC}(t) = \frac{k!}{(k-l+1)!}\frac{1}{l_1!l_2!\cdots l_k!}.$$

With this notation, the higher-order derivatives of a composite function can be written in a compact form as

$$\frac{\partial^n}{\partial x^n}\{f(g(x))\} = \sum_{k=1}^{n} f^{(k)}(g(x)) \sum_{t\in T(n,k)} \xi_P(t) \prod_{i=1}^{n} \left(g^{(i)}(x)\right)^{m_i}. \tag{4.42}$$

Furthermore, let us recall the following result giving the coefficients of a composition of two formal power series in terms of the coefficients of the two series (see e.g. [Sta97]).

**Lemma 4.3** *Define the formal power series* $f(z) = 1 + \sum_{n=1}^{\infty} \frac{f_n}{n!} z^n$, $g(z) = \sum_{n=1}^{\infty} \frac{g_n}{n!} z^n$ *and* $h(z) = 1 + \sum_{n=1}^{\infty} \frac{h_n}{n!} z^n$, *such that* $f(g(z)) = h(z)$. *Then, the* $k$*th coefficient of the formal power series composition* $h(z)$ *can be obtained as*

$$h_k = \sum_{l=1}^{k} f_l \sum_{t \in T(l,k)} \xi_P(t)\ g_1^{l_1} g_2^{l_2} \cdots g_k^{l_k}. \tag{4.43}$$

We depart from the result stated in Proposition 4.2, where the following $M, N$-estimators are provided, namely,

$$\frac{1}{M} \operatorname{Tr}\left[\mathbf{R}^k\right] \asymp \sum_{l=1}^{k} \mu_S(l,k) \frac{1}{M} \operatorname{Tr}\left[\hat{\mathbf{R}}^l\right], \quad \mathbf{s}^H \mathbf{R}^k \mathbf{s} \asymp \sum_{l=1}^{k} \mu_S(l,k)\, \mathbf{s}^H \hat{\mathbf{R}}^l \mathbf{s},\ k = 0,1,2,\ldots, \tag{4.44}$$

where the coefficients $\mu_S(l,k)$ are given in terms of the eigenvalue-moments of the SCM. The relation in (4.44) can be straightforwardly inverted as

$$\frac{1}{M} \operatorname{Tr}\left[\hat{\mathbf{R}}^k\right] \asymp \sum_{l=1}^{k} \mu_T(l,k) \frac{1}{M} \operatorname{Tr}\left[\mathbf{R}^l\right], \quad \mathbf{s}^H \hat{\mathbf{R}}^k \mathbf{s} \asymp \sum_{l=1}^{k} \mu_T(l,k)\, \mathbf{s}^H \mathbf{R}^l \mathbf{s},\ k = 0,1,2,\ldots, \tag{4.45}$$

for a set of coefficients $\mu_T(l,k)$ depending on the moments of the theoretical covariance matrix. The asymptotic limit of $\frac{1}{M} \operatorname{Tr}\left[\hat{\mathbf{R}}^k\right]$ is provided by Lemma 4.2, since, clearly, $\frac{1}{M} \operatorname{Tr}\left[\hat{\mathbf{R}}^k\right] = c^{-1} \hat{m}_k$. However, the expression in (4.20) does not help us to find the coefficients $\mu_T(l,k)$ describing the asymptotic convergence of the eigenvalue moments of the SCM, as well as, accordingly, the limit of $\mathbf{s}^H \hat{\mathbf{R}}^k \mathbf{s}$, $k = 0,1,2,\ldots$, in (4.45). Thus, we provide an alternative representation of the (unique) limit in (4.20) revealing the structure of the expressions in (4.45).

Consider the formal power series expansion in (4.29). In order to avoid dealing with cumbersome negative signs, we define $G(z) = -S_G(z)$. The Laurent series $G(z)$ has an inverse for composition $K(z)$, with an expansion [Bia03]

$$K(z) = \frac{1}{z} + \sum_{k=1}^{\infty} \hat{r}_k\ z^{k-1}. \tag{4.46}$$

Analogously to the classical moment-cumulant problem, the coefficients $\hat{r}_k$ are usually regarded as the cumulants associated with the spectral distribution of the SCM. In the problem at hand, the cumulants turn out to be of special interest due to its relation with the eigenvalue moments of the theoretical covariance matrix. In particular, it can be shown that $\hat{r}_k \asymp m_k$ (cf. Appendix B and also [Li01]). Moreover, using the fact that $G(K(z)) = z$, the eigenvalue-moments of the SCM can be obtained using the Lagrange inversion formula in Theorem 4.1 in terms of the eigenvalue-cumulants, i.e.,

$$\hat{m}_k = \frac{1}{k+1} \left[z^{-1}\right] K(z)^{k+1}.$$

In particular, for $t = (l_1, l_2, \ldots, l_k) \in T(l,k)$, the coefficient of $\hat{r}_1^{l_1}\hat{r}_2^{l_2}\cdots\hat{r}_k^{l_k}$ in $\hat{m}_k$ is equal to $\xi_{NC}(t)$. Hence, we can write

$$\hat{m}_k = \sum_{l=1}^{k} \sum_{t\in T(l,k)} \xi_{NC}(t)\ \hat{r}_1^{l_1}\hat{r}_2^{l_2}\cdots\hat{r}_k^{l_k}. \tag{4.47}$$

A unifying combinatorial interpretation of the functional relation between cumulants and moments is afforded by Speicher's work [Nic06], founded on the theory of lattices of non-crossing partitions. Indeed, as the relation between classical moments and cumulants in probability theory happens to be ruled by the lattice of all partitions, the structure of (4.47) is governed by the lattice of non-crossing partitions.

Let us now define $M(z)$ and $C(z)$ as the following two formal power series, namely,

$$M(z) = 1 + \sum_{n=1}^{\infty} \hat{m}_n\ z^n, \tag{4.48}$$

and

$$C(z) = 1 + \sum_{n=1}^{\infty} \hat{r}_n\ z^n. \tag{4.49}$$

From (4.48) and (4.49), we have that $M(z) = \frac{1}{z}G\left(\frac{1}{z}\right)$ and $C(z) = zK(z)$. Now, we observe that $C(zM(z)) = M(z)$. To see that, define $\omega = \frac{1}{z}$ and note that

$$C(zM(z)) = C\left(\frac{1}{\omega}M\left(\frac{1}{\omega}\right)\right) = C(G(\omega)) = G(\omega)\,K(G(\omega)) = M\left(\frac{1}{\omega}\right) = M(z).$$

Furthermore, define

$$\tilde{M}(z) = zM(z) = \sum_{n=1}^{\infty} \tilde{m}_n\ z^n,$$

with $\tilde{m}_n = \hat{m}_{n-1}$ and $\tilde{m}_0 = 1$. Then, since $\tilde{M}(z)$ has no constant term, we can use the formula in (B.9) for the composition of two formal power series in order to identify

$$\hat{m}_k = \sum_{l=1}^{k} \eta(l,k)\ \hat{r}_l, \tag{4.50}$$

with

$$\eta(l,k) = \sum_{t\in T(l,k)} \binom{l}{l_1, l_2, \ldots, l_k} \hat{m}_1^{l_2}\hat{m}_2^{l_3}\cdots\hat{m}_{k-1}^{l_k}.$$

An alternative expression for the coefficients $\eta(l,k)$ in (4.50) is given in [Nic06], where a recurrent version of the moment-cumulant formula in (4.47) is obtained using specific properties of the lattice of non-crossing partitions, as

$$\eta(l,k) = \sum_{\substack{i_1,i_2,\ldots,i_l\in\{0,1,\ldots,k-l\}\\ i_1+i_2+\cdots+i_l=k-l}} \hat{m}_{i_1}\hat{m}_{i_2}\cdots\hat{m}_{i_l}. \tag{4.51}$$

Using the fact that $\hat{r}_k \asymp m_k$, we readily have that $\mu_T(l,k) = \eta(l,k)$.

# Appendix 4.C Proof of Proposition 4.2

In order to obtain an $M,N$-estimator in (4.22), we can directly exploit the inverse functional relation between the cumulants and the moments of the spectral distribution of the SCM presented in Appendix B. Indeed, using the fact that $K(G(z)) = z$, we have

$$\hat{r}_k = -\frac{1}{k-1}\left[z^{-1}\right] G(z)^{-k+1}.$$

Specifically, by recovering the coefficients of $\hat{m}_1^{l_1}\hat{m}_2^{l_2}\cdots\hat{m}_k^{l_k}$ in $\hat{r}_k$ for all $t = (l_1, l_2, \ldots, l_k) \in T(l,k)$, we get

$$\hat{r}_k = \sum_{l=1}^{k} (-1)^{1+l}\frac{(k-2+l)!}{(k-1)!}\sum_{t\in T(l,k)}\frac{1}{l_1!l_2!\cdots l_k!}\,\hat{m}_1^{l_1}\hat{m}_2^{l_2}\cdots\hat{m}_k^{l_k}. \tag{4.52}$$

Speicher's framework introduced in Appendix B allows for a combinatorial interpretation of (4.52) as the explicit inversion of the relation in (4.47), which can be systematically obtained using the Möbius inversion formula (see e.g. [Sta97]) on the lattice of non-crossing partitions. In particular, the Möbius function of the lattice of non-crossing partitions is given by [Nic06]

$$\alpha(t) = s_1^{l_1}\cdots s_k^{l_k},$$

with $s_{n+1} = (-1)^n C_n$ being the $n$th (signed) Catalan number. (Note that a similar procedure applies for the classical moment-cumulant formula if we replace $\xi_{NC}(t)$ with $\xi_P(t)$ and use the Möbius function of the lattice of all partitions [Sta97].) Indeed, it can be readily checked that

$$\alpha(t) \equiv (-1)^{1+l}\frac{(k-2+l)!}{l_1!l_2!\cdots l_k!\,(k-1)!}.$$

Interestingly enough, the spectral cumulants of the SCM can be proved to be equivalent to

$$\hat{r}_k \asymp m_k. \tag{4.53}$$

To see that, we may use the formula in (4.20) for the asymptotic expression of $\hat{m}_k$ in terms of the eigenvalue moments of $\mathbf{R}$ provided by Lemma 4.2 in order to replace the moments $\hat{m}_k$ in (4.52). The result follows from an induction argument. Hence, using (4.53) in (4.52), we directly obtain an $M,N$-estimator of $\frac{1}{M}\operatorname{Tr}\left[\mathbf{R}^k\right]$.

In order to find an expression for the coefficients $\mu_S(l,k)$ defining not only the $M,N$-estimator of $\frac{1}{M}\operatorname{Tr}\left[\mathbf{R}^k\right]$ but also the estimator of the vector-valued quadratic forms $\mathbf{s}^H\mathbf{R}^k\mathbf{s}$, we follow the approach by Girko in [Gir98, Chapter 14]. First, note that we can express the $k$th eigenvalue moment as

$$\frac{1}{M}\operatorname{Tr}\left[\mathbf{R}^k\right] = (-1)^k\,k!^{-1}\frac{\partial^k}{\partial x^k}\frac{1}{M}\operatorname{Tr}\left[\left(\mathbf{I}_M + x\mathbf{R}\right)^{-1}\right]\Bigg|_{x=0}. \tag{4.54}$$

Girko's approach consists of replacing the nonrandom quantity $\frac{1}{M}\operatorname{Tr}\left[\left(\mathbf{I}_M + x\mathbf{R}\right)^{-1}\right]$ in (4.54), namely the so-called real Stieltjes transform of $\mathbf{R}$ in the framework of GSA (also known as the $\eta$-transform in the engineering literature [Tul04]) by its $M,N$-consistent estimator in terms of $\hat{\mathbf{R}}$. Thereafter, the expression is differentiated as above in order to obtain the estimator of the eigenvalue moment. G-analysis provides us with an $M,N$-consistent estimator of the real-valued Stieltjes transform of $\mathbf{R}$, namely the $G_2$-estimator, given by (cf. Section 14.2 in [Gir98])

$$G_2 = \frac{1}{M}\operatorname{Tr}\left[\left(\mathbf{I}_M + \theta\left(x\right)\hat{\mathbf{R}}\right)^{-1}\right] \asymp \frac{1}{M}\operatorname{Tr}\left[\left(\mathbf{I}_M + x\mathbf{R}\right)^{-1}\right],$$

with $\theta\left(x\right)$ being the positive solution of the following canonical equation, namely,

$$\theta\left(x\right) h\left(\theta\left(x\right)\right) = x, \tag{4.55}$$

where

$$h\left(\theta\right) = 1 - c + \frac{1}{N}\operatorname{Tr}\left[\left(\mathbf{I}_M + \theta\hat{\mathbf{R}}\right)^{-1}\right].$$

Only a (trivial) expression for the estimators of the first two moments are given by Girko in [Gir98]. In the following, we build upon these ideas and derive an explicit formula for moments of arbitrary order. Moreover, we extend the previous approach in order to equivalently prove (4.23).

For our purposes we can make use of the following identity, namely,

$$\mathbf{s}^H\mathbf{R}^k\mathbf{s} = \left(-1\right)^k k!^{-1}\frac{\partial^k}{\partial x^k}\,\mathbf{s}^H\left(\mathbf{I}_M + x\mathbf{R}\right)^{-1}\mathbf{s}\Big|_{x=0}, \tag{4.56}$$

where now we can use the $M,N$-consistent estimator for the quadratic forms of resolvents of covariance matrices (cf. Section 14.25 in [Gir98])

$$\mathbf{s}^H\left(\mathbf{I}_M + x\mathbf{R}\right)^{-1}\mathbf{s} \asymp \mathbf{s}^H\left(\mathbf{I}_M + \theta\left(x\right)\hat{\mathbf{R}}\right)^{-1}\mathbf{s}. \tag{4.57}$$

The RHS of (5.28) is regarded in the GSA literature as $G_{25}$ estimator. Thus, from (4.56) and (B.4), an expression for the estimator of $\mathbf{s}^H\mathbf{R}^k\mathbf{s}$ is obtained as

$$\mathbf{s}^H\mathbf{R}^k\mathbf{s} \asymp \sum_{l=1}^{k}\mu\left(l,k\right)\mathbf{s}^H\hat{\mathbf{R}}^l\mathbf{s}, \tag{4.58}$$

where we have defined

$$\mu\left(l,k\right) = \left(-1\right)^{k+l}\frac{l!}{k!}\sum_{t\in T(l,k)}\xi_P\left(t\right)\;\theta_1^{l_1}\theta_2^{l_2}\cdots\theta_k^{l_k}, \tag{4.59}$$

and

$$\theta_i = \frac{\partial^i}{\partial x^i}\,\theta\left(x\right)\big|_{x=0}. \tag{4.60}$$

In the following lemma, we derive an explicit closed-form expression for the coefficients $\theta_i$, $i = 0, 1, 2, \ldots$. (Note that, since $h\left(\theta\right)$ is always different from zero, we straightforwardly find that $\theta_0 = \theta\left(0\right) = 0$ and $\theta_1 = 1$.)

**Lemma 4.4** *With the previous definitions, the coefficients $\theta_k$, $k > 1$, are given by*

$$\theta_{k+1} = \sum_{l=1}^{k} (-1)^l \, l! \sum_{t \in T(l,k)} \xi_P(t) \;\; \psi_1^{l_1} \psi_2^{l_2} \cdots \psi_k^{l_k},$$

*where*

$$\psi_k = \sum_{l=1}^{k} (-1)^l \frac{(k+1)!}{(k-l+1)!} \sum_{t \in T(l,k)} \binom{l}{l_1, l_2, \ldots, l_l} \; \hat{m}_1^{l_1} \hat{m}_2^{l_2} \cdots \hat{m}_l^{l_l}.$$

**Proof.** From the definition in (4.60), the $i$th coefficient can be recovered from the formal power series expansion of $\theta(x)$ as $i! \left[x^i\right] \theta(x)$ around 0. Now, let (4.55) be written as

$$\theta(x) = x\phi(\theta(x)),$$

where we have defined

$$\phi(u) = \frac{1}{h(u)} = \frac{1}{1 - c + \frac{1}{N} \operatorname{Tr}\left[\left(\mathbf{I}_M + u\hat{\mathbf{R}}\right)^{-1}\right]}.$$

Since $\phi(0) = 1 \neq 0$, we may use the version of the Lagrange inversion formula in Colorary 3 in order to obtain the series coefficients from

$$\left[x^i\right] \theta(x) = \frac{1}{i} \left[u^{i-1}\right] \phi(u)^i.$$

Hence, we have

$$\begin{aligned} \theta_i &= i! \left[x^i\right] \theta(x) \\ &= \left[u^{i-1}\right] \phi(u)^i \\ &= \frac{\partial^{i-1}}{\partial u^{i-1}} \left\{\phi(u)^i\right\}\Big|_{u=0}. \end{aligned} \tag{4.61}$$

In order to obtain (4.61), we further define $\phi(u)^k = f(g(h(u)))$, with $f(y) = \frac{1}{y}$ and $g(v) = v^k$, and apply (B.4) using

$$\frac{\partial^l}{\partial g^l} \{f(g)\}\Big|_{u=0} = (-1)^l \, l!,$$

$$\frac{\partial^l}{\partial h^l} \left\{g(h) = h^{k+1}\right\}\Big|_{u=0} = \frac{(k+1)!}{(k-l+1)!},$$

and

$$\frac{\partial^l}{\partial u^l} \left\{h(u) = 1 - c + \frac{1}{N} \operatorname{Tr}\left[\left(\mathbf{I}_M + u\hat{\mathbf{R}}\right)^{-1}\right]\right\}\Big|_{u=0} = (-1)^l \, l! \hat{m}_l,$$

to finally obtain the result in the lemma. ■

Observe that the $i$th coefficient turns out to be related to $\hat{m}_k$, $k = 1, \ldots, i-1$. In particular, for $i = 2, 3, 4$, we have

$$\begin{aligned} \theta^{(2)} &= 2\hat{m}_1 \\ \theta^{(3)} &= 12\hat{m}_1^2 - 6\hat{m}_2 \\ \theta^{(4)} &= 24\hat{m}_3 - 120\hat{m}_1\hat{m}_2 + 120\hat{m}_1^3. \end{aligned}$$

Finally, we realize that the coefficients $\theta_k$ can also be obtained recursively by directly differentiating both sides of (4.55) $i$ times and letting $x = 0$. Indeed, from

$$\frac{\partial^i}{\partial x^i}\left\{\theta\left(x\right)h\left(\theta\left(x\right)\right)\right\} = \sum_{\upsilon=0}^{i}\binom{i}{\upsilon}\frac{\partial^\upsilon}{\partial x^\upsilon}\left\{\theta\left(x\right)\right\}\frac{\partial^{i-\upsilon}}{\partial x^{i-\upsilon}}\left\{h\left(\theta\left(x\right)\right)\right\},$$

the values of $\theta_i$, $i > 1$, can be readily shown to obey the following recurrent relation, namely,

$$\theta_i = \sum_{\upsilon=1}^{i-1}\frac{i!}{\upsilon!}\left(-1\right)^{i-\upsilon+1}\breve{m}_{i-\upsilon}\ \theta_\upsilon, \tag{4.62}$$

where $\breve{m}_k$ is the M,N-consistent estimator of $m_k$ (note that the RHS of (4.62) only depends on $\theta_r$, $r < i - 1$).

The result in Proposition 4.2 follows by letting $\mu_S\left(l, k\right) = \mu\left(l, k\right)$.

# Chapter 5

# Doubly-Consistent Robust Spatial Filtering under Signature Mismatch

## 5.1 Summary

In this chapter, the consistency of sample robust Capon beamforming (RCB) solutions that are constructed under signature-mismatch constraints from a set of received array observations is revised and improved. Particular emphasis is placed on the class of robust filters heuristically modelling the adverse effects of practical finite sample-size conditions as due to an imperfect knowledge of the effective spatial signature. In contrast, and as in practice, a small sample-size relative to the array dimension is identified in this chapter as the actual source of filter estimation errors under unknown second-order statistics. Accordingly, a new alternative approach to RCB design is proposed in this work that explicitly addresses both the signature-mismatch problem and the limitations due to a finite sample-support. To that effect, based on results borrowed from random matrix theory (RMT) on the estimation of spectral functions of the observation covariance matrix, a class of RCB estimators is derived that generalizes conventional implementations by proving to be consistent even for a limited number of samples per observation dimension. As a result, an improved performance is demonstrated via numerical simulations in the context of source power estimation.

## 5.2 Introduction

The Capon spectral estimator has been widely applied in the array processing literature to the problem of detecting a given number of radiating sources by using an array of passive sensors [Sto05]. This type of problem finds applications in radar and sonar systems, communications, astrophysics, biomedical signal processing, seismology, underwater surveillance and many other fields. The derivation and properties of the Capon method as a spatial spectrum estimator is entirely analogous to the case dealing with the estimation problem of the power spectrum of time data series. When used as a (spatial-reference) beamformer, the knowledge of the array manifold and the second-order statistics of the array output signal is exploited in order to enhance angular response of the minimum variance distorsionless-response (MVDR) spatial filter towards the direction of the source-of-interest (SOI), while nulling out the unwanted contribution from interfering sources. In the array signal processing problem, the power and the direction of arrival (DoA) of the signals impinging on the antenna array are obtained from the evaluation of the estimated angular spectrum.

Under perfect knowledge of the source spatial signature and assuming an infinite number of snapshots is available in order to estimate the theoretical covariance matrix of the received observations, the Capon beamformer is known to offer better resolution and interference rejection capabilities than existing data-independent beamformers. However, these two assumptions are quickly violated in realistic scenarios and so is the filter performance known to suffer from a severe degradation in practical implementations [Ger99]. Regarding the first type of mismatch, the assumption of perfect knowledge of the SOI steering vector is usually not satisfied due to e.g. inaccuracies in the angular (pointing) information, mutual coupling between antenna elements and small array response (calibration) errors. In this situations, the MVDR beamformer may suppress the SOI as an interference, which results in a underestimated SOI power and a drastically reduced signal-to-interference-plus-noise ratio (SINR). Consequently, significant effort has been devoted during the past years to the problem of improving the performance of optimum filters under imprecise knowledge of the steering signature vector. In order to specifically cope with the detrimental effects due to this problem, different robust designs of the Capon beamformer particularly involving a diagonal loading factor [Tre02] and essentially based on the vast mathematical theory of optimization have been recently presented in the literature (see [Li05]).

As for the second source of error, an insufficient sample-support may cause a considerable mismatch between the true and the sample covariance matrix (SCM). In order to consider also the negative consequences of a having a limited number of observed samples available, the main stream of proposed methods heuristically model the small-sample constraint as also due to spatial signature errors. However, the practical implementation of the optimal robust solution relies on the sample estimate of the unknown second-order statistics, namely the sample covariance matrix (SCM). The SCM represents a suitable approximation of the actual array covariance matrix under the assumption of a sufficiently large ratio between sample size and dimension [1].

[1] Indeed, the SCM is the minimum variance unbiased estimator of the theoretical covariance matrix (as well as the maximum likelihood estimator for Gaussian observations). Moreover, the SCM is a consistent estimator whenever the observation dimension remains bounded as the sample-size grows without bound.

Hence, in practice, the major source of errors in the statistical estimation of filtering solutions can be actually identified with a low sample-size relative to the dimension of the array observation. Improving on the traditional presumption of an infinite sample-support for array covariance approximation, a consistent estimator of the optimum diagonal loading factor has been reported in [Mes06c] (see also [Li05, Chapter 4]) that is consistent even for arbitrarily large arrays.

In this chapter, we follow a similar approach as in [Mes06c] and propose a new alternative RCB design that explicitly addresses both the signature-mismatch problem and the limitations due to a limited sample-size. For our purposes, motivated by Girko's general statistical analysis (GSA) of large-dimensional observations [Gir95] and provide a class of RCB estimators that generalizes conventional implementations by proving to be consistent for ar high-dimensional observations for a limited number of samples per observation dimension. Consequently, the proposed construction generalizes conventional RCB implementations that, based on directly replacing the theoretical covariance matrix with its sample estimate As a result, an improved performance in the context of source power estimation is numerically demonstrated.

The chapter is organized as follows. In Section 5.3, a class of robust Capon beamformers is introduced that is particularly suitable for problem of power spectral estimation. In Section 5.4, an asymptotic performance characterization of the conventional SCM-based RCB implementation in the limiting regime defined by both sample size and dimension going to infinity, is provided. Section 5.5 presents the proposed generalized consistent estimator of a family of RCB solutions, which is numerically evaluated in Section 5.6 in the context of the SOI power spectrum. After a short discussion and the concluding remarks in Section 5.7, the theoretical framework for the derivation of the results in the chapter is provided.

## 5.3 Robust Capon spatial filtering

Consider a collection of $N$ multivariate observations $\{\mathbf{y}(n) \in \mathbb{C}^M\}$ obtained by sampling across an antenna array with $M$ sensors, namely, $\{y_m(n), n = 1, \ldots, N, m = 1, \ldots, M\}$, such that $\mathbf{y}(n) = \begin{bmatrix} y_1(n) \cdots y_M(n) \end{bmatrix}^T$. A number of $K$ different sources are assumed to impinge on the antenna array from different directions. Under the assumption of narrowband signals and linear antenna elements, the array observation $\mathbf{y}(n) = \begin{bmatrix} y_1(n) \cdots y_M(n) \end{bmatrix}^T \in \mathbb{C}^M$ can be additively decomposed as

$$\mathbf{y}(n) = x(n)\mathbf{s} + \mathbf{n}(n), \tag{5.1}$$

where $x(n) \in \mathbb{C}$ models the signal waveform (or fading channel coefficient) associated with a given signal of interest at the $n$th discrete-time instant and $\mathbf{s} \in \mathbb{C}^M$ is its spatial signature vector (also steering vector or array transfer vector); furthermore, $\mathbf{n}(n) \in \mathbb{C}^M$ is the additive contribution of the interfering sources and background noise, which can be additively decomposed as $\mathbf{n}(n) = \sum_{k=1}^{K-1} x_k(n)\,\mathbf{s}_k + \mathbf{v}(n)$, where, for $k = 1, \ldots, K-1$, $x_k(n) \in \mathsf{C}$ and $\mathbf{s}_k \in \mathsf{C}^M$ are, respectively, the interfering signal processes and associated steering signatures, and $\mathbf{v}(n) \in \mathsf{C}^M$ is the system noise and out-of-system interference. Conventionally, the signals and the noise are assumed to be independent and jointly distributed wide-sense stationary random processes, with SOI power and noise covariance given, respectively, by $\mathsf{E}[x^*(n)\,x(n)] = \sigma_x^2 \delta_{m,n}$ and $\mathsf{E}\left[\mathbf{n}(m)\,\mathbf{n}^H(n)\right] = \mathbf{R}_{\mathsf{n}}\delta_{m,n}$.

In this work, we focus on the problem of estimating the signal waveform of the intended source and, specifically, on the statistical approximation of the SOI power using optimal spatial filtering techniques. In particular, the Capon beamformer is obtained from the following linearly constrained quadratic optimization problem, namely, [Sto05]

$$\mathbf{w}_{\mathsf{CAPON}} = \arg \min_{\mathbf{w} \in \mathsf{C}^M} \mathbf{w}^H \mathbf{R} \mathbf{w} \;\; \text{subject to}\, \mathbf{w}^H \mathbf{s} = 1, \tag{5.2}$$

where $\mathbf{R}$ is the theoretical covariance matrix of the array observation, which under the previous statistical assumptions is given by

$$\mathbf{R} = \sigma_x^2 \mathbf{s}\mathbf{s}^H + \mathbf{R}_{\mathsf{n}}. \tag{5.3}$$

The solution to (5.2) can be straigthforwardly found as

$$\mathbf{w}_{\mathsf{CAPON}} = \frac{\mathbf{R}^{-1}\mathbf{s}}{\mathbf{s}^H \mathbf{R}^{-1} \mathbf{s}}. \tag{5.4}$$

Furthermore, from above, the SOI power can be approximated by $\mathsf{E}\left[|\hat{x}(n)|^2\right] = \mathbf{w}^H \mathbf{R} \mathbf{w}$. Hence, with some abuse of notation, the Capon SOI power estimate is defined as

$$\sigma^2_{\mathsf{CAPON}} = \frac{1}{\mathbf{s}^H \mathbf{R}^{-1} \mathbf{s}}. \tag{5.5}$$

The rationale behind this procedure is that a natural (indirect) solution for the SOI power approximant must be possibly obtained by minimizing the power of the interference-plus-noise received contribution in (5.2) while keeping the intended signal unchanged.

The Capon SOI power estimate in (5.5) is well-known to significantly outperform solutions obtained from data-independent beamforming methods, provided that the actual SOI spatial signature is precisely known. However, in the case that only an inaccurate version of the SOI steering vector is available, as it usually happens in practice, a relatively significant performance degradation is in constrast to be expected. In order to alleviate this problem, a number of robust adaptive beamforming techniques has been proposed in the literature that provide a generalization of the original Capon beamformer above with the purpose of allowing for an improved estimation performance even under an imprecise knowledge of the SOI spatial signature. In particular, different robust solutions have been published over the past recent years that extend on the diagonal loading approach by providing an optimum loading level based on pressumed information about the uncertainty of the array steering vector [Li03, Vor03a, Sha03, Li04a, Lor05, Bec07] (see also [Hon95, Wan98a, Wan98b, Mad98] for a related approach against code signature waveform mismatch in the context of code-division multiple-access (CDMA) multiuser detection). More specifically, the imperfectly known spatial signature is most often assumed to belong to an uncertainty ellipsoidal set, according to which the corresponding amount of diagonal loading is explicitely calculated.

As in [Sto03], we are here interested in the problem of estimating directly the power of the SOI as in (5.5) robustly against a mismatch in the spatial signature. Building upon a contrained-covariance-fitting direct derivation of the Capon SOI power estimate provided in [Sto03], a robust Capon estimate is proposed in [Li03] that uniquely relies on the available imperfect knowledge about the steering vector and the pressumed uncertainty level. In particular, assuming that the steering vector is contained in an ellipsod described by a given positive definite matrix $\mathbf{C} \in \mathsf{C}^{M\times M}$ and centered at a nominal steering vector $\tilde{\mathbf{s}} \in \mathsf{C}^{M}$, the solution is given by the expression in (5.5), with the unknown steering vector being replaced by

$$\mathbf{s}_o = \arg\min_{\mathbf{s}\in\mathsf{C}^M} \mathbf{s}^H\mathbf{R}^{-1}\mathbf{s} \ \text{subject to} \left(\mathbf{s}-\tilde{\mathbf{s}}\right)^H \mathbf{C}^{-1}\left(\mathbf{s}-\tilde{\mathbf{s}}\right) \leq 1. \tag{5.6}$$

In [Li03], the solution of the optimization problem in (5.6) is found for an uncertainty set $\mathbf{C} = \epsilon\mathbf{I}_M$, yielding as a constraint the sphere $\|\mathbf{s}-\tilde{\mathbf{s}}\| \leq \epsilon$, with $\epsilon \in \mathsf{R}^+$ being a given user parameter. Thus, the optimum robust steering vector is obtained as

$$\mathbf{s}_o = \left(\mathbf{I}_M - \left(\mathbf{I}_M + \lambda_o\mathbf{R}\right)^{-1}\right)\tilde{\mathbf{s}}, \tag{5.7}$$

where the parameter $\lambda_o$ is found as the real positive solution of the following equation in $\lambda$, namely,

$$g\left(\lambda\right) = \epsilon, \tag{5.8}$$

where we have defined

$$g\left(\lambda\right) = \tilde{\mathbf{s}}^H\left(\mathbf{I}_M + \lambda\mathbf{R}\right)^{-2}\tilde{\mathbf{s}}.$$

Note that $g\left(\lambda\right)$ is a monotonically decreasing function of $\lambda$ for $\lambda > 0$ (see [Li03] for further details). Hence, using the available erroneous version of the true steering vector in (5.7), the robust estimate of the SOI power is given by

$$\sigma^2_{\mathsf{CAPON}} = \frac{1}{\mathbf{s}_o^H\mathbf{R}^{-1}\mathbf{s}_o}. \tag{5.9}$$

The previous approach to robust Capon beamforming can be extended to a broader class of solutions including an additional norm-constraint (or white-noise gain constraint) on the weight vector as classically proposed in order to optimally select an appropriate diagonal-loading factor (see [Hud81]). Based equivalently on the formulation in [Sto03] of the Capon beamforming problem, a doubly-constrained RCB is proposed in [Li04a] that is similarly obtained as a function of the spectrum of the covariance matrix and the associated eigensubspaces as well as the given parametrization of the uncertainty region (see also discussion in [Sto05, Section 6.5] and [Li05, Chapter 3]).

In practice, the array observation covariance matrix is most often not available, and so must the RCB necessarily rely on its sample estimate, namely the sample covariance matrix (SCM), i.e.,

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{y}(n) \mathbf{y}^{H}(n). \tag{5.10}$$

The SCM is known to be a consistent estimator of the theoretical covariance matrix when the number of samples available for its computation is infinitely larger than the size of the array. Obviously, such an assumption does hardly match realistic scenario conditions given in a practical setting. Consequently, a considerable performance degradation is to be expected when implementing the RCB solution above by directly replacing the true covariance matrix with the SCM.

In the next section, we provide a characterization of the performance of sample estimates in an asymptotic regime defined by a comparatively large sample size and dimension.

## 5.4 Asymptotic Convergence of Conventional Implementations

In this section, we derive the asymptotic expression of the power level approximant in (5.9) in Section 5.3 when constructed using the available SCM under the realistic assumption of an array sample observation of comparably large size and dimension. To that effect, we resort to the theory of the spectral analysis of large-dimensional random matrices or RMT. Specifically, we build upon results involving the Stieltjes transform of spectral probability measures. This fundamental building block allows us to characterize the asymptotic eigenspectrum of the SCM in terms of the limiting spectral distribution of the theoretical covariance matrix as all dimensions of the random matrix model increase without bound at a constant rate. For our purposes, not only the asymptotic spectrum but also the limiting behavior of the associated eigensubspaces are of interest.

Before presenting the main result of this section, observe that, from (5.1) and the statistical assumptions in Section 5.3, we can statistically model the observed samples as $\mathbf{y}(n) = \mathbf{R}^{1/2}\mathbf{u}(n)$, where $\mathbf{u}(n) \in \mathbb{C}^{M}$, $n = 1, \ldots, N$, is a collection of i.i.d. random vectors, whose entries have zero mean real and imaginary parts with variance 1/2 and bounded higher moments. Thus, the SCM in (5.10) will be modeled in the subsequent derivations as

$$\hat{\mathbf{R}} = \frac{1}{N} \mathbf{R}^{1/2} \mathbf{U} \mathbf{U}^{H} \mathbf{R}^{1/2}, \tag{5.11}$$

where the matrix $\mathbf{U} \in \mathbb{C}^{M \times N}$ is constructed using as its columns the vectors $\mathbf{u}(n)$, $n = 1, \ldots, N$.

For technical reasons, we assume that the nominal steering vector is normalized, such that $\|\tilde{\mathbf{s}}\| = 1$.

The next two propositions characterize the asymptotic behavior of the conventional SCM-based construction of the RCB.

**Proposition 5.1** *Let $\hat{\mathbf{R}}$ be given by (5.11). Moreover, consider an $M$ dimensional nonrandom vectors $\tilde{\mathbf{s}}$ with uniformly bounded Euclidean norm for all $M$. Then, as $M, N \to c < +\infty$, the solution in $\lambda$ to*

$$\tilde{\mathbf{s}}^H \left(\mathbf{I}_M + \lambda \hat{\mathbf{R}}\right)^{-2} \tilde{\mathbf{s}} = \epsilon,$$

*converges to the solution in $\lambda$ to the following equation, denoted as $\bar{\lambda}_o$, for $\eta = -\frac{1}{\lambda}$, namely,*

$$\eta f_1(\eta) f_2(\eta) \tilde{\mathbf{s}}^H (\mathbf{R} - f_1(\eta)\mathbf{I}_M)^{-2} \tilde{\mathbf{s}} + \eta (\eta f_2(\eta) - f_1(\eta)) \tilde{\mathbf{s}}^H (\mathbf{R} - f_1(\eta)\mathbf{I}_M)^{-1} \tilde{\mathbf{s}} = \epsilon,$$

*where*

$$f_2(\eta) = \frac{1}{1 - \frac{c}{M}\sum_{m=1}^{M}\left(\frac{\lambda_m(\mathbf{R})}{\lambda_m(\mathbf{R}) - f_1(\eta)}\right)^2},$$

*and $f = f_1(\eta)$ is the solution in $f$ to the following equation, namely,*

$$f = \frac{\eta}{1 - \frac{c}{M}\sum_{m=1}^{M}\frac{\lambda_m(\mathbf{R})}{\lambda_m(\mathbf{R}) - f}}.$$

**Proof.** See Appendix A. ■

Regarding the asymptotic convergence of the direct SCM-based implementation of the SOI power estimate, we have

**Proposition 5.2** *Let $\hat{\mathbf{R}}$, $\tilde{\mathbf{s}}$ and $\bar{\lambda}_o$ be defined as in Proposition 5.1. Moreover, define $\hat{\mathbf{s}}_o = \left(\mathbf{I}_M - \left(\mathbf{I}_M + \bar{\lambda}_o \hat{\mathbf{R}}\right)^{-1}\right)\tilde{\mathbf{s}}$. Then, as $M, N \to c < +\infty$, we have for $\eta = -\frac{1}{\lambda}$,*

$$\frac{1}{\hat{\mathbf{s}}_o^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{s}}_o} \asymp f_2(\eta) \tilde{\mathbf{s}}^H (\mathbf{R} - f_1(\eta)\mathbf{I}_M)^{-1} \mathbf{R} (\mathbf{R} - f_1(\eta)\mathbf{I}_M)^{-1} \tilde{\mathbf{s}},$$

*where $f_1(\eta)$ and $f_2(\eta)$ are given in Proposition 5.1.*

**Proof.** See Appendix B. ■

Clearly, from Propositions 5.1 and 5.2, the estimators of the optimum robust parameter and the robust power estimate are not consistent for arbitrarily large dimensional array observations. In order to further improve the power estimaton performance under situations characterized by a large array, or, equivalently, a relatively small number of observations, in the next section we derive an estimator of the RCB solution in (5.9) that is consistent even in the case of a limited number of samples per observation dimension.

## 5.5 Improved consistent RCB estimation

In this section, we propose a generalized consistent estimator of (5.9) for comparatively large sample size and dimension. The new RCB estimator generalizes the conventional implementation based on directly replacing the covariance matrix by its sample estimate by being consistent even for comparably large sample size ($N$) and dimension ($M$). To that effect, as in Section 5.4, we resort to Stieltjes transform methods dealing with the asymptotic spectrum of SCM-type matrices. Specifically, we derive an estimator of the optimum parameter $\lambda_o$ as well as the class of spectral functions of $\mathbf{R}$ defined by (5.9) that is consistent in the asymptotic regime described by $M, N \to \infty$, with $M/N \to c < +\infty$. We will refer to this estimators as $M, N$-consistent as a generalization of traditional $N$-consistent estimators.

The following two propositions provide asymptotic equivalents of the optimum parameter defining the robust steering vector in (5.7) and the SOI power estimate in (5.9), respectively, as a function of only the SCM.

**Proposition 5.3** *Let $\epsilon$, $\tilde{\mathbf{s}}$, $\mathbf{R}$, $\lambda_o$ and $\hat{\mathbf{R}}$ be defined as above. Under the previous statistical assumptions, as $M, N \to \infty$, with $M/N \to c < +\infty$, we have that $\lambda_o \asymp \breve{\lambda}$, where*

$$\breve{\lambda} = \frac{\breve{\eta}}{g_2(\breve{\eta})}, \tag{5.12}$$

*where $\breve{\eta}$ is the solution to the following equation in $\eta$, namely, $\breve{g}(\eta) = \epsilon$, where*

$$\breve{g}(\eta) = \frac{\eta^2 \left[ g_2(\eta)\, \tilde{\mathbf{s}}^H \left( \hat{\mathbf{R}} - \eta \mathbf{I}_M \right)^{-2} \tilde{\mathbf{s}} - g_1(\eta)\, \tilde{\mathbf{s}}^H \left( \hat{\mathbf{R}} - \eta \mathbf{I}_M \right)^{-1} \tilde{\mathbf{s}} \right]}{g_2(\eta) + \eta g_1(\eta)}, \tag{5.13}$$

*and*

$$g_1(\eta) = \frac{1}{N} \operatorname{Tr} \left[ \hat{\mathbf{R}} \left( \hat{\mathbf{R}} - \eta \mathbf{I}_M \right)^{-2} \right],$$
$$g_2(\eta) = 1 - \frac{c}{M} \operatorname{Tr} \left[ \hat{\mathbf{R}} \left( \hat{\mathbf{R}} - \eta \mathbf{I}_M \right)^{-1} \right].$$

**Proof.** See Appendix C. ∎

Note that the function $\breve{g}(\eta)$ is also monotonically decreasing for $\eta$ smaller than the minimum eigenvalue of $\hat{\mathbf{R}}$. Moreover, regarding the estimation of the robust SOI power approximant, we have

**Proposition 5.4** *Under the assumptions and definitions above, as $M, N \to \infty$, with $M/N \to c < +\infty$, we have that $\sigma^2_{\mathsf{CAPON}} \asymp \breve{\sigma}^2_{\mathsf{CAPON}}$, where*

$$\breve{\sigma}^2_{\mathsf{CAPON}} = \frac{\frac{1}{\breve{\mu}\breve{\lambda}} \left( \frac{g_1(\breve{\mu})}{\lambda_o} - 1 \right)}{\tilde{\mathbf{s}}^H \left( \hat{\mathbf{R}} - \breve{\mu} \mathbf{I}_M \right)^{-1} \hat{\mathbf{R}} \left( \hat{\mathbf{R}} - \breve{\mu} \mathbf{I}_M \right)^{-1} \tilde{\mathbf{s}}}, \tag{5.14}$$

*where $\breve{\lambda}$ is given in Proposition 1 and $\breve{\mu}$ is the solution to the following equation in $\mu$, namely,*

$$\mu = -\frac{1}{\breve{\lambda}} g_2(\mu).$$

**Proof.** See Appendix D. ■

Therefore, based on the result in Proposition 5.3, we can establish that

**Corollary 5.1** *The random quantity $\breve{\sigma}^2_{\mathsf{CAPON}}$ in (5.14) is a strongly consistent estimator of $\sigma^2_{\mathsf{CAPON}}$.*

## 5.6 Numerical evaluations

In the following, we consider the typical array processing application concerning the estimation of the SOI power. Instead of relying on the availability of an accurately known SOI spatial signature, we assume that an erroneous measurement or estimate of the steering vector is available and allow for a certain degree of uncertainty level in its knowledge. In particular, we numerically compare the performance of both the *conventional* (based on the direct substitution of $\mathbf{R}$ with $\hat{\mathbf{R}}$) and *proposed* (based on Propositions 5.3 and 5.4) implementations of the robust Capon power estimate in (5.9). Specifically, we consider a scenario consisting of $K = 5$ sources impinging on an array with $M = 30$ sensor elements from angles (degrees) $\{0, 20, 30, 50, 60\}$ and powers (dB) $\{10, 5, 30, 10, 25\}$ over the noise level $\boldsymbol{\sigma}_n^2 = 1$. Moreover, a number of observed samples equal to $N = 20$ is assumed to be available for SCM computation (note that $N < M$). Finally, the constant scalar defining the uncertainty level is fixed to $\epsilon = 1$.

Figure 5.1 shows the simulations results for the estimation of the optimum parameter $\lambda_o$. In particular, the empirical probability density function (PDF) of both conventional and proposed estimators is depicted versus the theoretical value of $\lambda$ obtained by solving equation (5.8) using the true covariance matrix. Observe that, even in the considered adverse (undersampled) estimation conditions, an apparent nearly unbiased behavior and a lower variance can be empirically appreciated for the proposed estimator against a highly biased and variant performance of the conventional implementation.

On the other hand, in Figure 5.2 the normalized histograms of the robust SOI power estimate obtained via the conventional as well as the proposed methods are depicted. The actual SOI power value is also shown. As before, the conventional estimator is characterized by an empirically observed unbiased behavior against the conventional implementation. Finally, regarding the algorithmic complexity, the leading computational constraint representing the bulk of the computation of the proposed generalized $M, N$-consistent estimator essentially involves implementing the EVD of the SCM. Thus, the number of required arithmetic operations is of the same order of magnitude as that corresponding to the original $N$-consistent RCB solution.

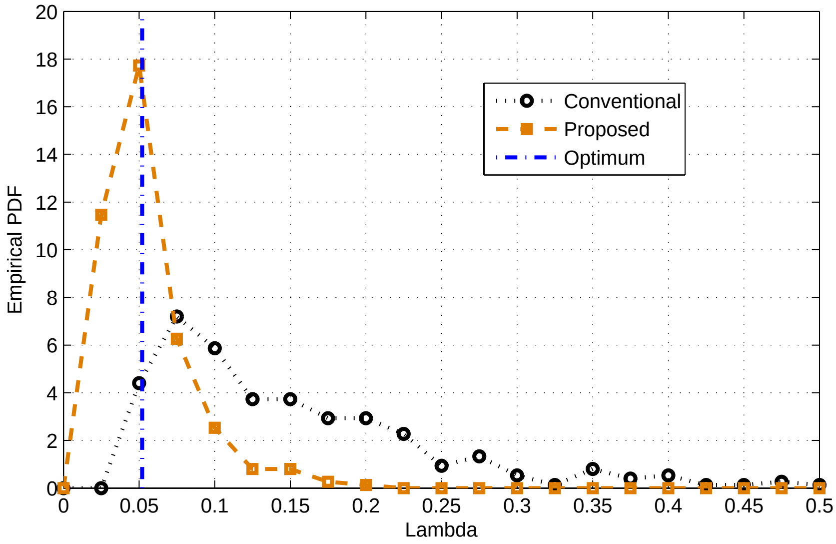


Figure 5.1: Normalized histogram of conventional and proposed estimates of $\lambda_o$ under a number of realizations of the SCM versus actual value from theoretical covariance matrix.

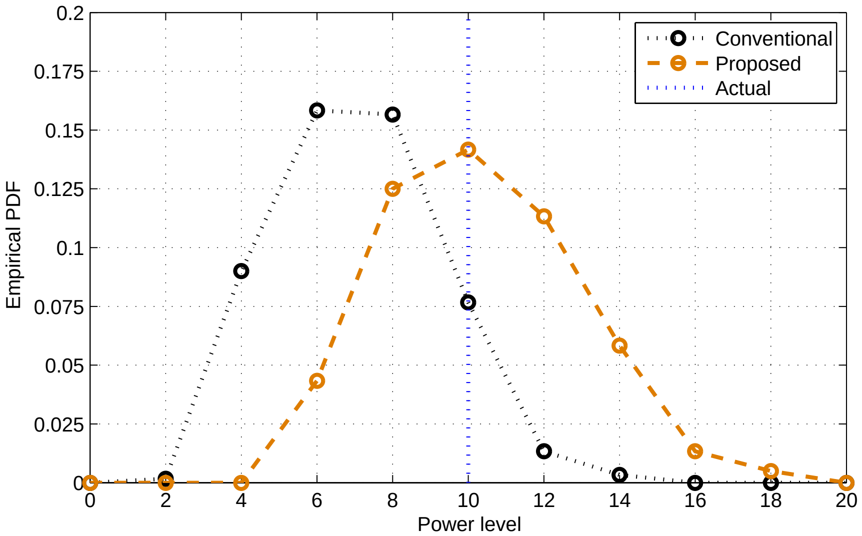


Figure 5.2: Normalized histogram of a number of realizations of the conventional and proposed estimates of (5.9) versus actual SOI power value

## 5.7 Conclusions

In this chapter, the consistency of robust Capon beamforming solutions have been revised and an improved alternative to the conventional implementation of the RCB has been proposed that is consistent for an arbitrary number of samples per array element. We have focused on the problem of avoiding the performance degradation observed in practice and associated with RCB implementations based on an erroneous SOI spatial signature vector and the sample approximation of the theoretical covariance matrix of the array observations. In the array processing literature, adverse effects of a practical finite sample-support are most often heuristically modelled as also due to an imperfect knowledge of the SOI steering information. On the contrary, and as it happens to be the case in realistic settings, a small sample-size relative to the array dimension is identified in this work as the actual source of filter estimation errors under unknown second-order statistics. Thus, the problem of improving the RCB performance under practical non-ideal operation conditions has been distilled to the separate optimization of the beamformer design under signature-mismatch constraints and the generalized consistent estimation of the optimal robust solution. Accordingly, a new alternative approach to RCB design is proposed in this work that explicitly addresses both the signature-mismatch problem and the limitations due to a finite sample-size. To that effect, we have resorted to the theory of the spectral analysis of large-dimensional random matrices in order to derive an estimator that generalizes conventional implementations by proving to be consistent even for a limited number of samples per observation dimension. Interestingly enough, the proposed RCB construction nearly presents the same computational complexity than the conventional RCB based on the direct application of the SCM, as well as the traditional SMI implementation of the standard Capon beamformer. As a result, an improved performance is demonstrated via numerical simulations in the context of a typical application of SOI power estimation.

## Appendix 5.A Proof of Proposition 5.1

Recall from Section 1.1 the following results regarding the asymptotic behavior of the eigenvalues and eigenvectors of SCM-type matrices as $M, N \to c < +\infty$, for all $z \in \mathbb{C}$ outside the limiting eigenvalue support of $\hat{\mathbf{R}}$, namely,

$$\frac{1}{M}\operatorname{Tr}\left[\left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1}\right] \asymp \frac{1}{M}\operatorname{Tr}\left[\left(w\left(z\right)\mathbf{R} - z\mathbf{I}_M\right)^{-1}\right],$$

where $w\left(z\right) = 1 - c - czm\left(z\right)$ and $m = m\left(z\right)$ is the unique solution in the set $\left\{m \in \mathbb{C} : -\left(1-c\right)/z + cm \in \mathbb{C}^{+}\right\}$ to the following equation, namely,

$$m = \frac{1}{M}\sum_{m=1}^{M}\frac{\lambda_m\left(\mathbf{R}\right)}{\lambda_m\left(\mathbf{R}\right)\left(1-c-czm\right) - z}.$$

Furthermore, for two $M$ dimensional deterministic vectors $\mathbf{a}, \mathbf{b}$ with uniformly bounded Euclidean norm for all $M$,

$$\mathbf{a}^H\left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1}\mathbf{b} \asymp \mathbf{a}^H\left(w\left(z\right)\mathbf{R} - z\mathbf{I}_M\right)^{-1}\mathbf{b}. \tag{5.15}$$

Equivalently, if we consider $f\left(z\right) = z/w\left(z\right)$, we have

$$w\left(z\right) \asymp \hat{w}\left(z\right), \tag{5.16}$$

$$f\left(z\right) \asymp \hat{f}\left(z\right), \tag{5.17}$$

where

$$\hat{w}\left(z\right) = 1 - c - cz\frac{1}{M}\operatorname{Tr}\left[\left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1}\right] \tag{5.18}$$

$$= 1 - \frac{c}{M}\sum_{m=1}^{M}\frac{\lambda_m\left(\hat{\mathbf{R}}\right)}{\lambda_m\left(\hat{\mathbf{R}}\right) - z}, \tag{5.19}$$

and

$$\hat{f}\left(z\right) = \frac{z}{\hat{w}\left(z\right)}, \tag{5.20}$$

respectively.

In order to prove Proposition 5.1, observe that

$$\tilde{\mathbf{s}}^H\left(\mathbf{I}_M + \lambda\hat{\mathbf{R}}\right)^{-2}\tilde{\mathbf{s}} = \frac{1}{\lambda^2}\frac{\partial}{\partial z}\left\{\tilde{\mathbf{s}}^H\left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1}\tilde{\mathbf{s}}\right\}\Big|_{z=-\frac{1}{\lambda}}.$$

On the other hand, note that

$$\tilde{\mathbf{s}}^H\left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1}\tilde{\mathbf{s}} \asymp \frac{1}{w\left(z\right)}\tilde{\mathbf{s}}^H\left(\mathbf{R} - f\left(z\right)\mathbf{I}_M\right)^{-1}\tilde{\mathbf{s}}, \tag{5.21}$$

where we can express $w(z)$ as

$$w(z) = \frac{1}{1 - \frac{c}{M} \sum_{m=1}^{M} \frac{\lambda_m(\mathbf{R})}{\lambda_m(\mathbf{R}) - f(z)}}.$$

Thus, the result follows by replacing $\tilde{\mathbf{s}}^H \left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1} \tilde{\mathbf{s}}$ with its asymptotic equivalent in (5.21), after differentiating and noting that

$$f'(z) = \frac{1}{1 - \frac{c}{M} \sum_{m=1}^{M} \left(\frac{\lambda_m(\mathbf{R})}{\lambda_m(\mathbf{R}) - f(z)}\right)^2}.$$

## Appendix 5.B Proof of Proposition 5.2

In order to prove Proposition 5.2, observe first that

$$\hat{\mathbf{s}}_o^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{s}}_o = \tilde{\mathbf{s}}^H \left(\hat{\mathbf{R}} + \frac{1}{\lambda}\mathbf{I}_M\right)^{-1} \hat{\mathbf{R}} \left(\hat{\mathbf{R}} + \frac{1}{\lambda}\mathbf{I}_M\right)^{-1} \tilde{\mathbf{s}},$$

so that we can write

$$\hat{\mathbf{s}}_o^H \hat{\mathbf{R}}^{-1} \hat{\mathbf{s}}_o = \tilde{\mathbf{s}}^H \left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1} \hat{\mathbf{R}} \left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1} \tilde{\mathbf{s}} \Big|_{z=-\frac{1}{\lambda}}.$$

Then, the limiting expression can be readily obtained from the following asymptotic equivalent of $\tilde{\mathbf{s}}^H \left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1} \hat{\mathbf{R}} \left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1} \tilde{\mathbf{s}}$ that can be derived following the lines of the proof in Appendix G of Chapter 2, namely,

$$\tilde{\mathbf{s}}^H \left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1} \hat{\mathbf{R}} \left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1} \tilde{\mathbf{s}} \asymp f'(z)\, \tilde{\mathbf{s}}^H \left(\mathbf{R} - f(z)\,\mathbf{I}_M\right)^{-1} \mathbf{R} \left(\mathbf{R} - f(z)\,\mathbf{I}_M\right)^{-1} \tilde{\mathbf{s}},$$

and using the results outlined at the beginning of Appendix A.

## Appendix 5.C Proof of Proposition 5.3

Based on the results in Appendix A, we let $f(z) = -\frac{1}{\lambda}$ and write (5.8) as

$$f(z)^2\, \tilde{\mathbf{s}}^H \left(\mathbf{R} - f(z)\,\mathbf{I}_M\right)^{-2} \tilde{\mathbf{s}} = \epsilon. \tag{5.22}$$

Furthermore, we note that the identity in (5.22) can be written as

$$\frac{f(z)^2}{f'(z)} \frac{\partial}{\partial z} \left\{\tilde{\mathbf{s}}^H \left(\mathbf{R} - f(z)\,\mathbf{I}_M\right)^{-1} \tilde{\mathbf{s}}\right\} = \epsilon. \tag{5.23}$$

Moreover, the LHS of (5.23) can be (asymptotically) equivalently expressed as

$$\frac{f(z)^2}{f'(z)}\frac{\partial}{\partial z}\left\{\tilde{\mathbf{s}}^H\left(\mathbf{R}-f(z)\mathbf{I}_M\right)^{-1}\tilde{\mathbf{s}}\right\} \asymp \frac{\hat{f}(z)^2}{\hat{f}'(z)}\frac{\partial}{\partial z}\left\{\hat{w}(z)\,\tilde{\mathbf{s}}^H\left(\hat{\mathbf{R}}-z\mathbf{I}_M\right)^{-1}\tilde{\mathbf{s}}\right\}. \tag{5.24}$$

The RHS of (5.24) depends only on the SCM and defines an asymptotically equivalent equation (5.8) that can be expressed, after some analysis and defining $z=\eta$, as $\breve{g}(\eta)=\epsilon$, with $\breve{g}(\eta)$ being given by (5.13) in Proposition 5.3. Thus, the $M,N$-consistent estimator of the optimum parameter can be obtained by first solving for the value of $\eta$ satisfying $\breve{g}(\eta)=\epsilon$, denoted by $\breve{\eta}$, and then finding an asymptotic equivalent of $-\frac{1}{f(\breve{\eta})}$ (cf. (5.12) in Proposition 5.3).

## Appendix 5.D Proof of Proposition 5.4

First, observe that

$$\mathbf{s}_o^H\mathbf{R}^{-1}\mathbf{s}_o = \tilde{\mathbf{s}}^H\left(\mathbf{R}-f(z)\mathbf{I}_M\right)^{-1}\mathbf{R}\left(\mathbf{R}-f(z)\mathbf{I}_M\right)^{-1}\tilde{\mathbf{s}}, \tag{5.25}$$

where we have fixed $f(z)=-\frac{1}{\lambda_o}$. As in Appendix B, an asymptotic equivalent of the RHS of (5.25) can be found as

$$\tilde{\mathbf{s}}^H\left(\mathbf{R}-f(z)\mathbf{I}_M\right)^{-1}\mathbf{R}\left(\mathbf{R}-f(z)\mathbf{I}_M\right)^{-1}\tilde{\mathbf{s}} \asymp \frac{1}{\hat{f}'(z)}\tilde{\mathbf{s}}^H\left(\hat{\mathbf{R}}-z\mathbf{I}_M\right)^{-1}\hat{\mathbf{R}}\left(\hat{\mathbf{R}}-z\mathbf{I}_M\right)^{-1}\tilde{\mathbf{s}}. \tag{5.26}$$

Then, using (5.26), along with $\lambda_o \asymp \breve{\lambda}$ (cf. Proposition 5.3) and

$$\hat{f}'(z) = \frac{1-\hat{f}(z)\,\hat{w}'(z)}{\hat{w}(z)},$$

with

$$\hat{w}'(z) = \frac{1}{N}\operatorname{Tr}\left[\hat{\mathbf{R}}\left(\hat{\mathbf{R}}-z\mathbf{I}_M\right)^{-2}\right],$$

we obtain the solution in (5.14).

A final remark about the optimality of the previous procedure is in order. The almost surely pointwise convergence stated in (5.15) can be in fact extended to uniform convergence (outside the eigenvalue support) by Egoroff's lemma [Dud02]. Then, by the Weierstrass convergence theorem [Ahl78], the limit of the RHS of (5.15) is also an analytic function (see alternatively Vitali's theorem on the uniform convergence of sequences of uniformly bounded holomorphic functions towards a holomorphic function [Rud87, Hil62]. Consequently, we are allowed to directly differentiate the RHS of (5.15) as in (5.24) in order to find the proposed estimator. Finally, observe that the sequence $\left\{\breve{\lambda}_M=\breve{\lambda}\right\}$ in Proposition 5.3 certainly converges towards $\lambda_o$ for an increasing $M$, as it follows indeed from the previous argumentation (i.e., the solutions of the proposed asymptotic equivalent of equation (5.8) converge to the actual solution of (5.8)).

# Appendix 5.E Proof of Proposition 5.3 and Proposition 5.4 using GSA

Recall from Section 1.3 the following convergence results based on the definition of the real-valued Stieltjes transform, namely, for $x \in \mathsf{R}^{+}$,

$$\frac{1}{M}\operatorname{Tr}\left[\left(\mathbf{I}_M + x\mathbf{R}\right)^{-1}\right] \asymp \frac{1}{M}\operatorname{Tr}\left[\left(\mathbf{I}_M + \theta\left(x\right)\hat{\mathbf{R}}\right)^{-1}\right], \tag{5.27}$$

and

$$\tilde{\mathbf{s}}^H \left(\mathbf{I}_M + x\mathbf{R}\right)^{-1} \tilde{\mathbf{s}} \asymp \tilde{\mathbf{s}}^H \left(\mathbf{I}_M + \theta\left(x\right)\hat{\mathbf{R}}\right)^{-1} \tilde{\mathbf{s}}, \tag{5.28}$$

as $M, N \to \infty$, with $M/N \to c < +\infty$, where $\theta\left(x\right)$ is the positive solution of the following canonical equation, namely,

$$\theta\left(x\right)\left(1 - c + \frac{1}{N}\operatorname{Tr}\left[\left(\mathbf{I}_M + \theta\left(x\right)\hat{\mathbf{R}}\right)^{-1}\right]\right) = x. \tag{5.29}$$

In particular, note that

$$f\left(z\right) w\left(z\right) = f\left(z\right)\left(1 - c + \frac{1}{N}\operatorname{Tr}\left[\left(\mathbf{I}_M - f\left(z\right)\mathbf{R}\right)^{-1}\right]\right) = z,$$

with $z$ being restricted to the real negative axis, resembles the canonical equation in (4.55) for $\theta\left(x\right) = f\left(-\frac{1}{x}\right)$.

Regarding the result in Proposition 5.3, note first that

$$g\left(\lambda\right) = \tilde{\mathbf{s}}^H \left(\mathbf{I}_M + \lambda\mathbf{R}\right)^{-2} \tilde{\mathbf{s}} = \tilde{\mathbf{s}}^H \left(\mathbf{I}_M + \lambda\mathbf{R}\right)^{-1} \tilde{\mathbf{s}} + \frac{\partial}{\partial x}\left\{\tilde{\mathbf{s}}^H \left(\mathbf{I}_M + x\mathbf{R}\right)^{-1} \tilde{\mathbf{s}}\right\}\Bigg|_{x=\lambda}.$$

Consequently, if we define

$$r_H\left(x\right) = \tilde{\mathbf{s}}^H \left(\mathbf{I}_M + \theta\left(x\right)\hat{\mathbf{R}}\right)^{-1} \tilde{\mathbf{s}},$$

we may use (5.28) in order to find an (asymptotically exact) approximation of $g\left(\lambda\right)$ in terms of $r_H\left(x\right)$ as

$$g\left(\lambda\right) \asymp t\left(\lambda\right) + x\frac{\partial}{\partial x}\left\{r_H\left(x\right)\right\}\Bigg|_{x=\lambda}, \tag{5.30}$$

and solve for the optimum parameter by letting the RHS of (5.30) be equal to $\epsilon$. Note that only the available sample estimate $\hat{\mathbf{R}}$ (as well as the given nominal steering vector $\tilde{\mathbf{s}}$ and the user parameter $\epsilon$) is involved in the calculation of the approximant of $\lambda_o$. In fact, such approximation clearly represent an $M, N$-consistent estimator of $\lambda_o$.

A major difficulty appears from the fact that, in the computation of the optimum parameter (usually following an interative procedure), an update of the solution $\theta\left(x\right)$ to the canonical equation in (4.55) is required at each step. An alternative derivation of the $M, N$-consistent estimator of $\lambda_o$ is provided next that avoids the previous computational requirement.

Finally, concerning the proof of Proposition 5.4, first note that the denominator of (5.9) can be written using the matrix inversion lemma as

$$\mathbf{s}_o^H \mathbf{R}^{-1} \mathbf{s}_o = \tilde{\mathbf{s}}^H \left( \mathbf{R} + \frac{1}{\lambda_o} \mathbf{I}_M \right)^{-1} \mathbf{R} \left( \mathbf{R} + \frac{1}{\lambda_o} \mathbf{I}_M \right)^{-1} \tilde{\mathbf{s}}. \tag{5.31}$$

Thus, using the estimator in (5.28), an $M, N$-consistent estimator of (5.31) can be readily obtained as

$$\mathbf{s}_o^H \mathbf{R}^{-1} \tilde{\mathbf{s}} \asymp - x^2 \frac{\partial}{\partial x} \left\{ t(x) \right\} \Big|_{x=\lambda_o} = \breve{\lambda}^2 \theta' \left( \breve{\lambda} \right) \tilde{\mathbf{s}}^H \left( \mathbf{I}_M + \theta \left( \breve{\lambda} \right) \hat{\mathbf{R}} \right)^{-1} \mathbf{R} \left( \mathbf{I}_M + \theta \left( \breve{\lambda} \right) \hat{\mathbf{R}} \right)^{-1} \tilde{\mathbf{s}},$$

where $\breve{\lambda}$ is the asymptotic equivalent of $\lambda_o$ given in Proposition 5.3 and $\theta\left(\breve{\lambda}\right)$ is the solution to the canonical equation in (5.29) at $x = \breve{\lambda}$, with first-order derivative

$$\theta' \left( \breve{\lambda} \right) = \frac{1}{1 - c + \frac{1}{N} \operatorname{Tr} \left[ \left( \mathbf{I}_M + \theta \left( \breve{\lambda} \right) \hat{\mathbf{R}} \right)^{-1} \right] - \theta \left( \breve{\lambda} \right) \frac{1}{N} \operatorname{Tr} \left[ \left( \mathbf{I}_M + \theta \left( \breve{\lambda} \right) \hat{\mathbf{R}} \right)^{-1} \mathbf{R} \left( \mathbf{I}_M + \theta \left( \breve{\lambda} \right) \hat{\mathbf{R}} \right)^{-1} \right]}.$$

# AFTERWORD and Upcoming Research Topics

This dissertation has dealt with the practical application of statistical signal processing to the fundamental problems of optimal signal waveform and power estimation in wireless communications and sensor array processing. In particular, special emphasis has been placed on the analysis and design under situations characterized by a finite sample-size, and a relatively large observation dimension. Based on the theory of the spectral analysis of large-dimensional random matrices, an analytical framework has been developed throughout this thesis that allowed us to derive a family of new statistical inference methods overcoming the limitations of traditional inferential schemes under the previous conditions. Specifically, a class of consistent estimators has been proposed that generalizes conventional implementations by proving to be consistent even for arbitrarily high-dimensional observations (i.e., for a limited number of samples per filtering degree-of-freedom).

In particular, the new theoretical framework have been shown to properly characterize the performance of multi-antenna and multi-channel signal processing systems with training preambles in the more meaningful asymptotic regime. Moreover, the problem of optimum reduced-rank linear filtering has been reviewed and extended to satisfy the generalized consistency definition. On the other hand, a double-limit asymptotic characterization of a class of vector-valued quadratic forms involving the negative powers of the observation covariance matrix has been provided that generalizes existing results on the limiting eigenvalue moments.of the inverse Wishart distribution. Using these results, a new generalized consistent eigenspectrum estimator based on the inverse-shifted power method has been derived that uniquely relies on the SCM and does not require eigendecomposition operation. The effectiveness of the previous spectral estimator has been demonstrated upon its application to the construction of an improved source power estimator that is robust to inaccuracies in the knowledge of both the noise level and the true covariance matrix.

In order to alleviate the computation complexity issue associated with practical implementations involving matrix inversions, a solution to the two previous problems was afforded in terms of the positive powers of the SCM. To that effect, a class of generalized consistent estimators of the covariance eigenspectrum and the power level were obtained on the Krylov subspace defined by the true covariance matrix and the signature vector associated with the intended parameter. In practice, filtering solutions are very often required to robustly operate not only under sample-size constraints but also under the availability of an imprecise knowledge of the signature vector. As a final contribution of this thesis, a signal-mismatch robust filtering architecture has been proposed that is consistent in the doubly-asymptotic regime characterizing a sample of comparably large size and dimension.

In conclusion, the application of random matrix theory to reviewing and characterizing the performance and accuracy of traditional statistical signal processing techniques in situations where the sample size and dimension are comparable in magnitude, as well as to designing and evaluating suitable statistical inference methods appropriately operating in this general regime undoubtedly represents a promising line of prominent current and future research. In the following, some topics of upcoming work are shortly outlined.

### First-order analysis of optimal signal processing schemes under general SCM structure

Throughout this dissertation, results from random matrix theory have been extended and applied in order to provide an asymptotic first-order characterization of traditional signal processing architectures. Based on a more meaningful double-limit regime, the conventional consistency definition has been reviewed and a class of new statistical signal processing schemes has been provided to overcome the inherent limitations of system implementations under low sample-support, relatively high-dimensional observations. The proposed statistical inference methods rely on a certain structure of the sample covariance matrix, namely resulting from the signal and noise processes being jointly distributed. However, in many scenarios of unquestionable practical relevance, the signal waveform cannot be assumed to be distributed according to the law from which the noise process is drawn. For instance, this is the case of a transmission in Gaussian noise of a sequence of symbols modeled as discrete random variables drawn from a finite probability space. In these situations, the information-plus-noise-type model represents a more suitable random matrix model for the SCM. Hence, a new family of results regarding the asymptotic characterization of optimum signal processing schemes based on second-order statistics can be obtained that further generalize the existing findings.

### Large-system performance analysis of training-based communication architectures

Chapter 1 presents an asymptotic characterization of the transient SNR performance of pilot-aided MIMO systems under the assumption of a training phase length and a system size of comparable magnitude. The training beamvector sequence is assumed to satisfy the WBE. Along with the case of orthogonal training, of theoretical interest for comparison purposes is also the case of i.i.d. pilot beamvector entries. This study allows for an analytical insight into the actual degree of diversity gain achieved over arbitrary block-fading MIMO channels in practice. More stress has been definintely laid in the wireless communication literature on the performance characterization of energy-constrained MIMO systems with limited training in terms of the mutual information measure. In this context, the theoretical framework developed in this thesis can be used to properly investigate the actual capacity gains from the statistical inference perspective in the large-system regime.

### Statistical characterization of linear filtering and subspace methods in signal processing systems

The asymptotic characterization of traditional signal processing systems provided in this work is essentially based on the derivation of asymptotic equivalents obtained from a first-order limiting analysis in the general regime. Moreover, the inferential methods proposed throughout the thesis basically come off as consequence of inverting the limiting results from this first-order analysis. However, as intuitively expected, such a first-order asymptotic characterization does not prove to effectively explain crucial details describing the behavior of empirical procedures and practical signal processing schemes. As an example, although random matrix theory tools have proved to properly characterize the limiting behavior of statistical subspace MUSIC-like inferential methods, the performance breakdown effect of subspace-based parameter estimation methods escape to this characterization. Thus, the need for a more sophisticated study of the fluctuations of the limiting solutions is apparent. To that effect, CLT-like results recently appeared in the literature can be applied with some extensions in order to characterize the variance and the distribution of the previous methods [Mes06d, Mes06a].

# Appendix A

# Asymptotic Convergence Results

In this appendix, we provide some results on the asymptotic convergence of random sequences that are repeatedly used in the proofs throughout the dissertation. In the following presentation, basic elements from probability theory are involved, such as random variables inequalities, as well as the notion of convergence with probability one (or almost surely) of random sequences. For a textbook treatment of the fundamentals, see e.g. [Ash99, Bil95, Chu01, Dud02, Dur95].

The following result generalizes the concept of stochastic convergence of random sequences under certain operations. In particular, it involves the convergence of stochastic sequences under the application of continuous maps, such as additions and multiplications in linear spaces (see e.g. [Ser80, p. 24]). In the literature, the original result for convergence in distribution is regarded as continuous mapping or Mann-Wald theorem.

**Lemma A.1** *(Convergence of transformed random sequences) Let* $\mathbf{x}_1, \mathbf{x}_2, \ldots$ *and* $\mathbf{x}$ *be random vectors in* $\mathsf{C}^p$ *defined on a certain probability measure space* $(\Omega, \mathcal{F}, \mathsf{P})$*. Furthermore, let* $g : \mathsf{C}^p \to \mathsf{C}^q$ *be a Borel-measurable function and assume that* $g$ *is continuous almost everywhere or, equivalently, with* $\mathsf{P}_{\mathsf{x}}$*-probability one on a set* $A$ [1]*. Here, a complex-valued function is said to be measurable if both its real and imaginary parts are measurable, integrable functions. Then,*

$$\mathbf{x}_N \asymp \mathbf{x} \Rightarrow g(\mathbf{x}_N) \asymp g(\mathbf{x}) .$$

**Proof.** The assertion follows immediately from the definitions of almost sure convergence and continuity (see e.g. [Bar66, p. 78, Exercise 7.R]). Let $\Omega_0 = \{\omega : \mathbf{x}_n(\omega) \to \mathbf{x}(\omega)\}$ and $\Omega_1 = \{\omega : \mathbf{x}(\omega) \in A\}$. Thus, for $\omega \in (\Omega_0 \cap \Omega_1)$, ($\mathsf{P}_{\mathsf{x}}$-)continuity of $g$ ensures that $g(\mathbf{x}_n(\omega)) \to g(\mathbf{x}(\omega))$ (see e.g. [Mun00, p. 130, Theorem 21.3] and [Apo73, p. 78, Theorem 4.16]). Note that $(\Omega_0 \cap \Omega_1)^c = \Omega_0^c \cup \Omega_1^c$, which has probability zero because $\mathsf{P}(\Omega_0^c) = \mathsf{P}(\Omega_1^c) = 0$. It follows that $\Omega_0 \cap \Omega_1$ has probability one and the result is proved. ■

[1] The set $A \subseteq \mathsf{C}^p$ of continuity points of $g$ in $\mathsf{C}^p$ has measure one, where the measure is the probability distribution $\mathsf{P}_{\mathsf{x}}$ induced by $\mathsf{x}$ in $\mathsf{C}^p$, i.e. $\mathsf{P}_{\mathsf{x}}(A) = \mathsf{P}(\mathsf{x} \in A) = 1$.

For example, $\mathbf{z}_N \asymp \mathbf{z}$ in $\mathbb{C}^2$ implies e.g. $z_{1N} + z_{2N} \asymp z_1 + z_2$ and $z_{1N} z_{2N} \asymp z_1 z_2$, where $z_{1N}, z_{2N}$ are the the entries of the vector $\mathbf{z}_N$ and $z_1, z_2$ are the corresponding elements of $\mathbf{z}$.

**Lemma A.2** *a) Let $\left\{y_1^{(N)}, \ldots, y_N^{(N)}\right\}$ denote a collection of (possibly dependent) random variables such that*

$$\max_{1 \leq m \leq N} \mathsf{E}\left[\left|y_m^{(N)}\right|^p\right] \leq \frac{C_a}{N^{1+\delta_a}},$$

*for some constants $C_a$, $\delta_a > 0$ and $p \geq 2$ not depending on $N$. Then,*

$$\frac{1}{N} \sum_{m=1}^{N} \left|y_m^{(N)}\right| \to 0,$$

*almost surely as $N \to \infty$.*

*b) (Double array) Let $X_{m,n}^{(N)}$, $m, n = 1, \ldots, N$ denote a double array of (possibly dependent) random variables such that*

$$\max_{1 \leq m,n \leq N} \mathsf{E}\left[\left|X_{m,n}^{(N)}\right|^p\right] \leq \frac{C_b}{N^{1+\delta_b}},$$

*for some constants $C_b$, $\delta_b > 0$ and $p \geq 2$ not depending on $N$. Then,*

$$\frac{1}{N^2} \sum_{m=1}^{N} \sum_{n=1}^{N} \left|X_{m,n}^{(N)}\right| \to 0,$$

*almost surely as $N \to \infty$.*

**Proof.** Note that, given $\epsilon > 0$, the Chebyshev inequality implies that

$$\Pr\left[\frac{1}{N} \sum_{m=1}^{N} \left|y_m^{(N)}\right| > \epsilon\right] \leq \frac{1}{\epsilon^p} \mathsf{E}\left[\left(\frac{1}{N} \sum_{m=1}^{N} \left|y_m^{(N)}\right|\right)^p\right].$$

On the other hand, using Jensen's inequality and the convexity of $f(x) = |x|^p$ for $p \geq 2$, we have

$$\mathsf{E}\left[\left(\frac{1}{N} \sum_{m=1}^{N} \left|y_m^{(N)}\right|\right)^p\right] \leq \frac{1}{N} \sum_{m=1}^{N} \mathsf{E}\left[\left|y_m^{(N)}\right|^p\right].$$

Consequently,

$$\Pr\left[\frac{1}{N} \sum_{m=1}^{N} \left|y_m^{(N)}\right| > \epsilon\right] \leq \frac{1}{\epsilon^p} \frac{1}{N} \sum_{m=1}^{N} \mathsf{E}\left[\left|y_m^{(N)}\right|^p\right] \leq \frac{1}{\epsilon^p} \max_{1 \leq m \leq N} \mathsf{E}\left[\left|y_m^{(N)}\right|^p\right] \leq \frac{1}{\epsilon^p} \frac{C_a}{N^{1+\delta_a}}.$$

The result in a) follows from the Borel-Cantelli lemma. In order to prove b), exactly the same line of reasoning can be followed with $y_m^{(N)} = \frac{1}{N} \sum_{n=1}^{N} \left|X_{m,n}^{(N)}\right|$. In particular, we just need to note that

$$\mathsf{E}\left[\left(\frac{1}{N^2} \sum_{m=1}^{N} \sum_{n=1}^{N} \left|X_{m,n}^{(N)}\right|\right)^p\right] \leq \frac{1}{N^2} \sum_{m=1}^{N} \sum_{n=1}^{N} \mathsf{E}\left[\left|X_{m,n}^{(N)}\right|^p\right],$$

so that we finally have

$$\Pr\left[\frac{1}{N^2}\sum_{m=1}^{N}\sum_{n=1}^{N}\left|X_{m,n}^{(N)}\right|>\epsilon\right]\leq\frac{1}{\epsilon^p}\frac{1}{N^2}\sum_{m=1}^{N}\sum_{n=1}^{N}\mathsf{E}\left[\left|X_{m,n}^{(N)}\right|^p\right]\leq\frac{1}{\epsilon^p}\max_{1\leq m,n\leq N}\mathsf{E}\left[\left|X_{m,n}^{(N)}\right|^p\right]\leq\frac{1}{\epsilon^p}\frac{C_b}{N^{1+\delta_b}},$$

and the result is equivalently proved using the Borel-Cantelli lemma. ■

**Lemma A.3** *([Bai98, Lemma 2.7]) Let $\mathbf{u}_n$ denote an $M$-dimensional random vector with i.i.d. complex random entries with zero mean and unit variance and $\mathbf{C}$ an $M\times M$ complex matrix. Then, we have, for any $p\geq 2$,*

$$\mathsf{E}\left[\left|\mathbf{u}_n^H\mathbf{C}\mathbf{u}_n-\mathrm{Tr}\left[\mathbf{C}\right]\right|^p\right]\leq K_p\left[\left(\mathsf{E}\left[|\xi|^4\right]\mathrm{Tr}\left[\mathbf{C}\mathbf{C}^H\right]\right)^{p/2}+\mathsf{E}\left[|\xi|^{2p}\right]\mathrm{Tr}\left[\left(\mathbf{C}\mathbf{C}^H\right)^{p/2}\right]\right],$$

*where $\xi$ denotes a particular entry of $\mathbf{u}_n$ and $K_p$ is a given constant that does not depend on $\mathbf{C}$.*

**Lemma A.4** *Let $\mathbf{u}_n$ be a random vector defined as in Lemma A.3 and $\mathbf{A}$ an $M\times M$ complex matrix such that $\|\mathbf{A}\|_{\mathrm{tr}}$ is uniformly bounded for all $M$. Then, for any finite $p$,*

$$\mathsf{E}\left[\left|\mathbf{u}_n^H\mathbf{A}\mathbf{u}_n\right|^p\right]<+\infty, \tag{A.1}$$

*for all $M$.*

**Proof.** First, we use the Jensen inequality in order to write

$$\mathsf{E}\left[\left|\mathbf{u}_n^H\mathbf{A}\mathbf{u}_n\right|^p\right]<2^{p-1}\left\{\mathsf{E}\left[\left|\mathbf{u}_n^H\mathbf{A}\mathbf{u}_n-\mathrm{Tr}\left[\mathbf{A}\right]\right|^p\right]+\left|\mathrm{Tr}\left[\mathbf{A}\right]\right|^p\right\}. \tag{A.2}$$

Then, the second term in the RHS of (A.2) is bounded by assumption since, clearly, $\left|\mathrm{Tr}\left[\mathbf{A}\right]\right|\leq\|\mathbf{A}\|_{\mathrm{tr}}$. The first term is bounded by Lemma A.3, since $\|\mathbf{A}\|_F\leq\|\mathbf{A}\|_{\mathrm{tr}}$. ■

**Remark A.1** *The following two cases of special interes in our derivations are included in Lemma A.4, namely $\mathbf{A}=\mathbf{b}\mathbf{a}^H$ and $\mathbf{A}=\frac{1}{M}\mathbf{B}$, where $\mathbf{a},\mathbf{b}\in\mathbb{C}^M$ are two determistic vectors with uniformly bounded Euclidean norm and $\mathbf{B}$ is an $M\times M$ complex matrix with uniformly bounded spectral radius.*

**Lemma A.5** *(SLLN for sequences of non-independent variables) Let $\mathbf{u}_n$ be an $M$-dimensional random vector as defined in Lemma A.3 and $\mathbf{U}_n$ an $M\times M$ complex random matrix depending on all vectors of the set $\{\mathbf{u}_1,\ldots,\mathbf{u}_{n-1},\mathbf{u}_{n+1},\ldots,\mathbf{u}_N\}$ and with bounded spectral norm for all $M$. Let further consider an arbitrary matrix $\mathbf{C}$ such that $\|\mathbf{C}\|_F$ is uniformly bounded for all $M$. Then,*

$$\frac{1}{N}\sum_{n=1}^{N}\mathbf{u}_n^H\mathbf{U}_n\mathbf{C}\mathbf{u}_n\asymp\mathrm{Tr}\left[\mathbf{U}_n\mathbf{C}\right]. \tag{A.3}$$

**Proof.** Let us first rewrite (A.3) as

$$\frac{1}{N}\sum_{n=1}^{N}\eta_n \asymp 0,$$

where we have defined $\eta_n = \mathbf{u}_n^H\mathbf{U}_n\mathbf{C}\mathbf{u}_n - \mathrm{Tr}\left[\mathbf{U}_n\mathbf{C}\right]$. We note that $\{\eta_n\}$ is a martingale difference sequence with respect to the increasing $\sigma$-fields $\{\mathcal{F}_n\}$, where $\mathcal{F}_n$ is generated by the random vectors $\{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$. Indeed, observe that

$$\mathsf{E}\left[\eta_n\left|\mathcal{F}_{n-1}\right.\right] = \mathsf{E}\left[\mathbf{u}_n^H\mathbf{U}_n\mathbf{C}\mathbf{u}_n\left|\mathcal{F}_{n-1}\right.\right] - \mathsf{E}\left[\mathrm{Tr}\left[\mathbf{U}_n\mathbf{C}\right]\left|\mathcal{F}_{n-1}\right.\right] = 0,$$

since

$$\mathsf{E}\left[\mathbf{u}_n\mathbf{u}_n^H\mathbf{U}_n\left|\mathcal{F}_{n-1}\right.\right] = \mathsf{E}\left[\mathbf{U}_n\left|\mathcal{F}_{n-1}\right.\right].$$

Consequently, we can apply the Burkholder inequality [Bur73], namely

$$\mathsf{E}\left[\left|\sum_{n=1}^{N}\eta_n\right|^p\right] \leq K_p\left\{\mathsf{E}\left[\left(\sum_{n=1}^{N}\mathsf{E}\left[\left|\eta_n\right|^2\left|\mathcal{F}_{n-1}\right.\right]\right)^{p/2}\right] + \mathsf{E}\left[\sum_{n=1}^{N}\left|\eta_n\right|^p\right]\right\},$$

for any $p \geq 2$ and some constant $K_p$. Applying Lemma A.3 we can write, for $p = 2q$ and $q \in \mathsf{N}$,

$$\begin{aligned}\mathsf{E}\left[\left|\eta_n\right|^{2q}\right] &= \mathsf{E}\left[\left|\mathbf{u}_n^H\mathbf{U}_n\mathbf{C}\mathbf{u}_n - \mathrm{Tr}\left[\mathbf{U}_n\mathbf{C}\right]\right|^{2q}\right] \\ &\leq K\left(\mathsf{E}^q\left[\left|\xi\right|^4\right] + \mathsf{E}\left[\left|\xi\right|^{4q}\right]\right),\end{aligned}$$

where $\xi$ denotes a particular entry of $\mathbf{u}_n$, and we have used the following inequality involving the Frobenius and strong norm, namely

$$\left\|\mathbf{U}_n\mathbf{C}\right\|_F \leq \left\|\mathbf{U}_n\right\|\left\|\mathbf{C}\right\|_F, \tag{A.4}$$

which is uniformly bounded by assumption for all $M$. On the other hand, using again Lemma A.3,

$$\mathsf{E}\left[\left|\eta_n\right|^2\left|\mathcal{F}_{n-1}\right.\right] = \mathsf{E}\left[\left|\mathbf{u}_n^H\mathbf{U}_n\mathbf{C}\mathbf{u}_n - \mathrm{Tr}\left[\mathbf{U}_n\mathbf{C}\right]\right|^2\left|\mathcal{F}_{n-1}\right.\right] \leq K\mathsf{E}\left[\left|\xi\right|^4\right].$$

Therefore,

$$\mathsf{E}\left[\left|\frac{1}{N}\sum_{n=1}^{N}\eta_n\right|^{2q}\right] \leq \frac{1}{N^{2q}}\left(K\left(N\mathsf{E}\left[\left|\xi\right|^4\right]\right)^q + NK'\left(\mathsf{E}^q\left[\left|\xi\right|^4\right] + \mathsf{E}\left[\left|\xi\right|^{4q}\right]\right)\right) = \frac{1}{N^q}K'' + \frac{1}{N^{2q-1}}K''',$$

and the result is readily proved by direct application of the Borel-Cantelli lemma with $q > 1$. ■

**Lemma A.6** *Let $\mathbf{u}_m, \mathbf{u}_n$ be two independent random vectors defined as in Lemma A.3 and $\mathbf{S}, \mathbf{T}$ be two $M \times M$ complex matrices with, respectively, uniformly bounded trace and spectral norm. Then, for any finite p,*

$$\mathsf{E}\left[\left|\mathbf{u}_m^H\mathbf{S}\mathbf{u}_n\mathbf{u}_n^H\mathbf{T}\mathbf{u}_m\right|^p\right] < +\infty, \tag{A.5}$$

*for all $M$.*

**Proof.** In order to prove the result in (A.5), we are forced to resort to complex martingale difference techniques and the Burkholder inequality as in Lemma A.5. For that purpose, we begin using the singular value decomposition of $\mathbf{T}$ to write

$$\mathbf{u}_m^H \mathbf{S} \mathbf{u}_n \mathbf{u}_n^H \mathbf{T} \mathbf{u}_m = \sum_{k=1}^{M} \mathbf{v}_k^H \mathbf{u}_m \mathbf{u}_m^H \mathbf{S} \mathbf{u}_n \mathbf{u}_n^H \mathbf{s}_k,$$

where $\{\sigma_k\}, \{\mathbf{s}_k\}$ and $\{\mathbf{v}_k\}$, $k = 1, \ldots, M$, are the sets of $M$ singular values and left and right singular vectors of the matrix $\mathbf{T}$, having by assumption uniformly bounded absolute value and Euclidean norm, respectively, for all $M$. Further, we define $\zeta_k = \mathbf{v}_k^H \mathbf{u}_m \mathbf{u}_m^H \mathbf{S} \mathbf{u}_n \mathbf{u}_n^H \mathbf{s}_k - \mathbf{v}_k^H \mathbf{S} \mathbf{s}_k$ and observe, using Jensen's inequality, that proving (A.5) is equivalent to showing that

$$\mathsf{E}\left[\left|\sum_{k=1}^{M} \zeta_k + \mathbf{v}_k^H \mathbf{S} \mathbf{s}_k\right|^p\right] \leq K \left\{ \mathsf{E}\left[\left|\sum_{k=1}^{M} \zeta_k\right|^p\right] + \left|\mathrm{Tr}\left[\mathbf{S}\mathbf{T}\right]\right|^p \right\} < +\infty.$$

Hence, since $\left|\mathrm{Tr}\left[\mathbf{S}\mathbf{T}\right]\right|^p \leq \left(\mathrm{Tr}\left[\left|\mathbf{S}\mathbf{T}\right|\right]\right)^p \leq \left(\rho\left(\mathbf{T}\right) \mathrm{Tr}\left[\left|\mathbf{S}\right|\right]\right)^p = \left\|\mathbf{T}\right\|^p \left\|\mathbf{S}\right\|_{\mathrm{tr}}^p < +\infty$, the problem in (A.5) is reduced to

$$\mathsf{E}\left[\left|\sum_{k=1}^{M} \zeta_k\right|^p\right] < +\infty. \tag{A.6}$$

where we have defined $\zeta_k = \mathbf{v}_k^H \mathbf{u}_m \mathbf{u}_m^H \mathbf{S} \mathbf{u}_n \mathbf{u}_n^H \mathbf{s}_k - \mathbf{v}_k^H \mathbf{S} \mathbf{s}_k$. Now, let us consider the sequence of increasing $\sigma$-fields $\{\mathcal{G}_n\}$ generated by the random variables $\{\zeta_1, \ldots, \zeta_n\}$, $n = 1, \ldots, M$. Since

$$\mathsf{E}\left[\zeta_k \left|\mathcal{G}_{k-1}\right.\right] = 0,$$

we note that the sequence$\{\zeta_k\}$, $k = 1, \ldots, M$, forms a martingale difference with respect to $\{\mathcal{G}_n\}$. Therefore, we may apply Burkholder's inequality to bound the expectation in (A.6) as

$$\mathsf{E}\left[\left|\sum_{k=1}^{M} \zeta_k\right|^p\right] \leq K_p \left\{ \mathsf{E}\left[\left(\sum_{k=1}^{M} \mathsf{E}\left[\left|\zeta_k\right|^2 \left|\mathcal{G}_{k-1}\right.\right]\right)^{p/2}\right] + \mathsf{E}\left[\sum_{k=1}^{M} \left|\zeta_k\right|^p\right] \right\}.$$

Furthermore, using Jensen's inequality, we have for any $q > 1$ that

$$\mathsf{E}\left[\left|\zeta_k\right|^q\right] \leq K \left\{ \mathsf{E}\left[\left|\mathbf{u}_n^H \mathbf{s}_k \mathbf{v}_k^H \mathbf{u}_m \mathbf{u}_m^H \mathbf{S} \mathbf{u}_n - \mathrm{Tr}\left[\mathbf{s}_k \mathbf{v}_k^H \mathbf{u}_m \mathbf{u}_m^H \mathbf{S}\right]\right|^q\right] + \mathsf{E}\left[\left|\mathbf{u}_m^H \mathbf{S} \mathbf{s}_k \mathbf{v}_k^H \mathbf{u}_m - \mathrm{Tr}\left[\mathbf{S} \mathbf{s}_k \mathbf{v}_k^H\right]\right|^q\right] \right\}. \tag{A.7}$$

Now, we apply Lemma A.3 and find that

$$\mathsf{E}\left[\left(\sum_{k=1}^{M} \mathsf{E}\left[\left|\zeta_k\right|^2 \left|\mathcal{G}_{k-1}\right.\right]\right)^{p/2}\right] \leq K \mathsf{E}^{p/2}\left[\left|\xi\right|^4\right] \mathsf{E}\left[\left(\mathbf{u}_m^H \mathbf{T}^H \mathbf{T} \mathbf{u}_m \mathbf{u}_m^H \mathbf{S} \mathbf{S}^H \mathbf{u}_m + \mathrm{Tr}\left[\mathbf{S}\mathbf{S}^H \mathbf{T}^H \mathbf{T}\right]\right)^{p/2}\right].$$

Noting that $\mathrm{Tr}\left[\mathbf{S}\mathbf{S}^H \mathbf{T}^H \mathbf{T}\right] = \left\|\mathbf{S}^H \mathbf{T}\right\|_F \leq \left\|\mathbf{S}\right\|_F \left\|\mathbf{T}\right\| \leq \left\|\mathbf{S}\right\|_{\mathrm{tr}} \left\|\mathbf{T}\right\| < +\infty$ and using the binomial theorem along with the Cauchy-Schwarz inequality, we can finally write

$$\mathsf{E}\left[\left(\sum_{k=1}^{M} \mathsf{E}\left[\left|\zeta_k\right|^2 \left|\mathcal{G}_{k-1}\right.\right]\right)^{p/2}\right] \leq K \mathsf{E}^{p/2}\left[\left|\xi\right|^4\right] \sum_{l=1}^{p/2} \mathsf{E}^{1/2}\left[\left(\mathbf{u}_m^H \mathbf{T}^H \mathbf{T} \mathbf{u}_m\right)^{2l}\right] \mathsf{E}^{1/2}\left[\left(\mathbf{u}_m^H \mathbf{S} \mathbf{S}^H \mathbf{u}_m\right)^{2l}\right] < +\infty,$$

since all expectations are readily identified to be bounded by Lemma A.4. On the other hand, by the law of iterated expectations,

$$\mathsf{E}\left[g\left(\mathbf{u}_m, \mathbf{u}_n\right)\right] = \mathsf{E}\left[\mathsf{E}\left[g\left(\mathbf{u}_m, \mathbf{u}_n\right)|\mathbf{u}_n\right]\right],$$

where, here $g\left(\mathbf{u}_m, \mathbf{u}_n\right) = \left|\mathbf{v}_k^H \mathbf{u}_m \mathbf{u}_m^H \mathbf{S} \mathbf{u}_n \mathbf{u}_n^H \mathbf{s}_k - \mathbf{v}_k^H \mathbf{u}_m \mathbf{u}_m^H \mathbf{S} \mathbf{s}_k\right|^q$. Hence, the second expectation on the RHS of (A.7) can be bounded using Jensen's inequality and Lemma A.3 as

$$\sum_{k=1}^{M} \mathsf{E}\left[|\zeta_k|^p\right] \leq K \left(\mathsf{E}^{p/2}\left[|\xi|^4\right] + \mathsf{E}\left[|\xi|^{2p}\right]\right) \left\{\mathsf{E}\left[\left(\mathbf{u}_m^H \mathbf{T}^H \mathbf{T} \mathbf{u}_m\right)^{p/2} \left(\mathbf{u}_m^H \mathbf{S} \mathbf{S}^H \mathbf{u}_m\right)^{p/2}\right] + \left\|\mathbf{S}^H \mathbf{T}\right\|_F^{p/2}\right\} < +\infty,$$

where we have followed the same arguments as above. ■

**Lemma A.7** *Let* $\mathbf{B}$ *be a* $M \times M$ *complex Hermitian matrix and define*

$$\mathbf{B} = \frac{1}{N} \sum_{n=1}^{N} \mathrm{y}_n \mathrm{y}_n^H,$$

$$\mathbf{B}_n = \mathbf{B} - \frac{1}{N} \sum_{n=1}^{N} \mathrm{y}_n \mathrm{y}_n^H.$$

*Then, for any* $z \in \mathsf{C}^+$ *and* $c \in \mathsf{R}$ *such that* $c = M/N$,

$$1 - c - cz\frac{1}{M} \operatorname{Tr}\left[\left(\mathbf{B} - z\mathbf{I}_M\right)^{-1}\right] = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{\mathrm{y}_n^H \left(\mathbf{B}_n - z\mathbf{I}_M\right)^{-1} \mathrm{y}_n}.$$

**Proof.** First, note that

$$\begin{aligned}
1 &= \frac{1}{M} \operatorname{Tr}\left[\left(\mathbf{B} - z\mathbf{I}_M\right)\left(\mathbf{B} - z\mathbf{I}_M\right)^{-1}\right] \\
&= \frac{1}{M} \operatorname{Tr}\left[\frac{1}{N} \sum_{n=1}^{N} \mathrm{y}_n \mathrm{y}_n^H \left(\mathbf{B} - z\mathbf{I}_M\right)^{-1} - z\left(\mathbf{B} - z\mathbf{I}_M\right)^{-1}\right] \\
&= \frac{1}{M} \sum_{n=1}^{N} \frac{1}{N} \mathrm{y}_n^H \left(\mathbf{B}_n + \frac{1}{N} \mathrm{y}_n \mathrm{y}_n^H - z\mathbf{I}_M\right)^{-1} \mathrm{y}_n - z\frac{1}{M} \operatorname{Tr}\left[\left(\mathbf{B} - z\mathbf{I}_M\right)^{-1}\right] \\
&= \frac{1}{M} \sum_{n=1}^{N} \frac{\frac{1}{N} \mathrm{y}_n^H \left(\mathbf{B}_n - z\mathbf{I}_M\right)^{-1} \mathrm{y}_n}{1 + \frac{1}{N} \mathrm{y}_n^H \left(\mathbf{B}_n - z\mathbf{I}_M\right)^{-1} \mathrm{y}_n} - z\frac{1}{M} \operatorname{Tr}\left[\left(\mathbf{B} - z\mathbf{I}_M\right)^{-1}\right], \qquad \text{(A.8)}
\end{aligned}$$

where, in the last equality, we have used the Sherman-Morrison inversion formula for rank-one matrix updates. Furthermore, by expanding the sum in (A.8), we get

$$1 = \frac{N}{M} - \frac{1}{M} \sum_{n=1}^{N} \frac{1}{1 + \frac{1}{N} \mathrm{y}_n^H \left(\mathbf{B}_n - z\mathbf{I}_M\right)^{-1} \mathrm{y}_n} - z\frac{1}{M} \operatorname{Tr}\left[\left(\mathbf{B} - z\mathbf{I}_M\right)^{-1}\right]. \qquad \text{(A.9)}$$

Finally, multiplying both sides in (A.9) by $c$ and rearranging terms, we obtain the result in the lemma. ■

**Lemma A.8** *[Sil95a, Lemma 2.3] For $z \in \mathbb{C}^{+}$ let $m_1(z)$, $m_2(z)$ be Stieltjes transforms of any two probability distribution functions. Moreover, let $\mathbf{A}$ and $\mathbf{B}$ be two $M \times M$ complex matrices, with $\mathbf{A}$ Hermitian non-negative definite, and $\mathbf{r} \in \mathbb{C}^{M}$. Then,*

$$\left\| \left(m_1(z)\,\mathbf{A} + \mathbf{I}_M\right)^{-1} \right\| \leq \max\left(\frac{4\,\|\mathbf{A}\|}{\operatorname{Im}\{z\}}, 2\right), \tag{A.10}$$

$$\begin{aligned} &\left|\operatorname{Tr}\left[\mathbf{B}\left(\left(m_1(z)\,\mathbf{A} + \mathbf{I}_M\right)^{-1} - \left(m_2(z)\,\mathbf{A} + \mathbf{I}_M\right)^{-1}\right)\right]\right| \\ &\leq M \left|m_2(z) - m_2(z)\right| \|\mathbf{A}\| \|\mathbf{B}\| \left(\max\left(\frac{4\,\|\mathbf{A}\|}{\operatorname{Im}\{z\}}, 2\right)\right)^2, \end{aligned} \tag{A.11}$$

$$\begin{aligned} &\left|\mathbf{r}^H \left(m_1(z)\,\mathbf{A} + \mathbf{I}_M\right)^{-1} \mathbf{r} - \mathbf{r}^H \left(m_2(z)\,\mathbf{A} + \mathbf{I}_M\right)^{-1} \mathbf{r}\right| \\ &\leq \left|m_2(z) - m_2(z)\right| \|\mathbf{A}\| \|\mathbf{r}\|^2 \left(\max\left(\frac{4\,\|\mathbf{A}\|}{\operatorname{Im}\{z\}}, 2\right)\right)^2. \end{aligned} \tag{A.12}$$

**Lemma A.9** *[Sil95b, Lemma 2.6] Let $\mathbf{A}$ and $\mathbf{B}$ be two $M \times M$ complex matrices, with $\mathbf{B}$ Hermitian, $\tau \in \mathbb{R}$ and $\mathbf{r} \in \mathbb{C}^{M}$. Then, for $z \in \mathbb{C}^{+}$,*

$$\left|\operatorname{Tr}\left[\left(\left(\mathbf{B} - z\mathbf{I}_M\right)^{-1} - \left(\mathbf{B} + \tau\mathbf{r}\mathbf{r}^H - z\mathbf{I}_M\right)^{-1}\right)\mathbf{A}\right]\right| \leq \frac{\|\mathbf{A}\|}{\operatorname{Im}\{z\}}, \tag{A.13}$$

**Lemma A.10** *[Mes06b, Lemma 7] Let $\mathbf{X}$ will denote an $M \times N$ complex random matrix, such that the real and imaginary parts of the entries of $N^{-1/2}\boldsymbol{\Xi}$ are i.i.d. random variables with mean zero, variance $1/2$ and bounded moments. Moreover, let $\mathbf{R}$ be a $M \times M$ Hermitian non-negative definite matrix, whose eigenvalues are uniformly bounded for all $M$, and define $\hat{\mathbf{R}} = \mathbf{R}^{1/2}\mathbf{X}\mathbf{X}^H\mathbf{R}^{1/2}$ with $\mathbf{R}^{1/2}$ denoting any Hermitian square-root of the matrix $\mathbf{R}$, such that $\hat{\mathbf{R}} = \frac{1}{N}\sum_{n=1}^{N} \mathrm{y}_n \mathrm{y}_n^H$, and $\hat{\mathbf{R}}_n = \hat{\mathbf{R}} - \frac{1}{N}\mathrm{y}_n \mathrm{y}_n^H$. Then, for any $z \in \mathbb{C}^{+}$,*

$$\left|\frac{1}{1 + \frac{1}{N}\mathrm{y}_n^H\left(\hat{\mathbf{R}}_n - z\mathbf{I}_M\right)^{-1}\mathrm{y}_n}\right| \leq \frac{|z|}{\operatorname{Im}\{z\}}, \tag{A.14}$$

$$\left|\frac{1}{1 + \frac{1}{N}\operatorname{Tr}\left[\mathbf{R}\left(\hat{\mathbf{R}} - z\mathbf{I}_M\right)^{-1}\right]}\right| \leq \max\left(2, \frac{4}{\operatorname{Im}\{z\}}\frac{1}{N}\operatorname{Tr}\left[\mathbf{R}\right]\right). \tag{A.15}$$

# Appendix B

# Combinatorics of Set Partitions

In this appendix, we recall some results regarding the problem of combinatorial enumeration on lattices of set partitions.

Given a finite set $[n] = \{1, \ldots, n\}$, we will be interested in the enumeration of the set of all partitions, denoted henceforth by $P(n)$, as well as a specific subset of $P(n)$, namely the set of non-crossing partitions $NC(n)$ [Kre72]. Both the set of all partitions and the set of non-crossing partitions have a lattice structure when partially ordered (see further [Sta97, Cam94] for definitions on set partitions and partial ordered sets). Some notational aspects regarding the definition of a generic partition are introduced next. We will note a partition as $\pi$ , $\{B_1, \ldots, B_k\} \in L(n)$, where $L(n)$ is one of the two lattices defined above, and say that the partition $\pi$ consists of $|\pi| = k$ non-empty blocks with cardinalities $|B_l| = i_l$, $l = 1, \ldots, k$. From the number of blocks total, up to $k$ blocks with different cardinality may be identified. The type of a partition is specified by defining, for each $i \in \mathsf{Z}^+$, the number $m_i$ of blocks having cardinality $i$. Thus, if $\pi$ has $k$ non-empty blocks, the following equalities hold true

$$m_1 + m_2 + \ldots + m_n = k \tag{B.1}$$

$$1m_1 + 2m_2 + \ldots + nm_n = n. \tag{B.2}$$

For our purposes, it will also be of interest to define the set of possibly different types of a partition of an $n$-set characterized by having $k$ blocks, denoted by $T(k, n)$. The elements of $T(k, n)$ can be found as the solutions to the system of linear Diophantine equations defined by (B.1) and (B.2). Using $m_1 = k - (m_2+, \ldots, m_n)$ and substituting in (B.2), the problem is reduced to finding the solutions to the following (non-negative) integer equation, namely,

$$1m_2 + 2m_3 + \ldots + (n-1)\, m_n = n - k,$$

which are given by the partitions of the integer $k - m$. Note that $m_l = 0$, for $l > n - k + 1$. Accordingly, the number of elements in $T(k, n)$ equals the number of integer partitions $p(n-k)$, with $p(1) = 1$, $p(2) = 2$, $p(3) = 3$, $p(4) = 5$, $p(5) = 7$, $p(6) = 11$, ….

In particular, the number of elements in $P(n)$ is given by the Bell number $B(n)$, which satisfies the recursion

$$B(n+1) = \sum_{k=0}^{n} \binom{n}{k} B(k).$$

Furthermore, the number of $k$-block partitions of an $n$-set is given by the Stirling number of second kind, denoted here as $S(k,n)$, and defined in closed-form as

$$S(k,n) = \frac{1}{k!} \sum_{j=1}^{k} (-1)^{k-j} \binom{k}{j} j^n.$$

Clearly, the following identity must hold, namely, $B(n) = \sum_{k=1}^{n} S(k,n)$. Additionally, when the partition type $t = \{m_1, m_2, \ldots, m_n\} \in T(k,n)$ is specified, the number partitions of the same type is delivered by the Faa di Bruno's coefficient, namely,

$$\xi_P(t) = \frac{m!}{m_1! m_2! \cdots m_n! 1!^{m_1} 2!^{m_2} \cdots k!^{m_n}}.$$

Indeed, observe that

$$S(k,n) = \sum_{t \in T(k,n)} \xi_P(t),$$

where the sum runs over all the elements in $T(k,n)$.

The previous combinatorial characterization of the lattice of all partitions of a set allows us to write the expression for the higher-order derivatives of a composite function (Faa di Bruno's formula) in a compact form as

$$\frac{\partial^n}{\partial x^n} \{f(g(x))\} = \sum_{\pi \in P(n)} f^{(|\pi|)}(g(x)) \prod_{B \in \pi}^{n} g^{(|B|)}(x) \tag{B.3}$$

$$= \sum_{k=1}^{n} f^{(k)}(g(x)) \sum_{t \in T(n,k)} \xi_P(t) \prod_{i=1}^{n} \left(g^{(i)}(x)\right)^{m_i}. \tag{B.4}$$

Regarding the lattice of non-crossing partitions, the total number of elements in $NC(n)$ is counted by the Catalan number, defined by

$$C(n) = \frac{1}{n+1} \binom{2n}{n}. \tag{B.5}$$

Moreover, the number of partitions with $k$ blocks is given by the Narayana number [Kre72] as

$$N(k,n) = \frac{1}{n} \binom{n}{k} \binom{n}{k-1}. \tag{B.6}$$

The number of non-crossing partitions of a certain type can be found as

$$\xi_{NC}(t) = \frac{n!}{(n-k+1)!} \frac{1}{m_1! m_2! \cdots m_n!}. \tag{B.7}$$

Again, the following identity must hold, namely, $C(n) = \sum_{k=1}^{n} N(k,n)$, and we observe that

$$N(k,n) = \sum_{t \in T(k,n)} \xi_{NC}(t).$$

As a last remark, in (B.6) and (B.7), blocks with equal number of elements are indistinguishable. In the case blocks are considered distinguishable, the number of partitions with $k$ blocks from a set of $n$ elements is

$$N_{dis}(k,n) = \frac{n!}{(n-k+1)!}. \tag{B.8}$$

Next, we recall two results regarding the operation with formal power series that will be of special interest for the derivations in the sequel. In particular, the following lemma and subsequent theorem give the coefficients of a composition of two formal power series resp. a compositional inverse[1] in terms of the coefficients of the two series (see e.g. [Sta97, Gou83]).

**Lemma B.1** *Define the formal power series* $f(z) = 1 + \sum_{n=1}^{\infty} \frac{f_n}{n!} z^n$, $g(z) = \sum_{n=1}^{\infty} \frac{g_n}{n!} z^n$ *and* $h(z) = 1 + \sum_{n=1}^{\infty} \frac{h_n}{n!} z^n$, *such that* $f(g(z)) = h(z)$. *Then, the coefficients of the formal power series composition* $h(z)$ *can be obtained as*

$$h_n = \sum_{k=1}^{n} f_n \sum_{t \in T(k,n)} \xi_P(t)\; g_1^{l_1} g_2^{l_2} \cdots g_n^{l_k}. \tag{B.9}$$

In effect, when considered as the (convergent) series expansion of a function rather than a formal power series, the coefficients $h_n$ are clearly the $n$th derivative of the composite function $h(z) = f(g(z))$ evaluated at $z = 0$. The later observation can be used to derive the combinatorial interpretation of the classical moment-cumulant formula. Indeed, the coefficients of the moments as polynomials in the cumulants are precisely those occurring in the Faà di Bruno's formula in (B.3).

**Theorem B.1** *(Lagrange inversion formula) Consider the formal power series* $f(z) = \sum_{n\geq 1} f_n z^n$, *with* $f_1 \neq 0$, *and let* $g(z) = \sum_{n\geq 1} g_n z^n$ *be its inverse for composition, i.e.,* $f(g(z)) = g(f(z)) = z$. *Then,*

$$g_n = [z^n]\{g(z)\} = \left[z^{-1}\right]\left\{\frac{1}{nf(z)^n}\right\}, \tag{B.10}$$

*where* $\left[z^{-l}\right]$ *denotes the operator extracting the coefficient of* $z^{-l}$ *in a series expansion.*

**Proof.** Since $g(f(z)) = z$, we may write

$$1 = \frac{\partial}{\partial z}\left\{\sum_{k\geq 1} g_k f(z)^k\right\} = \sum_{k\geq 1} g_k k f(z)^{k-1} f'(z).$$

[1] See [Hen74] for the algebraic definition of formal power series, and the existence (and uniqueness) of their functional composition and the compositional inverse.

On the other hand, we equivalently have

$$\frac{1}{nf(z)^n} = \sum_{k\geq 1} \frac{k}{n} g_k f(z)^{k-1-n} f'(z).$$

Further, observe that

$$\left[z^{-1}\right]\left\{\frac{1}{nf(z)^n}\right\} = \sum_{k\geq 1} \frac{k}{n} g_k \ \left[z^{-1}\right]\left\{f(z)^{k-1-n} f'(z)\right\}. \tag{B.11}$$

Now, evaluating the argument of the coefficient extraction operator as

$$f(z)^{k-1-n} f'(z) = \begin{cases} \frac{1}{k-n}\frac{\partial}{\partial z}\left\{f(z)^{k-n}\right\}, & k \neq n \\ \frac{f'(z)}{f(z)}, & k = n, \end{cases}$$

and using

$$\left[z^{-1}\right]\left\{h'(z)\right\} = 0 \tag{B.12}$$

$$\left[z^{-1}\right]\left\{h'(z)/h(z)\right\} = m, \tag{B.13}$$

for $h(z)$ a formal Laurent series $h(z) = \sum_{n\geq m} h_n z^n$, $m \in \mathsf{Z}$, $h_m \neq 0$, we find that

$$\left[z^{-1}\right]\left\{f(z)^{k-1-n} f'(z)\right\} = \begin{cases} 0, \ k \neq n \\ 1, k = n. \end{cases} \tag{B.14}$$

The equality in (B.12) is trivial. To see (B.13), we write $h(z) = z^m (h_m + h_{m+1} z + \ldots)$ such that $h^{-1}(z) = z^{-m}(g_0 + g_1 z + \ldots)$ with $g_0 = 1/h_m$. Then, we noting that

$$\frac{h'(z)}{h(z)} = z^{-m}(g_0 + g_1 z + \ldots) h'(z) = z^{-m}(g_0 + g_1 z + \ldots)\left(\sum_{n\geq m} n h_n z^{n-1}\right),$$

we finally get

$$\left[z^{-1}\right]\left\{\frac{h'(z)}{h(z)}\right\} = \left[z^{m-1}\right](g_0 + g_1 z + \ldots) h'(z) = g_0 m h_m = m.$$

By pluging (B.14) in (B.11), we finally obtain the result in the theorem. ■

**Corollary B.1** *Let $f(z)$ be defined by $f(z) = z\phi(f(z))$, with $\phi(0) \neq 0$. Then,*

$$f_n = \left[z^{-n}\right]\{f(z)\} = \frac{1}{n}\left[z^{n-1}\right]\{\phi(z)^n\}.$$

**Proof.** Define the formal power series $\psi(z) = z/\phi(z)$. Since

$$\psi(f(z)) = \frac{f(z)}{\phi(f(z))} = z,$$

we can regard $f(z)$ as the compositional inverse of $\psi(z)$. Thus, by just applying the Lagrange inversion formula in (B.10) we readily obtain

$$[z^n]\{f(z)\} = \left[z^{-1}\right]\left\{\frac{1}{n\psi(z)^n}\right\} = \left[z^{-1}\right]\left\{\frac{\phi(z)^n}{nz^n}\right\} = \frac{1}{n}\left[z^{n-1}\right]\{\phi(z)^n\}.$$

■

# Bibliography


[Abr81] Y.I. Abramovich, and A.I. Nevrev, "An analysis of effectiveness of adaptive maximisation of the signal-to-noise ratio which utilises the inversion of the estimated correlation matrix.", *Radio Engineering and Electronic Physics. Translated from: Radiotekhnika i Elektronika*, vol. 26, no. 12, pp. 67–74, December 1981.

[Ahl78] L. Ahlfors, *Complex Analysis. An introduction to the theory of analytic functions of one complex variable*, McGraw-Hill, New York, 3rd ed., 1978.

[Akh61] N.I. Akhiezer, and I.M. Glazman, *Theory of Linear Operators in Hilbert Spaces*, vol. 1, Frederick Ungar Publishing Co., New York, 1961.

[Akh65] N.I. Akhiezer, *The Classical Moment Problem*, Hafner Publishing Co., New York, 1965.

[And03] T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, 2003.

[Apo73] T. Apostol, *Mathematical Analysis*, Addison-Wesley, Reading, MA, 2nd ed., 1973.

[Ash99] R. Ash, *Probability and Measure Theory*, Academic Press, 2nd ed., 1999.

[Axe94] O. Axelsson, *Iterative Solution Methods*, Cambridge University Press, Cambridge, 1994.

[Bai98] Z.D. Bai, and Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices", *Annals of Probability*, vol. 26, no. 1, pp. 316–345, 1998.

[Bai99] Z. D. Bai, "Methodologies in spectral analysis of large dimensional random matrices, a review", *Statistica Sinica*, vol. 9, pp. 611–677, 1999.

[Bai05] Z.D. Bai, "High dimensional data analysis", *COSMOS*, vol. 37, pp. 17–27, 2005.

[Bai06] Z.D. Bai, and J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, Science Press, Beijing, 2006.

[Bai07] Z.D. Bai, G.M. Pan, and B.Q. Miao, "On asymptotics of eigenvectors of large sample covariance matrix", *accepted for publication in Annals of Probability*, 2007.

[Bar66] Robert G. Bartle, *The Elements of Integration*, John Wiley and Sons, New York, 1966.

[Bec07] A. Beck, and Y.C. Eldar, "Doubly constrained robust capon beamformer with ellipsoidal uncertainty sets", *IEEE Transactions on Signal Processing*, vol. 55, no. 2, pp. 753–758, January 2007.

[Bia03] P. Biane, "Characters of symmetric groups and free cumulants", A.M. Vershik (ed.), *Asymptotic Combinatorics with Applications to Mathematical Physics*, pp. 185–200, Springer, 2003.

[Bil95] P. Billingsley, *Probability and Measure*, Wiley-Interscience, 3rd ed., 1995.

[Bry06] W. Bryc, A. Dembo, and T. Jiang, "Spectral measure of large random Hankel, Markov and Toeplitz matrices", *Annals of Probability*, vol. 34, no. 1, pp. 1–38, 2006.

[Bur73] D.L. Burkholder, "Distribution function inequalities for martingales", *Annals of Probability*, , no. 1, pp. 19–42, 1973.

[Bur02] S. Burykh, and K. Abed-Meraim, "Reduced-rank adaptive filtering using Krylov subspace", *EURASIP Journal on Applied Signal Processing*, vol. 2002, no. 12, pp. 1387–1400, 2002.

[Cam94] P.J. Cameron, *Combinatorics*, Cambridge University Press, 1994.

[Che02] W. Chen, U. Mitra, and P. Schniter, "On the equivalence of three reduced rank linear estimators with applications to DS-CDMA", *IEEE Transactions on Information Theory*, vol. 48, no. 9, pp. 2609–2614, September 2002.

[Cho00] S. Chowdhurry, M.D. Zoltowski, and J.S. Goldstein, "Application of reduced-rank chip-level MMSE equalization to forward link DS-CDMA with frequency selective multipath", *Proceedings of the 38th Annual Allerton Conference on Communications, Systems, and Computing*, October 2000.

[Chu01] K.L. Chung, *A Course in Probability Theory*, Academic Press, San Diego, 3rd ed., 2001.

[Chu05] M.T. Chu, and G.H. Golub, *Inverse Eigenvalue Problems: Theory, Algorithms, and Applications*, Oxford University Press, 2005.

[Col07] M. Coldrey, and P. Bohlin, "Training-based MIMO systems part i: Performance comparison", *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5464–5476, November 2007.

[Cot05] L. Cottatellucci, and R. Müller, "A systematic approach to multistage detectors in multipath fading channels", *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3146–3158, September 2005.

[Die07] G.K. Dietl, *Linear Estimation and Detection in Krylov Subspaces*, vol. 1, Springer, Berlin, 2007.

[Doz07a] R.B. Dozier, and J.W. Silverstein, "Analysis of the limiting spectral distribution of large dimensional information-plus-noise type matrices", *Journal of Multivariate Analysis*, vol. 98, no. 4, pp. 1099–1122, April 2007.

[Doz07b] R.B. Dozier, and J.W. Silverstein, "On the empirical distribution of eigenvalues of large dimensional information-plus-noise type matrices", *Journal of Multivariate Analysis*, vol. 98, no. 4, pp. 678–694, April 2007.

[Dud02] R.M. Dudley, *Real Analysis and Probability*, Cambridge University Press, 2nd ed., 2002.

[Dum07] C. Dumard, F. Kaltenberger, and K. Freudenthaler, "Low-cost approximate LMMSE equalizer based on Krylov subspace methods for HSDPA", *IEEE Transactions on Wireless Communications*, vol. 55, no. 7, pp. 1610–1614, May 2007.

[Dur95] R. Durrett, *Probability: Theory and Examples*, Duxbury Press, Belmont, Calif., 2nd ed., 1995.

[Erm93] V. Ermolaev, and A. Flaksman, "Signal processing in adaptive arrays using power basis", *International Journal of Electronics*, vol. 75, no. 4, pp. 753–765, 1993.

[Erm94] V.T. Ermolaev, and A.B. Gershman, "Fast algorithm for minimum-norm direction-of-arrival estimation", *IEEE Transactions on Signal Processing*, vol. 42, no. 9, pp. 2389–2394, September 1994.

[Eve00] R. Everson, and S. Roberts, "Inferring the eigenvalues of covariance matrices from limited, noisy data", *IEEE Transactions on Signal Processing*, vol. 48, no. 7, pp. 2083–2091, July 2000.

[Fel66] W. Feller, *An Introduction to Probability Theory and Its Applications, Volume II*, John Wiley and Sons, Inc., New York, 1966.

[Ge04] H. Ge, M. Lundberg, and L.L. Scharf, "Warp convergence in reduced-rank conjugate gradient wiener filters", *Proceedings of the Third IEEE Sensor Array and Multichannel Signal Processing Workshop*, July 2004.

[Ge06] H. Ge, I.P. Kirsteins, and L.L. Scharf, "Data dimension reduction using Krylov subspaces: Making adaptive beamformers robust to model order-determination", *Proceedings of the 2006 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2006.

[Ger99] A.B. Gershman, "Robust adaptive beamforming in sensor arrays", *AEU Int. Journal of Electronics and Communications*, vol. 53, no. 6, pp. 305–314, December 1999.

[Ger03] J.S. Geronimo, and T.P. Hill, "Necessary and sufficient condition that the limit of Stieltjes transforms is a Stieltjes transform", *Journal of Approximation Theory*, vol. 121, no. 1, pp. 54–60, March 2003.

[Gir95] V.L. Girko, *Statistical Analysis of Observations of Increasing Dimension*, Kluwer, 1995.

[Gir98] V.L. Girko, *An Introduction to Statistical Analysis of Random Arrays*, VSP, The Netherlands, 1998.

[Gol96] G.H. Golub, and C.F. Van Loan, *Matrix computations*, John Hapkins, Univ. Press, Baltimore, Maryland, 3rd ed., 1996.

[Gol97a] J.S. Goldstein, and I.S. Reed, "Reduced-rank adaptive filtering", *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 492–496, February 1997.

[Gol97b] J.S. Goldstein, and I.S. Reed, "Subspace selection for partially adaptive sensor array processing", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 2, pp. 539–544, April 1997.

[Gol97c] J.S. Goldstein, and I.S. Reed, "Theory of partially adaptive radar", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 4, pp. 1309–1325, October 1997.

[Gol98] J.S. Goldstein, I.S. Reed, and L.L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections", *IEEE Trans. on Information Theory*, vol. 44, pp. 2943–2959, November 1998.

[Gol99] J.S. Goldstein, I.S. Reed, and P.A. Zulch, "Multistage partially adaptive STAP CFAR detection algorithm", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 2, pp. 645–661, April 1999.

[Gol03] A. Goldsmith, S.A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels", *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, June 2003.

[Gou83] I.P. Goulden, and D.M. Jackson, *Combinatorial Enumeration*, John Wiley, New York, 1983.

[Gra01] A. Grant, and C. Schlegel, "Convergence of linear interference cancellation multiuser receivers", *IEEE Transactions on Communications*, vol. 49, no. 10, pp. 1824–1834, October 2001.

[Gra03] P. Graczyk, G. Letac, and H. Massam, "The complex Wishart distribution and the symmetric group", *Annals of Statistics*, vol. 31, no. 1, pp. 287–309, 2003.

[Gue00] J.R. Guerci, J.S. Goldstein, and I.S. Reed, "Optimal and adaptive reduced-rank STAP", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 2, pp. 647–663, April 2000.

[Guo99] D. Guo, L.K. Rasmussen, and T.J. Lim, "Linear parallel interference cancellation in long-code CDMA multiuser detection", *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 12, pp. 2074–2081, December 1999.

[Guo00] D. Guo, L.K. Rasmussen, and T.J. Lim, "A matrix-algebraic approach to linear parallel interference cancellation in CDMA", *IEEE Transactions on Communications*, vol. 48, no. 1, pp. 152–161, January 2000.

[Hac04] W. Hachem, "Simple polynomial detectors for CDMA downlink transmissions on frequency-selective channels", *IEEE Transactions on Information Theory*, vol. 50, no. 1, pp. 164–171, January 2004.

[Hac07] W. Hachem, P. Loubaton, and J. Najim, "Deterministic equivalents for certain functionals of large random matrices", *Annals of Applied Probability*, vol. 17, no. 3, pp. 875–930, 2007.

[Hal05] P. Hall, J.S. Marron, and A. Neeman, "Geometric representation of high dimension, low sample size data", *Journal Of The Royal Statistical Society Series B*, vol. 67, no. 3, pp. 427–444, 2005.

[Han97] C.P. Hansen, *Rank-deficient and discrete ill-posed problems*, SIAM, Philadelphia, PA, 1997.

[Has03] B. Hassibi, and B.M. Hochwald, "How much training is needed in multiple-antenna wireless links?", *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, April 2003.

[Hen74] P. Henrici, *Applied and Complex Computational Analysis, Vol. 1*, Wiley, New York, 1974.

[Hia00] F. Hiai, and D. Petz, *The semicircle law, free random variables and entropy*, American Mathematical Society, 2000.

[Hil62] E. Hille, *Analytic Function Theory, Vol. II*, Ginn and Company, Boston, 1962.

[Hon95] M.J. Honig, U. Madhow, and S. Verdú, “Blind adaptive multiuser detection”, *IEEE Transactions on Information Theory*, vol. 41, pp. 944–960, July 1995.

[Hon01] M.J. Honig, and W. Xiao, “Performance of reduced-rank linear interference suppression”, *IEEE Transactions on Information Theory*, vol. 47, no. 5, pp. 1938–1946, July 2001.

[Hon02] M.J. Honig, and J.S. Goldstein, “Adaptive reduced-rank interference suppression based on the multistage Wiener filter”, *IEEE Transactions on Communications*, vol. 50, no. 6, pp. 986–994, June 2002.

[Hon06] M.L. Honig, and Y. Sun, “Performance of reduced-rank equalization”, *IEEE Transactions on Information Theory*, vol. 52, no. 10, pp. 4548–4562, October 2006.

[Hor73] L. Hormander, *An Introduction to Complex Analysis in Several Variables*, North Holland Publishing Co., Amsterdam, 1973.

[Hor85] R. Horn, and C.R. Johnson, *Matrix analysis*, Cambridge Univ. Press, 1985.

[Hot46] H. Hotelling, “Analysis of a complex of statistical variables into principal components”, *Journal of Educational Psychology*, vol. 24, pp. 417–441, 488–520, 1946.

[Hud81] J.E. Hudson, *Adaptive Array Principles*, Peter Peregrinus Ltd., London, 1981.

[Joh93] D.H. Johnson, and D.E. Dudgeon, *Array Signal Processing: Concepts and Techniques*, Prentice-Hall PTR, Englewood Cliffs, NJ, 1993.

[Joh00] M. Joham, and M. D. Zoltowski, “Interpretation of the multi-stage nested Wiener filter in the Krylov subspace framework”, Tech. rep. tum-lns-tr-00-6, Munich University of Technology, 2000.

[Joh07] I.M. Johnstone, “High dimensional statistical inference and random matrices”, *Proc. of International Congress of Mathematicians*, vol. I, pp. 307–333, Eur. Math. Soc., Zurich, 2007.

[Kar46] K. Karhunen, “Zur Spektraltheorie stochastischer Prozesse”, *Ann. Acad. Sci. Fennicae*, vol. 37, 1946.

[Kar07] N. El Karoui, “Spectrum estimation for large dimensional covariance matrices using random matrix theory”, *Annals of statistics, accepted for publication*, 2007.

[Kay93] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, 1993.

[Kol05] T. Kollo, and D. von Rosen, *Advanced Multivariate Statistics with Matrices*, Springer, 2005.

[Kre72] G. Kreweras, "Sur les partitions non croisees d'un cycle", *Discrete Mathematics*, vol. 1, no. 4, pp. 333–350, 1972.

[Kre77] M.G. Krein, and A.A. Nudelman, *The Markov Moment Problem and Extremal Problems*, vol. 50, American Mathematical Society, Providence, RI, transl. math. monographs ed., 1977.

[Lag86] M. Lagunas, M. Santamaria, A. Gasull, , and A. Moreno, "Maximum likelihood filters in spectral estimation problems", *Signal Processing*, vol. 10, pp. 19–34, 1986.

[Lag08] M.A. Lagunas, M.A. Rojas, and P. Stoica, "New spectral estimation based on filterbank for spectrum sensing", *Proc. of 33rd IEEE International Conference on Acoustics, Speech, and Signal Processing*, March 2008.

[Lax02] P.D. Lax, *Functional analysis*, vol. 1, John Wiley and Sons, New York, 2002.

[Li01] L. Li, A. Tulino, and S. Verdú, "Asymptotic eigenvalue moments for linear multiuser detection", *Communications in Information and Systems*, vol. 1, no. 3, pp. 273–304, March 2001.

[Li03] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading", *IEEE Trans. on Signal Processing*, vol. 51, no. 7, pp. 1702–1715, July 2003.

[Li04a] J. Li, P. Stoica, and Z. Wang, "Doubly constrained robust capon beamformer", *IEEE Transactions on Signal Processing*, vol. 52, no. 9, pp. 2407–2423, September 2004.

[Li04b] L. Li, A. Tulino, and S. Verdú, "Design of reduced-rank MMSE multiuser detectors using random matrix methods", *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 986–1008, June 2004.

[Li05] J. Li, and P. Stoica (eds.), *Robust Adaptive Beamforming*, Wiley-Interscience, John Wiley and Sons, 2005.

[Lor05] R.G. Lorenz, and S.P. Boyd, "Robust minimum variance beamforming", *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1684–1696, May 2005.

[Lou03] P. Loubaton, and W. Hachem, "Asymptotic analysis of reduced rank Wiener filters", *Proceedings of Information Theory Workshop 2003*, pp. 328–331, March 2003.

[Lov04] D.J. Love, Jr. R.W. Heath, W. Santipach, and M. L. Honig, "What is the value of limited feedback for mimo channels?", *IEEE Commun. Mag.*, vol. 42, no. 10, pp. 54–59, October 2004.

[Lue84] D.G. Luenberger, *Linear and Nonlinear Programming*, Addison-Wesley, MA, 2nd ed., 1984.

[Mad94] U. Madhow, and M.J. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA", *IEEE Transactions on Communications*, vol. 42, pp. 317–3188, December 1994.

[Mad98] U. Madhow, "Blind adaptive interference suppression for direct-sequence CDMA", *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2049–2069, October 1998.

[Mai00] D. Maiwald, and D. Kraus, "Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices", *IEE Proceedings - Radar, Sonar and Navigation*, vol. 147, no. 4, pp. 162–168, August 2000.

[Mar67] V.A. Marçenko, and L.A. Pastur, "Distribution of eigenvalues for some sets of random matrices", *Mathematics of the USSR - Sbornik*, vol. 1, no. 4, pp. 457–483, 1967.

[Mar83] T.L. Marzetta, "A new interpretation for capon's maximum likelihood method of frequency-wavenumber spectral estimation", *IEEE Transactions on Signal Processing*, vol. 31, no. 2, pp. 445–449, April 1983.

[Mar99a] J.E. Marsden, and M.J. Hoffman, *Basic Complex Analysis*, W.H. Freeman, New York, 1999.

[Mar99b] T. Marzetta, "BLAST training: Estimating channel characteristics for high-capacity space-time wireless", *Proc. 37th Annual Allerton Conference on Communications, Control, and Computing*, September 1999.

[Mar99c] T.L. Marzetta, and B.M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading", *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139–157, January 1999.

[Mas93] J. L. Massey, and Th. Mittelholzer, "Welch's bound and sequence sets for code-division multiple access systems", R. Capocelli, A. De Santis, U. Vaccaro (eds.), *in Sequences II, Methods in Commun., Security and Computer Science*, pp. 200–266, Springer-Verlag, New York, 1993.

[McC02] M.L. McCloud, and L.L. Scharf, "A new subspace identification algorithm for high-resolution doa estimation", *IEEE Transactions on Signal Processing*, vol. 50, no. 10, pp. 1382–1390, October 2002.

[Mes05] X. Mestre, and M.A. Lagunas, "Diagonal loading for finite sample size beamforming: an asymptotic approach", J. Li, P. Stoica (eds.), *Robust Adaptive Beamforming*, pp. 200–266, Wiley-Interscience, John Wiley and Sons, 2005.

[Mes06a] X. Mestre, "Improved estimation of eigenvalues of covariance matrices and their associated subspaces using their samples estimates", *submitted for publication to IEEE Transactions on Information Theory*, July 2006.

[Mes06b] X. Mestre, "On the asymptotic behavior of quadratic forms of the resolvent of certain covariance-type matrices.", Tech. Rep. CTTC/RC/2006-01, CTTC, 2006.

[Mes06c] X. Mestre, and M.A. Lagunas, "Finite sample size effect on MV beamformers: optimum diagonal loading factor for large arrays", *IEEE Trans. on Signal Processing*, vol. 54, no. 1, pp. 69–82, January 2006.

[Mes06d] X. Mestre, and F. Rubio, "An improved subspace-based algorithm in the small sample size regime", *Proc. of the IEEE International Conf. on Acoustics, Speech and Signal Processing, Toulouse (France), May 2006.*, May 2006.

[Mes07] X. Mestre, B.A. Johnson, and Y.I. Abramovich, "Source power estimation for array processing applications under low sample size constraints", *Proc. of 32nd IEEE International Conference on Acoustics, Speech and Signal Processing*, April 2007.

[Mey00] C.D. Meyer, *Matrix analysis and applied linear algebra*, SIAM, 2000.

[Mos96] S. Moshavi, E.G. Kanterakis, and D.L. Schilling, "Multistage linear receivers for DS-CDMA systems", *International Journal of Wireless Information Networks*, vol. 3, pp. 1–17, January 1996.

[Mou07] B. Mouhouche, P. Loubaton, and W. Hachem, "Asymptotic analysis of chip level MMSE equalizers in the downlink of CDMA systems", *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 3048–3060, June 2007.

[Mui82] R.J. Muirhead, *Aspects of Multivariate Statistical Theory*, John Wiley, New York, 1982.

[Mun00] J. Munkres, *Topology*, Prentice-Hall, Upper Saddle River, NJ, 2000.

[Myr00] W.L. Myrick, M.D. Zoltowski, and J.S. Goldstein, "Low-sample performance of reduced-rank power minimization based jammer suppression for GPS", *Proceedings of the IEEE Sixth International Symposium on Spread Spectrum Techniques and Applications*, pp. 93–97, September 2000.

[Mül01] R. Müller, and S. Verdú, "Design and analysis of low-complexity interference mitigation on vector channels", *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 8, pp. 1429–1441, August 2001.

[Nic06] A. Nica, and R. Speicher, *Lectures on the Combinatorics of Free Probability*, Cambridge University Press, 2006.

[Ora97] F. Oravecz, and D. Petz, "On the eigenvalue distribution of some symmetric random matrices", *Acta Sci. Math. (Szeged)*, vol. 63, pp. 383–395, 1997.

[Ott93] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample ml techniques for parameter estimation and detection in array processing", Litva Haykin, Shepherd (eds.), *Radar Array Processing*, pp. 99–151, Springer-Verlag, Berlin, 1993.

[Pad97] D.A. Pados, and S.N. Batalama, "Low-complexity blind detection of DS-CDMA signals: Auxiliary-vector receivers", *IEEE Transactions on Communications*, vol. 45, pp. 1586–1594, December 1997.

[Pad99] D.A. Pados, and S.N. Batalama, "Joint space-time auxiliary vector filtering for DS-CDMA systems with antenna arrays", *IEEE Transactions on Communications*, vol. 47, pp. 1406–1415, 1999.

[Pad01] D.A. Pados, and G.N. Karystinos, "An iterative algorithm for the computation of the MVDR-filter", *IEEE Transactions on Signal Processing*, vol. 49, no. 2, pp. 290–300, February 2001.

[Pan07] G. Pan, M. Guo, and Y.-C. Liang, "Asymptotic performance of reduced-rank linear receivers with principal component filter", *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 1148–1151, March 2007.

[Pap91] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, $3^{\underline{\text{rd}}}$ ed., 1991.

[Par80] B.N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, NJ, 1980.

[Pau07] D. Paul, and J.W. Silverstein, "No eigenvalues outside the support of limiting empirical spectral distribution of a separable covariance matrix", *submitted*, 2007.

[Pec00] C.D. Peckham, A.M. Haimovich, T.F. Ayoub, J.S. Goldstein, and I.S. Reed, "Reduced-rank STAP performance analysis", *IEEE Trans. on Aerospace and Electronic Systems*, vol. 36, no. 2, pp. 664–676, April 2000.

[Poo98] H.V. Poor, and G.W. Wornell, *Wireless Communications: Signal Processing Perspectives*, Prentice Hall, 1998.

[Qia03] H. Qian, and S.N. Batalama, "Data record-based criteria for the selection of an auxiliary vector estimator of the MMSE/MVDR filter", *IEEE Transactions on Communications*, vol. 51, no. 10, pp. 1700–1708, October 2003.

[Rao73] C.R. Rao, *Linear Statistical Inference and Its Applications*, John Wiley, $2^{\underline{\text{nd}}}$ ed., 1973.

[Rao07] N.R. Rao, J. Mingo, R. Speicher, and A. Edelman, "Statistical eigen-inference from large Wishart matrices", *Annals of Statistics, accepted for publication*, 2007.

[Roh76] V.K. Rohatgi, *An Introduction to Probability Theory and Mathematical Statistics*, John Wiley and Sons, New York, 1976.

[Rub06] F. Rubio, and X. Mestre, "Analysis of multi-stage receivers under finite sample-support", *Proc. of 31st IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 2006.

[Rub08] F. Rubio, D. Guo, M. Honig, and X. Mestre, "On optimal training and beamforming in uncorrelated MIMO systems with feedback", *Proc. 42nd Annual Conference on Information Sciences and Systems*, March 2008.

[Rud87] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, New York, $3^{\underline{\text{rd}}}$ ed., 1987.

[Rup94] M. Rupf, and J.L. Massey, "Optimum sequence multisets for synchronous code-division multiple-access channels", *IEEE Transactions on Information Theory*, vol. 40, pp. 1261–1266, July 1994.

[Saa96] Y. Saad, *Iterative Methods for Sparse Linear Systems*, PWS, 1996.

[Sam03] D. Samardzija, and N. Mandayam, "Pilot-assisted estimation of MIMO fading channel response and achievable data rates", *IEEE Transactions on Signal Processing*, vol. 51, no. 11, pp. 2882–2890, November 2003.

[San03] E.L. Santos, and M.D. Zoltowski, "Matrix conjugate gradients for the generation of high-resolution spectrograms", *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, November 2003.

[San07] E.L. Santos, M.D. Zoltowski, and M. Rangaswamy, "Indirect dominant mode rejection: A solution to low sample support beamforming", *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3283–3293, July 2007.

[Sch91] L.L. Scharf, *Statistical Signal Processing*, Addison-Wesley, New York, 1991.

[Sch98] L. Scharf, and J.K. Thomas, "Wiener filters in canonical coordinates for transform coding, filtering, and quantizing", *IEEE Transactions on Signal Processing*, vol. 46, no. 3, pp. 647–654, March 1998.

[Sch00] M.K. Schneider, and A.S. Willsky, "Krylov subspace estimation", *SIAM Journal on Scientific Computing*, vol. 22, no. 5, pp. 1840–1864, October 2000.

[Sch02] L.L. Scharf, and M.L. McCloud, "Blind adaptation of zero forcing projections and oblique pseudo-inverses for subspace detection and estimation when interference dominates noise", *IEEE Transactions on Signal Processing*, vol. 50, no. 12, pp. 2938–2946, December 2002.

[Sch03a] L.L. Scharf, L.T. McWhorter, E.K.P. Chong, J.S. Goldstein, and M.D. Zoltowski, "Algebraic equivalence of conjugate direction and multistage Wiener filters", *Proc. of Workshop on ASAP*, March 2003.

[Sch03b] M.K. Schneider, and A.S. Willsky, "A Krylov subspace method for covariance approximation and simulation of random processes and fields", *Multidimensional Systems and Signal Processing*, vol. 14, no. 4, pp. 295–318, October 2003.

[Sch06] P.J. Schreier, and L.L. Scharf, "Canonical coordinates for transform coding of noisy sources", *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 235–243, January 2006.

[Ser80] R. Serfling, *Approximation theorems in mathematical statistics*, Wiley, 1980.

[Sha03] S. Shabazpanahi, A. Gershman, Z.-Q. Luo, and K.M. Wong, "Robust adaptive beamforming for general-rank signal models", *IEEE Trans. on Signal Processing*, vol. 51, no. 9, pp. 2257–2269, September 2003.

[Sil85] J.W. Silverstein, "The limiting eigenvalue distribution of a multivariate F matrix", *SIAM Journal on Mathematical Analysis*, vol. 16, no. 3, pp. 641–646, 1985.

[Sil86] J.W. Silverstein, "Eigenvalues and eigenvectors of large dimensional sample covariance matrices", *Random Matrices and their Applications, Contemporary Mathematics*, vol. 50, pp. 153–159, 1986.

[Sil95a] J.W. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices", *Journal of Multivariate Analysis*, vol. 55, pp. 331–339, August 1995.

[Sil95b] J.W. Silverstein, and Z.D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices", *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, February 1995.

[Sil95c] J.W. Silverstein, and S.-I. Choi, "Analysis of the limiting spectral distribution of large dimensional random matrices", *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 295–309, August 1995.

[Spe95] R. Speicher, and R. Woroudi, "Boolean convolution", D. Voiculescu (ed.), *Proceedings of Workshop on Random Matrices and Operator Algebra Free Products*, vol. 12, pp. 267–279, American Mathematical Society, Toronto, 1995.

[Spr79] M.D. Springer, *The Algebra of Random Variables*, John Wiley and Sons, New York, 1979.

[Sri07] M.S. Srivastava, "Multivariate theory for analyzing high dimensional data", *Journal of Japan Sta. Soc.*, vol. 37, pp. 53–86, 2007.

[Sta97] R.P. Stanley, *Enumerative Combinatorics, Vols. I and II*, Cambridge University Press, 1997.

[Sto03] P. Stoica, Z. Wang, and J. Li, "Robust capon beamforming", *IEEE Signal Processing Letters*, vol. 10, no. 6, pp. 172–175, June 2003.

[Sto05] P. Stoica, and R.L. Moses, *Spectral Analysis of Signals*, vol. 1, Prentice Hall, 2005.

[Str96] P. Strobach, "Low-rank adaptive filters", *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 2932–2947, December 1996.

[Tre97] L.L. Trefethen, and D. Bau III, *Numerical Linear Algebra*, SIAM, 1997.

[Tre02] H.L. Van Trees, *Optimum Array Processing*, John Wiley and Sons, New York, NY, USA, 2002.

[Tri00] L.G.F. Trichard, I.B. Colling, and J.S. Evans, "Parameter selection for multiuser receivers based on iterative methods", *Proceedings of the IEEE 51st Vehicular Technology Conference*, vol. 2, pp. 926–930, May 2000.

[Tri05] L.G.F. Trichard, J.S. Evans, and I.B. Collings, "Optimal multistage linear multiuser receivers", *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 1092–1101, May 2005.

[Tuf82] D.W. Tufts, R. Kumaresan, and I. Kirsteins, "Data adaptive signal estimation by singular value decomposition of a data matrix", *Proceedings of the IEEE*, vol. 70, no. 6, pp. 684–685, June 1982.

[Tuf93] D.W. Tufts, and A.A. Shah, "Estimation of a signal waveform from noisy data using low-rank approximation to a data matrix", *IEEE Transactions on Signal Processing*, vol. 41, no. 4, pp. 1716–1721, April 1993.

[Tul04] A. Tulino, and S. Verdú, "Random matrix theory and wireless communications", S. Verdú (ed.), *Foundations and Trends in Communications and Information Theory*, vol. 1, NOW Publishers, June 2004.

[Ver98] S. Verdú, *Multiuser Detection*, Cambridge University Press, NY, 1998.

[Voi87] D. Voiculescu, “Multiplication of certain noncommuting random variables”, *Journal of Operator Theory*, vol. 18, no. 2, pp. 223–235, 1987.

[Voi92] D.V. Voiculescu, K. Dykema, and A. Nica, *Free random variables*, vol. 1, CRM Monograph Series. American Mathematical Society, 1992.

[Vor03a] S. Vorobyov, A. Gershman, and Z.-Q. Luo, “Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem”, *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 313–324, February 2003.

[Vor03b] H. van der Vorst, *Iterative Krylov Methods for Large Linear systems*, Cambridge University Press, Cambridge, 2003.

[Wan98a] X. Wang, and H.V. Poor, “Blind equalization and multiuser detection in dispersive CDMA channels”, *IEEE Transactions on Communications*, vol. 46, no. 1, pp. 91–103, January 1998.

[Wan98b] X. Wang, and H.V. Poor, “Blind multiuser detection: A subspace approach”, *IEEE Transactions on Information Theory*, vol. 44, pp. 677–690, March 1998.

[Wei02] M.E. Weippert, J.D. Hiemstra, J.S. Goldstein, and M.D. Zoltowski, “An improved subspace-based algorithm in the small sample size regime”, *Proceedings of the Second IEEE Sensor Array and Multichannel Signal Processing Workshop*, August 2002.

[Wel74] L.R. Welch, “Lower bounds on the maximum cross correlation of signals”, *IEEE Transactions on Information Theory*, vol. 20, pp. 397–399, May 1974.

[Wig67] E.P. Wigner, “Random matrices in physics”, *SIAM Review*, vol. 9, no. 1, pp. 1–23, January 1967.

[Wil88] J.H. Wilkinson, *The algebraic eigenvalue problem*, Oxford University Press, New York, NY, USA, 1988.

[Xia05] W. Xiao, and M.L. Honig, “Large system transient analysis of adaptive least squares filtering”, *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2447–2474, July 2005.

[Xu90] G. Xu, and T. Kailath, “Fast signal-subspace decomposition without eigendecomposition”, *Conference Record Twenty-Fourth Asilomar Conference on Signals, Systems and Computers*, November 1990.

[Xu94] G. Xu, and T. Kailath, “Fast estimation of principal eigenspace using Lanczos algorithm”, *SIAM Journal on Matrix Analysis and Applications*, vol. 15, no. 3, pp. 974–994, July 1994.

[YH01] M. Nikpour Y. Hua, and P. Stoica, "Optimal reduced-rank estimation and filtering", *IEEE Transactions on Signal Processing*, vol. 49, no. 3, pp. 457–469, March 2001.

[Yin86] Y.Q. Yin, "Limiting spectral distribution for a class of random matrices", *Journal of Multivariate Analysis*, vol. 20, no. 1, pp. 50–68, October 1986.

[Zhe02] L. Zheng, and D.N.C. Tse, "Communication on the Grassmann manifold: a geometric approach to the noncoherent multiple-antenna channel", *IEEE Trans. on Information Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[Zul98] P.A. Zulch, J.S. Goldstein, J.R. Guerci, and I.S. Reed, "Comparison of reduced-rank signal processing techniques", *Conference Record of the Thirty-Second Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 421–425, November 1998.