Random Matrix Theory Tutorial – Introduction to Deterministic Equivalents

Axel Müller [1], **Mérouane Debbah** [2]

Mathematical and Algorithmic Sciences Lab
France Research Center, Huawei Technologies Co. Ltd.
Arcs de Seine Bâtiment B, 18 Quai du Point du Jour
92100 Boulogne-Billancourt, France

{axel.mueller,merouane.debbah}@huawei.com

RÉSUMÉ. *L'ensemble des consignes rassemblées ci-dessous s'organise en plusieurs rubriques. La rédaction remercie les auteurs pour le strict respect qu'ils accorderont à ces dispositions.*

ABSTRACT. *In the following we provide a tutorial on the practical application of random matrix theory (RMT) in communication problems, in order to facilitate utilization of the analytic RMT approach for interested researchers. To this end, we first state the necessary basic theoretical concepts, lemmas and tools from RMT. After this we will build intuition, confidence, and insight into RMT concepts and their applications, by putting the introduced theoretical results into a tutorial like context. To familiarize the reader with the introduced tools we rely on an example of a step by step derivation of the deterministic equivalent for a non-trivial rate problem. Thus, we provide the theory needed to soundly use the framework of RMT in the design of future large scale (w.r.t. the numbers of users, base stations and antennas) communication networks.*

MOTS-CLÉS : *sur trois lignes maximum. un ensemble de mots significatifs doit être isolé sous forme de mots-clés (et terminés par un point).*

KEYWORDS: *Random Matrix Theory, Deterministic Equivalent, Communication Technology, Tutorial, MIMO.*

DOI:10.3166/TS.x.1-25

Extended Abstract (in English)

In order to satisfy increasing throughput and other performance demands, future networks will need to adopt new approaches of network design, which ultimately result in very large systems (w.r.t. the numbers of users, base stations (BSs) and antennas). Most of the analytic tools used today in the communications community were originally developed for the analysis of point-to-point communication or small MIMO systems. Therefore, it is not a surprise that they often fail to provide meaningful insight into possible solutions for a new era of large dense heterogeneous multi user multi cell communication systems. New tools, which are adapted to the large nature of the systems, need to be developed and applied to give insight into and find the right combination of future system design solutions. Fortunately, the mathematical tool of large scale random matrix theory (RMT) has matured in recent years, to a point where it is of excellent use in this task. Publications documenting this maturing process (e.g., (Bai, Silverstein, 1998b ; Tulino, Verdú, 2004 ; Hachem, Khorunzhy *et al.*, 2008 ; Bai, Silverstein, 2009 ; Couillet, Debbah, 2011 ; Hoydis, 2012)) are numerous and they are not necessarily adapted to helping interested researchers to enter the field. This article introduces the theory needed to soundly use the framework of RMT, especially the technique known as *deterministic equivalent* (DE) approach, in the form of a tutorial and adheres to a predominately pedagogical style.

DEs are mathematical objects that are able to provide accurate deterministic approximations of random quantities, which represent important system performance indicators in cellular networks. For example, the capacity of large dimensional multi antenna channels. We define the DE of a sequence of random quantities as follows:

DÉFINITION 1 (Deterministic Equivalent (Hachem *et al.*, 2007)). — *The deterministic equivalent of a sequence of random complex values $(X_n)_{n\geq 1}$ is a deterministic sequence $(\overline{X}_n)_{n\geq 1}$, which approximates X_n such that*

$$X_n - \overline{X}_n \xrightarrow[n\to+\infty]{\text{a.s.}} 0$$

where $a.s.$ is taken to mean almost sure convergence.

Quite often the quantity X_n is going to be a functional of the resolvent of a Hermitian matrix. Also, even relatively simple problems can result in a DE $\overline{X}_n$, which is not guaranteed to converge itself. Yet, it is possible to deterministically calculate a DE $\overline{X}_n$. Intuitively, one can remark that $\overline{X}_n$ is still "contains" the factor n, even as $n \to \infty$. In fact, the DE gives us an approximation for each value of N, which becomes more precise for increasing N. The realizations of the random variable, almost surely ($a.s.$) fall within a increasingly narrow bound around the DE. Thus, DEs tend to be much more accurate for finite (and even small) system dimensions than the classical limits. Assume we analyse a random quantity involving a random matrix $\mathbf{X} \in \mathbb{C}^{M\times K}$. RMT handles cases in which both dimensions (M, K) grow large, while classic limit approximations (e.g., the strong law of large numbers) can only treat the case where M grows large. As a consequence, RMT results exploit more degrees of freedom than classical approaches and, thus, usually far outperform them w.r.t. convergence speed.

For example, even an 8×8 matrix offers already up to 64 degrees of freedom, which mostly leads to quite acceptable convergence. In general, RMT achieves impressive convergence rates for linear functionals of eigenvalues, e.g., for central limit theorems in $1/M$ (i.e., $M(X_M - \overline{X}_M) \to \mathcal{N}(0,1)$) (Hachem, Loubaton, Najim, 2008, Remark 3.1) and for expectations in $1/M^2$ (i.e., $\mathbb{E}[X_M] = \overline{X}_M + \mathcal{O}(1/M^2)$) (Pastur, Shcherbina, 2011)(Vallet *et al.*, 2010, Proposition 4), when the random quantities are complex Gaussian distributed. Quadratic forms are usually slower, for example central limit theorems in $1/\sqrt{M}$ (Kammoun *et al.*, 2009, Theorem 3) and expectations in $1/M^{\frac{3}{2}}$ (Rubio *et al.*, 2012, Lemma 1). The well-known RMT Lemmas 15, 17, 20 and 22 are of frequent use in finding such DEs in the first place. Practical usage of DEs is very intuitive (i.e., summation and multiplication behave as one would expect), as is shown by the results stated in Theorems 12, 13, and Lemma 14.

1. Introduction

In order to satisfy increasing throughput and other performance demands, it is generally accepted that future communication networks will need to adopt new complementary approaches of network design, for example, densification (small cells), large scale antenna systems (massive multiple input multiple output (MIMO)), heterogeneous architectures, and cooperation/coordination (coordinated multi point). All of these new design paradigms ultimately result in overall very large communications systems with respect to the numbers of users, base stations and antennas in general. Most of the analytic tools used today in the communications community were originally developed for the analysis of point-to-point communication or small MIMO systems. Therefore, it is not a surprise that they often fail to provide meaningful insight into possible solutions for a new era of large dense heterogeneous multi user multi cell communication systems. New tools, which are adapted to the large nature of the systems, need to be developed and applied to give insight into and find the right combination of future system design solutions. Fortunately, the mathematical tool of large scale random matrix theory (RMT) has matured in recent years, to a point where it is of excellent use in this task. Publications documenting this maturing process (e.g., (Bai, Silverstein, 1998b ; Tulino, Verdú, 2004 ; Hachem, Khorunzhy *et al.*, 2008 ; Bai, Silverstein, 2009 ; Couillet, Debbah, 2011 ; Hoydis, 2012)) are numerous and they are not necessarily adapted to helping interested researchers to enter the field.

This article provides the theory needed to soundly use the framework of RMT, especially the technique known as *deterministic equivalent* approach. To this end, we first state the necessary basic theoretical concepts, lemmas and tools of the field. After this we will build intuition and insight into RMT concepts and their applications, by putting the introduced theoretic results into a tutorial like context. In order to familiarize the reader with the introduced tools, we will give a step by step derivation of the deterministic equivalent for the not-trivial example of determining the signal to interference plus noise ratio (SINR) given a certain linear receiver structure. Most of the concepts in this article have already been discussed in many other works (e.g., (Couillet, Debbah, 2011 ; Hoydis, 2012)). We will distinguish ourselves from these works by adhering to a more pedagogical style, as was done in our previous work (Müller, 2014). This means that we give more guidance than usual on how to arrive at a given result. Also details, which are only of mathematical interest are left out, when they are not essential. Hence, this tutorial might be most suited to future generations of researchers interested in the analytic RMT approach, than to experts of this field.

2. The Stieltjes Transform

The canonical introduction to the field of RMT is to begin with the definition of the Stieltjes transform. This is in part due to the history of the field, where Marčenko and Pastur first used this approach (Marčenko, Pastur, 1967) to find the distribution of the eigenvalues for certain random matrices. Others followed suit by using (e.g., (Telatar, 1999 ; Tse, Hanly, 1999 ; Verdú, Shamai, 1999)), extending (e.g., (Silverstein, Bai, 1995 ; Tulino, Verdú, 2004)) or building on (e.g., (Hachem *et al.*, 2007)) this approach in the context of communications systems [1]. Yet it also makes sense from an educational point of view, since Stieltjes transforms show up in many communications engineering problems and are relatively easy to handle, i.e., they serve as a good introduction to the framework of RMT. Let us start by defining some required terminology:

DÉFINITION 2. — *Given a measure μ that assigns finite measure to each bounded set on $\mathbb{R}$, we denote*

$$F_\mu(x) = \mu((-\infty, x]) \,.$$

If μ is a probability measure, then the associated F_μ is called the (cumulative) distribution function (cdf).

Now, we define the Stieltjes transform of a measure, by

DÉFINITION 3 (Stieltjes Transform). — *Let μ be a finite non negative measure with support* $\operatorname{supp}(\mu) \subset \mathbb{R}$*, i.e., $\mu(\mathbb{R}) < \infty$, and F_μ is given as in Definition 2. The Stieltjes transform $m(z)$ of μ is defined* $\forall\, z \in \mathbb{C} \setminus \operatorname{supp}(\mu)$ *as*

$$m(z) = \int_{\mathbb{R}} \frac{1}{\lambda - z} \mu(d\lambda) = \int_{\mathbb{R}} \frac{1}{\lambda - z} dF_\mu(\lambda) \,. \tag{1}$$

The second equality in (1) is not immediately evident and Billingsley (Billingsley, 1995) invites us to best regard $\int f(x)\mu(dx)$ and $\int f(x) dF_\mu(x)$ as merely notational variants [2]. Some literature uses $\int_{\mathbb{R}} \frac{1}{\lambda - z} d\mu(\lambda)$ as an alternative notation to (1) [3].

1. The work of Marčenko and Pastur on the spectra of random matrices in general is preceded by Wigner (Wigner, 1958). However, the first usage of the Stieltjes transform is generally attributed to Marčenko and Pastur.

2. The interested reader is invited to study (Billingsley, 1995, (17.22)ff.) for the subtle distinctions between th Riemann-Stieltjes Integral and the Lebesgue-Stieltjes Integral, which ultimately turn out to be unimportant in general measure theory.

3. This unfortunate practice seems to stem from a notational generalization of the known relation $\int dF(x) = \int F(dx)$.

The following important set of properties allows one to recover μ when only its Stieltjes transform $m(z)$ is known. These results can be found, for example, in (Hachem *et al.*, 2007) or (Krein, Nudelman, 1977).

PROPRIÉTÉ 4. — *Let $m(z)$ be the Stieltjes transform of a finite measure μ on $\mathbb{R}$. Then,*

(i) $\mu(\mathbb{R}) = \lim_{y\to\infty} -\mathbf{i}ym(\mathbf{i}y)$,

(ii) $\mu([a,b]) = \lim_{y\to 0^+} \frac{1}{\pi}\int_a^b \Im\{m(x+\mathbf{i}y)\}dx$, *if a, b are continuity points of μ.*

We proceed to define the empirical probability measure of the eigenvalues of an Hermitian matrix $\mathbf{X}$.

DÉFINITION 5 (Empirical Probability Measure of Eigenvalues). — *Let $\mathbf{X} \in \mathbb{C}^{N\times N}$ be a Hermitian matrix with the real valued eigenvalues $\lambda_1, \ldots, \lambda_N$. The empirical probability measure $\mu_{\mathbf{X}}$ of the eigenvalues of $\mathbf{X}$ is defined as*

$$\mu_{\mathbf{X}}(A) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i(\mathbf{X})\in A}\,.$$

The equivalent notation variants $\frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i(\mathbf{X})}(A)$ and $\frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_A(\lambda_i(\mathbf{X}))$ are also often found in the literature. This measure constitutes a point measure and can also be seen as a normalised counting measure. We define its corresponding distribution function (according to Definition 2) as

DÉFINITION 6 (Empirical Spectral Distribution (e.s.d.)). — *Let the empirical probability measure $\mu_{\mathbf{X}}(a)$ of the eigenvalues of $\mathbf{X}$ be defined as in Definition 5. The empirical (cumulative) distribution function, or empirical spectral distribution (e.s.d.) $F^{\mathbf{X}}(x)$ of the eigenvalues of $\mathbf{X}$ is then defined as*

$$F^{\mathbf{X}}(x) = \mu_{\mathbf{X}}((-\infty, x]) = \frac{1}{N}\sum_{i=1}^{N}\mathbb{1}_{\{\lambda_i(\mathbf{X})\le x\}}\,.$$

At this point some researchers might ask themselves, why one would be interested in the Stieltjes transform. It only seems to complicate and hide the information contained within the measure. Especially, taking the Stieltjes transform of an e.s.d. seems to only obscure the information about the eigenvalue distribution. However, this seemingly additional complication allows us to manipulate this information using existing tools, that would otherwise not be applicable. Or as the author of (Tao, 2012) put it: *As such, [the Stieltjes Transform] neatly packages the spectral information in a way that can be easily manipulated by the methods of complex analysis.*

To begin answering the common question about the practical connection between Stieltjes transforms and the spectra of Hermitian matrices, we introduce the notion of the resolvent $\mathbf{Q}$ of the Hermitian matrix $\mathbf{X}$:

$$\mathbf{Q}(z) = (\mathbf{X} - z\mathbf{I}_M)^{-1}\,.$$

Or, more generally

DÉFINITION 7 (Notation of Resolvents). — *The resolvent* $\mathbf{Q}_M$ *of a matrix* $\mathbf{A}_M \in \mathbb{C}^{M\times M}$ *is the complex-indexed matrix*

$$\mathbf{Q}_M(z) = (\mathbf{A}_M - z\mathbf{I})^{-1} .$$

It is defined for any $z \in \mathbb{C}$ *different from the eigenvalues of* $\mathbf{A}_M$.

The resolvent is a central object in spectral theory. Among other things, it indicates the eigenvalues of $\mathbf{X}$ by defining the support of the complex scalar variable z.
Taking our definition of the Stieltjes transform and using it with the empirical probability measure $\mu_{\mathbf{X}}$ from Definition 5, which we recall to be a point measure, one quickly finds:

$$m_{\mu_{\mathbf{X}}}(z) = \int_{\mathbb{R}} \frac{1}{\lambda - z} \mu_{\mathbf{X}}(d\lambda) = \frac{1}{M} \sum_{i=1}^{M} \frac{1}{\lambda_i(\mathbf{X}) - z} .$$

Abusing the diag notation in the sense of common computational software, it is possible to obtain

$$\begin{aligned} m_{\mu_{\mathbf{X}}}(z) &= \frac{1}{M} \mathrm{tr}\, \mathrm{diag} \left(\frac{1}{\lambda_1(\mathbf{X}) - z}, \cdots, \frac{1}{\lambda_M(\mathbf{X}) - z} \right) \\ &= \frac{1}{M} \mathrm{tr} \left\{ \left[\mathrm{diag}\left(\lambda_1(\mathbf{X}), \cdots, \lambda_M(\mathbf{X}) \right) - z\mathbf{I}_M \right]^{-1} \right\} \\ &\triangleq \frac{1}{M} \mathrm{tr} \left[(\mathbf{\Lambda} - z\mathbf{I}_M)^{-1} \right] \end{aligned}$$

for any unitary matrix $\mathbf{U} \in \mathbb{C}^{M\times M}$

$$m_{\mu_{\mathbf{X}}}(z) = \frac{1}{M} \mathrm{tr} \left[(\mathbf{\Lambda} \mathbf{U} \mathbf{U}^{\mathrm{H}} - z\mathbf{U}\mathbf{U}^{\mathrm{H}})^{-1} \right] = \frac{1}{M} \mathrm{tr} \left[(\mathbf{U} \mathbf{\Lambda} \mathbf{U}^{\mathrm{H}} - z\mathbf{I}_M)^{-1} \right]$$

if now $\mathbf{U}$ is chosen to contain the eigenvectors of the Hermitian matrix $\mathbf{X}$, we finally have

$$m_{\mu_{\mathbf{X}}}(z) = \frac{1}{M} \mathrm{tr} \Big[\underbrace{(\mathbf{X} - z\mathbf{I}_M)^{-1}}_{\text{Resolvent of } \mathbf{X}} \Big] . \qquad (2)$$

For the sake of brevity, we will abbreviate $m_{\mu_{\mathbf{X}}}(z)$ by $m_{\mathbf{X}}(z)$ in the following, whenever it does not impede understanding.

3. The Deterministic Equivalent

We will now discuss one of the most important concepts in RMT for the purpose of wireless communications, when first order results are concerned – the definition of a *deterministic equivalent* (DE). In order to define the DE, it is necessary to introduce the concept of almost sure convergence of sequences of random variables first:

DÉFINITION 8 (Almost Sure Convergence). — *The sequence of random variables $(X_n)_{n\geq 1}$ converges almost surely to X, if*

$$P\left(\limsup_{n\to\infty} |X_n - X| = 0\right) = 1\,.$$

This is denoted by $X_n \xrightarrow[n\to\infty]{a.s.} X$ or $X_n \xrightarrow{a.s.} X$, if the context is unambiguous.

We define the DE of a sequence of random quantities as follows:

DÉFINITION 9 (Deterministic Equivalent). — *The deterministic equivalent of a sequence of random complex values $(X_n)_{n\geq 1}$ is a deterministic sequence $(\overline{X}_n)_{n\geq 1}$, which approximates X_n such that*

$$X_n - \overline{X}_n \xrightarrow[n\to+\infty]{\text{a.s.}} 0\,.$$

DEs were first proposed in this form by Hachem *et al.* in (Hachem *et al.*, 2007 ; Hachem, Khorunzhy *et al.*, 2008). There is was also argued that these objects are able to provide accurate deterministic approximations of important system performance indicators in cellular networks. For example, the capacity of large dimensional multi antenna channels. Quite often the quantity X_n is going to be a functional of the resolvent of a Hermitian matrix. For example a normalized trace, which we know from (2) to be a Stieltjes transform of a probability measure. However, usually we are interested in more complex forms related to spectral properties. the object X_n will often concentrate around $\overline{X}_n$ in the large n regime and if $\overline{X}_n$ has a limit, we even obtain (almost sure) convergence. Furthermore, even relatively simple problems can result in a DE $\overline{X}_n$, which is not guaranteed to converge itself. Yet, it is possible to deterministically calculate $\overline{X}_n$.

In the practical application of DEs, the terms of "(almost sure) limit" and "large-scale approximation" are also often used. The following remarks should help differentiate those terms from DEs.

REMARQUE 10 ((Almost Sure) Limit). — If a sequence of random complex variables $(X_n)_{n\geq 1}$ almost surely converges to a simple (non-sequence) deterministic quantity $\overline{X}$, i.e.,

$$X_n \xrightarrow[n\to+\infty]{\text{a.s.}} \overline{X}$$

then we call this quantity $\overline{X}$ the (almost sure) limit of X_n. Sometimes this is also denoted $\lim X_n = \overline{X}$. □

REMARQUE 11 (Large-Scale Approximation). — If a DE is used as an approximation at finite n, it is often referred to as a *large-scale approximation*. □

We want to re-iterate here, that even though the concepts of Stieltjes transform and DE are often introduced alongside each other, they are a-priori completely independent. The Stieltjes transform is a (precise and non-asymptotic) tool to open up the spectral analysis of matrices to the tools of complex analysis, often via the empirical spectral distribution. The DE is an (almost surely asymptotically precise) deterministic approximation to a sequence of random quantities, which often represents some performance indicator of some problem defined by random quantities. However, it turns out that DEs of Stieltjes transforms are often relatively easy to find and many performance indicators can be expressed in terms of Stieltjes transforms.

The following theorems and lemmas, pertaining to DEs give us the theoretical justifications to treat and work with DEs as one would intuitively expect. First, the continuous mapping theorem is a very useful result if an arbitrary function f, e.g., a performance metric, is continuous:

THÉORÈME 12 (Continuous mapping theorem (Van der Vaart, 2000, Theorem 2.3)). — *Let $(X_n)_{n\geq 1}$ be a sequence of real random variables and let $f : \mathbb{R} \mapsto \mathbb{R}$ be continuous at every point of a set A such that $P(X \in A) = 1$, for some random variable X. Then, if $X_n \xrightarrow{a.s.} X$, this implies $f(X_n) \xrightarrow{a.s.} f(X)$.*

This theorems basically states, that a function of a DE behaves, as it would for the values it approximates. In some cases, one is able to prove that $X_n \xrightarrow{\text{a.s.}} X$, but one would like to show that $(X_n)_{n\geq 1}$ converges also in mean to X, i.e., $\lim_n \mathbb{E}\left[|X_n - X|\right] = 0$. This can often be done using the dominated convergence theorem:

THÉORÈME 13 (Dominated Convergence Theorem (Billingsley, 1995, Theorem 16.4)). — *Let $(f_n)_{n\geq 1}$ be a sequence of real measurable functions such that the pointwise limit $f(x) = \lim_{n\to\infty} f_n(x)$ exists. Assume there is an integrable $g : \mathbb{R} \mapsto [0, \infty]$ with $|f_n(x)| \leq g(x)$ for each $x \in \mathbb{R}$. Then f is integrable, as is f_n for each n, and*

$$\lim_{n\to\infty} \int_{\mathbb{R}} f_n d\mu = \int_{\mathbb{R}} f d\mu\,.$$

The standard argument to show that almost sure convergence of the DE often entails convergence in the mean is then as follows: Define the functions $f_n = |X_n - X|$ for all n. Since $X_n \xrightarrow{\text{a.s.}} X$, it follows that $f_n \xrightarrow{\text{a.s.}} f = 0$. If one can show that $f_n \leq g$ and $\mathbb{E}\left[g\right] < \infty$, it follows from the dominated convergence theorem that $\lim_{n\to\infty} \mathbb{E}\left[|X_n - X|\right] = 0$. For instance, Stieltjes transforms are bounded by $1/|z|$ for real supported measures, e.g., the empirical probability measure of eigenvalues in Definition 5. Hence, Stieltjes transforms of this measure are bounded functions, which allows us to infer convergence in the mean from the convergence of the Stieltjes transform. The final lemma is important when one deals with products or ratios of DEs.

LEMME 14. — *(Peacock* et al.*, 2008, Lemma 1) Let $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ be two sequences of complex random variables. Let $(\overline{a}_n)_{n\geq 1}$ and $(\overline{b}_n)_{n\geq 1}$ be two deterministic sequences of complex quantities. Assume that $a_n - \overline{a}_n \xrightarrow[n\to\infty]{a.s.} 0$ and $b_n - \overline{b}_n \xrightarrow[n\to\infty]{a.s.} 0$.*

(i) If $|a_n|$, $|\bar{b}_n|$ *and/or* $|\bar{a}_n|$,$|b_n|$ *are almost surely bounded*[4], *then*

$$a_n b_n - \bar{a}_n \bar{b}_n \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

(ii) If $|a_n|$, $|\bar{b}_n|^{-1}$ *and/or* $|\bar{a}_n|$,$|b_n|^{-1}$ *are almost surely bounded, then*

$$a_n/b_n - \bar{a}_n/\bar{b}_n \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

This Lemma allows us to take a "mix and match" or "divide and conquer" approach to calculating DEs involving products; much like in the case of simple sums. To be more precise, Theorems 12 and 13, combined with Lemma 14, allow us to directly find a DE of some continuous function of the SINR, while only DEs for the interference and signal power terms have been derived.

4. Common RMT Related Tools and Lemmas

Prior to demonstrating some calculations involving RMT, we need a few more standard tools and lemmas that will be of constant use throughout the derivations.

LEMME 15 (Common Matrix Identities). — *Let* $\mathbf{A}$, $\mathbf{B}$ *be complex invertible matrices and* $\mathbf{C}$ *a rectangular complex matrix, all of proper size. We restate the following, well known, relationships:*
Woodbury Identity:

$$\left(\mathbf{A}+\mathbf{C}\mathbf{B}\mathbf{C}^H\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}\left(\mathbf{B}^{-1}+\mathbf{C}^H\mathbf{A}^{-1}\mathbf{C}\right)^{-1}\mathbf{C}^H\mathbf{A}^{-1}. \tag{3}$$

Searl Identity:

$$\left(\mathbf{I}+\mathbf{A}\mathbf{B}\right)^{-1}\mathbf{A} = \mathbf{A}\left(\mathbf{I}+\mathbf{B}\mathbf{A}\right)^{-1}. \tag{4}$$

Resolvent Identity:

$$\mathbf{A}^{-1}-\mathbf{B}^{-1} = -\mathbf{A}^{-1}\left(\mathbf{A}-\mathbf{B}\right)\mathbf{B}^{-1} = \mathbf{A}^{-1}\left(\mathbf{B}-\mathbf{A}\right)\mathbf{B}^{-1}. \tag{5}$$

The first lemma completely pertaining to the concept of RMT is commonly referred to as the *trace lemma*. It concerns itself with the convergence of quadratic forms and was introduced in (Bai, Silverstein, 1998a). We will continue looking at sequences of matrices and random vectors with growing dimensions, i.e., $(\mathbf{A}_M)_{M\geq 1} \in \mathbb{C}^{M\times M}$ and $(\mathbf{x}_M)_{M\geq 1} \in \mathbb{C}^M$ or $(\mathbf{y}_M)_{M\geq 1} \in \mathbb{C}^M$. However, in order to improve readability we often abbreviate $(\mathbf{A}_M)_{M\geq 1}$ as $\mathbf{A}_M$ or even as $\mathbf{A}$, if the meaning is unambiguous.

LEMME 16 (Preliminary Trace Lemma Result (Bai, Silverstein, 2009, Lemma B.26)). — *Let* $\mathbf{A} \in \mathbb{C}^{M\times M}$ *be deterministic and* $\mathbf{x} = [x_1 \dots x_M]^{\mathsf{T}} \in \mathbb{C}^M$

4. I.e., all quantities x_n conform to $\limsup |x_n| < \infty$ with probability one.

be a random vector of independent entries. Assume $\mathbb{E}[x_i] = 0$, $\mathbb{E}\left[|x_i|^2\right] = 1$, *and* $\mathbb{E}\left[|x_i|^\ell\right] \leq \upsilon_\ell < \infty$ *for each* $\ell \leq 2p$. *Then, for any* $p \geq 1$,

$$\mathbb{E}\left[|\mathbf{x}^H \mathbf{A}\mathbf{x} - \operatorname{tr}\mathbf{A}|^p\right] \leq C_p \left(\operatorname{tr}\mathbf{A}\mathbf{A}^H\right)^{\frac{p}{2}} \left(\upsilon_4^{\frac{p}{2}} + \upsilon_{2p}\right)$$

where C_p *is a constant which only depends on* p.

LEMME 17 (Trace Lemma (Bai, Silverstein, 1998a)). — *Let* $\mathbf{x}_M = [x_1, \ldots, x_M]^T$ *be an* $M \times 1$ *vector where the* x_m *are i.i.d. Gaussian complex random variables with unit variance. Let* $\mathbf{A}_M$ *be an* $M \times M$ *matrix independent of* $\mathbf{x}_M$. *If in addition* $\limsup_M \|\mathbf{A}\|_2 < \infty$, *then we have the standard result*

$$\frac{1}{M}\mathbf{x}^H \mathbf{A}_M \mathbf{x} - \frac{1}{M}\operatorname{tr}(\mathbf{A}_M) \xrightarrow[M\to+\infty]{a.s.} 0\,. \tag{6}$$

PREUVE 18. — Immediately from Lemma 16 we see that for any $p \geq 2$, there exists a constant C_p, depending only on p, such that

$$\mathbb{E}_{\mathbf{x}_M}\left[\left|\frac{1}{M}\mathbf{x}_M^{\mathrm{H}}\mathbf{A}_M\mathbf{x}_M - \frac{1}{M}\operatorname{tr}(\mathbf{A}_M)\right|^p\right] \leq$$
$$\frac{C_p}{M^p}\left(\left(\mathbb{E}|x_m|^4 \operatorname{tr}\left(\mathbf{A}\mathbf{A}^{\mathrm{H}}\right)\right)^{p/2} + \mathbb{E}|x_m|^{2p} \operatorname{tr}\left(\mathbf{A}\mathbf{A}^{\mathrm{H}}\right)^{p/2}\right)$$

where the expectation is taken over the distribution of $\mathbf{x}_M$. If in addition $\limsup_M \|\mathbf{A}\|_2 < \infty$ and noticing that $\operatorname{tr}\left(\mathbf{A}\mathbf{A}^{\mathrm{H}}\right) \leq M\|\mathbf{A}\|_2^2$ and that $\operatorname{tr}\left(\mathbf{A}\mathbf{A}^{\mathrm{H}}\right)^{p/2} \leq M\|\mathbf{A}\|_2^p$, we obtain the simpler inequality:

$$\mathbb{E}_{\mathbf{x}_M}\left[\left|\frac{1}{M}\mathbf{x}_M^{\mathrm{H}}\mathbf{A}_M\mathbf{x}_M - \frac{1}{M}\operatorname{tr}(\mathbf{A}_M)\right|^p\right] \leq \frac{C_p'\|\mathbf{A}\|_2^p}{M^{p/2}}$$

where $C_p' = C_p\left(\left(\mathbb{E}[|x_m|^4]\right)^{p/2} + \mathbb{E}[|x_m|^{2p}]\right)$. By choosing $p = 4$ (the required moments do exist) and using the Markov inequality (Billingsley, 1995, Equation (5.31)), we have

$$P\left(\left|\frac{1}{M}\mathbf{x}_M^{\mathrm{H}}\mathbf{A}_M\mathbf{x}_M - \frac{1}{M}\operatorname{tr}(\mathbf{A}_M)\right| > \varepsilon\right) \leq \frac{1}{\varepsilon^4}\frac{C_4'\|\mathbf{A}\|_2^4}{M^2}.$$

Seeing that this quantity is summable, i.e.,

$$\sum_M P\left(\left|\frac{1}{M}\mathbf{x}_M^{\mathrm{H}}\mathbf{A}_M\mathbf{x}_M - \frac{1}{M}\operatorname{tr}(\mathbf{A}_M)\right| > \varepsilon\right) < +\infty$$

we can apply the first Borel-Cantelli lemma (Billingsley, 1995, Theorem 4.3) to get the following almost sure convergence result:

$$\frac{1}{M}\mathbf{x}^{\mathrm{H}}\mathbf{A}_M\mathbf{x} - \frac{1}{M}\operatorname{tr}(\mathbf{A}_M) \xrightarrow[M\to+\infty]{a.s.} 0.$$

■

Other versions of this result exist, which are adapted to specific variations of the basic problem and assumptions. For example, (Wagner *et al.*, 2012, Lemma 4) showed that $\limsup_M \|\mathbf{A}\|_2 < \infty$, only needs to hold almost surely and the assumption of the elements in $\mathbf{x}_M$ being i.i.d. can be replaced by mere independence (see Lemma 16 and (Hoydis, 2012)).

A natural complement to Lemma 17 is the following lemma,

LEMME 19 ((Couillet, Debbah, 2011, Lemma 3.7)). — *Let $\mathbf{A}_M$ be as in Lemma 17, i.e., $\limsup_M \|\mathbf{A}\|_2 < \infty$, and $\mathbf{x}_M, \mathbf{y}_M$ be random, mutually independent with complex Gaussian entries of zero mean and variance* 1*. Then, for any $p \geq 2$ we have*

$$\mathbb{E}\left[\left|\frac{1}{M}\mathbf{y}_M^H \mathbf{A}_M \mathbf{x}_M\right|^p\right] = O(M^{-p/2})\,.$$

In particular,

$$\frac{1}{M}\mathbf{y}_M^H \mathbf{A}_M \mathbf{x}_M \xrightarrow[M\to+\infty]{a.s.} 0\,. \tag{7}$$

This lemma indicates, that many random quantities that are similar to quadratic forms, asymptotically vanish.

We have seen that the previous Lemmas need statistical independence between the matrix and the vectors of the analysed object. This is often not the case, thus the following two matrix inversion lemmas can often be used to remove interfering matrix columns. This is especially effective in Gram matrices, i.e., matrices of the form $\mathbf{X}\mathbf{X}^{\mathrm{H}} = \sum_m \mathbf{x}_m \mathbf{x}_m^{\mathrm{H}}$, for $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_M] \in \mathbb{C}^{M\times M}$.

LEMME 20 (Matrix Inversion Lemma I (Silverstein, Bai, 1995, Lemma 2.2)). — *Let $\mathbf{A}$ be an $M \times M$ invertible matrix and $\mathbf{x} \in \mathbb{C}^M, c \in \mathbb{C}$ for which $\mathbf{A} + c\mathbf{x}\mathbf{x}^H$ is invertible. Then, as an application of* (3)*, we have*

$$\mathbf{x}^H \left(\mathbf{A} + c\mathbf{x}\mathbf{x}^H\right)^{-1} = \frac{\mathbf{x}^H \mathbf{A}^{-1}}{1 + c\mathbf{x}^H \mathbf{A}^{-1}\mathbf{x}} \tag{8}$$

and

$$\left(\mathbf{A} + c\mathbf{x}\mathbf{x}^H\right)^{-1}\mathbf{x} = \frac{\mathbf{A}^{-1}\mathbf{x}}{1 + c\mathbf{x}^H \mathbf{A}^{-1}\mathbf{x}}\,. \tag{9}$$

LEMME 21 (Matrix Inversion Lemma II). — *Using the same definitions as in Lemma 20 and combining this lemma with* (5)*, one finds the relationship*

$$\left(\mathbf{A} + c\mathbf{x}\mathbf{x}^H\right)^{-1} = \mathbf{A}^{-1} - \frac{c\mathbf{A}^{-1}\mathbf{x}\mathbf{x}^H\mathbf{A}^{-1}}{1 + c\mathbf{x}^H \mathbf{A}^{-1}\mathbf{x}}\,. \tag{10}$$

The following rank-one perturbation lemma is particularly useful, if one has used a matrix inversion lemma to remove a statistical dependence before using the trace lemma. It shows that the DE is also valid for the complete matrix (see for example (15)).

LEMME 22 (Rank-One Perturbation Lemma (Bai, Silverstein, 2007, Lemma 2.1)). — *Let $z \in \mathbb{C} \setminus \mathbb{R}^+$, $\mathbf{A} \in \mathbb{C}^{M\times M}$, $\mathbf{B} \in \mathbb{C}^{M\times M}$ with $\mathbf{B}$ Hermitian non negative definite and $\mathbf{x} \in \mathbb{C}^M$. Then*

$$\left|\operatorname{tr}\left[\mathbf{A}\left((\mathbf{B}-z\mathbf{I}_M)^{-1}-(\mathbf{B}+\mathbf{x}\mathbf{x}^{H}-z\mathbf{I}_M)^{-1}\right)\right]\right| \leq \frac{\|\mathbf{A}\|_2}{dist(z,\mathbb{R}^+)}$$

where $dist()$ is the Euclidean distance. If $z \in \mathbb{R}^-$ and $\limsup_M \|\mathbf{A}\|_2 < \infty$, then this implies

$$\frac{1}{M}\left|\operatorname{tr}\left[\mathbf{A}\left((\mathbf{B}-z\mathbf{I}_M)^{-1}-(\mathbf{B}+\mathbf{x}\mathbf{x}^{H}-z\mathbf{I}_M)^{-1}\right)\right]\right| \leq \frac{1}{M}\frac{\|\mathbf{A}\|_2}{|z|} \xrightarrow[M\to\infty]{} 0\,.$$

We remark that the variable z often corresponds to the inverse of the signal to noise ratio (SNR) in communications problems that use regularized linear precoding schemes. This partly explains the sometimes observed deteriorating approximation performance in such applications at large SNR. In (Couillet, Debbah, 2011, Lemma 14.3) one can also find a variant of Lemma 22 for $z = 0$, under the assumption the smallest eigenvalue of the Hermitian matrix B bounded away from zero for all large M, i.e, $\liminf_{M\to\infty} \lambda_{min}(B) > 0$:

$$\frac{1}{M}\operatorname{tr}\mathbf{A}\mathbf{B}^{-1} - \frac{1}{M}\operatorname{tr}\mathbf{A}\left(\mathbf{B}+\mathbf{v}\mathbf{v}^{\mathrm{H}}\right)^{-1} \xrightarrow[M\to\infty]{a.s.} 0\,.$$

5. Application of RMT in Communications

In this section, we will motivate the usage of RMT and showcase the discussed tools and approaches for the analysis of advanced communication systems.

5.1. Advantages of Large Dimensional Analyses

A question many researchers ask before becoming interested in the field of RMT, is why it is necessary in the first place to go to abstract large dimensional (tending to infinite) analysis. Wireless communication systems are becoming more and more complicated, so we need to use tools that simplify the analysis. The standard approach today is to use Monte-Carlo (MC) simulations. However, the introduced DEs have several advantages over the MC approach. For one, as DEs do not contain any randomness, it is possible to simplify analysis and facilitate understanding of the underlying relationships within the respective research problems. Take, for example, a system whose performance is influence by several parameters in non-linear ways. The deterministic solution via DEs shows the direct causal relationships and interactions between the system parameters and performance; something that is impossible to achieve with MC analysis. Furthermore, the analytic formulations of DEs enable direct optimization using known mathematical tools. Also one might ask, why not go to finite dimensional theoretical analysis? The short answer is that such analyses are either too complicated to be useful or they are (usually) unsolvable. This becomes evident in the following simple example [5]:

We define a simple multi-user (MU) multiple input multiple output (MIMO) uplink system, in which the base station is comprised of a central processing station and M distributed antennas (or remote radio heads). We take K single antenna users that transmit at the same time and at the same frequency, using Gaussian signalling for the transmit symbols $x_i \sim \mathcal{CN}(0, P)$ that form the aggregate transmit symbol vector $\mathbf{x} = [x_1, \ldots x_K]$. We assume that $P = \mathcal{O}(1/K)$, such that the transmit power remains bounded for an increasing number of UTs. For the channel model, we employ Rayleigh fading $h_{i,j} \sim \mathcal{CN}(0, v_{i,j})$, $1 \leq i \leq M$, $1 \leq j \leq K$. In other words, the resulting aggregate channel matrix $\mathbf{H}$ has a variance profile $\mathbf{V} = \{v_{i,j}\}$, $1 \leq i \leq M$, $1 \leq j \leq K$. Taking additive white Gaussian receiver noise into account and without receive processing, we obtain the standard formula for the received signal:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}\,.$$

The usual first question concerning the analysis of this very simple system is to find its capacity. We know from Telatar (Telatar, 1999, Theorem 2) that in the case of a Gaussian normal i.i.d. channel (i.e., $v_{i,j} = 1\ \forall i, j$), the ergodic mutual information per receive antenna is given as [6]

$$C_M^{iid} = \mathrm{E}_{\mathbf{H}}\left[\frac{1}{M}\log\det\left(\mathbf{I}_M + P\mathbf{H}\mathbf{H}^{\mathrm{H}}\right)\right] = \int_0^\infty \log\left(1 + P\lambda\right) f(\lambda) d\lambda$$

where $f(\lambda)$ is the probability density function of an unordered eigenvalue λ of the Wishart matrix $\mathbf{H}\mathbf{H}^{\mathrm{H}}$ and it is given by

$$f(\lambda) = \frac{M-K}{M}\delta(\lambda) + \frac{K}{M}\frac{1}{K}\sum_{k=0}^{K-1}\frac{k!}{(k+M-K)!}\left[L_k^{M-K}(\lambda)\right]^2 \lambda^{M-K}\mathrm{e}^{-\lambda}\,.$$

Here, $L_k^{M-K}(\lambda)$ is the associated Laguerre polynomial of order k:

$$L_k^N(\lambda) = \frac{\lambda^{-N}\mathrm{e}^{\lambda}}{k!}\frac{d^k}{d\lambda^k}\left(\mathrm{e}^{-\lambda}\lambda^{k+N}\right) = \sum_{l=0}^{k}(-1)^l\frac{(k+N)!}{(k-l)!\,(N+l)!l!}\lambda^l\,.$$

(Dohler, 2003, Eq. (2.38) and (2.45)) described a way to calculate the integral in the capacity equation in closed form, e.g. for the case of $N = K = 2$ we have

$$f(\lambda) = \frac{1}{2}\sum_{k=0}^{1}\left[L_k^0(\lambda)\right]^2\mathrm{e}^{-\lambda} = \frac{1}{2}\left[1 + \left(1 - \lambda^2\right)\right]\mathrm{e}^{-\lambda}\,.$$

---

5. This example follows closely (Hoydis, 2010).

6. In the case of Gaussian channels with Rayleigh fading, Gaussian signalling with mean zero and covariance $\frac{PK}{K}\mathbf{I}_K$ maximises the mutual information which, thus, is equivalent to the capacity (Telatar, 1999, Theorem 1).

Realizing that $L_0^0 = 1$ and $L_1^0 = 1 - \lambda$ we arrive at

$$C_2^{iid} = \frac{1}{2}\int_0^{\infty} \log(1 + P\lambda)\left[1 + \left(1 - \lambda^2\right)\right] \mathrm{e}^{-\lambda} d\lambda \tag{11}$$

$$= \frac{1}{2} - \frac{1}{2P} + \left(1 + \frac{1}{2P^2}\mathrm{e}^{1/P}\mathrm{E}_1(1/P)\right) \tag{12}$$

where $\mathrm{E}_1(z) = \int_1^{\infty} \frac{e^{-tz}}{t}\, dt$ is the exponential integral for complex values and can be computed using numerical software. In summary it is possible to derive a closed form solution for the ergodic mutual information for simple systems featuring channels without variance profiles. However, the resulting formulations do not offer much insight any more. For example, one clearly struggles to predict the influence exerted by P in (12). Furthermore, the finite dimensional approach breaks down completely, once one tries to deviate from any of the ideal assumptions. For example, moving away from the assumption of Gaussian distributions makes problem impossible to solve. Even if we now start to consider a simple variance profile like

$$\mathbf{V} = \begin{pmatrix} 1 & \alpha \\ \alpha & 1 \end{pmatrix} \tag{13}$$

finding the corresponding ergodic mutual information becomes intractable. In other words, even for simple systems, the finite dimensional theory approach often results in unsolvable problems. Also, even if the finite dimensional problem is solvable the formulations usually become too complicated to offer much insight and/or they require numerical tools for solving.

Using the large dimensional approach on the other hand, we can relatively easily treat, e.g., the case of arbitrary variance profiles. From Hachem et al. (Hachem *et al.*, 2007) we have the following theorem

THÉORÈME 23 (Capacity under Variance Profile (Hachem *et al.*, 2007, Theorem 4.1) (also (Hoydis *et al.*, 2011) and (Hachem, Khorunzhy *et al.*, 2008, Theorem 1))). — *Let $M, K \to \infty$ such that $0 < \frac{K}{M} < \infty$ and $v_{i,j} < v_{max} < \infty$, $\forall i, j$. Then for the model used in Subsection 5.1, we have $C_M - \overline{C}_M \xrightarrow{\text{a.s.}} 0$, where*

$$\overline{C}_M = \frac{1}{M}\sum_{i=1}^{K} \log(1 + \delta_j) - \frac{1}{M}\sum_{i=1}^{M} \log\left(\frac{1}{PK}e_i\right) - \frac{1}{M}\sum_{j=1}^{K} \frac{\delta_j}{1 + \delta_j}$$

with $\delta_j = \frac{1}{K}\sum_{l=1}^{M} v_{l,j} e_l$ for $j = 1, \ldots, K$ and e_i for $i = 1, \ldots, M$ is given as the unique positive solution to the M implicit equations

$$e_i = \left(\frac{1}{PK} + \frac{1}{K}\sum_{j=1}^{K} \frac{v_{i,j}}{1 + \frac{1}{K}\sum_{l=1}^{M} v_{l,j} e_l}\right)^{-1}.$$

Though it might look daunting at first, Theorem 23 offers many analytical benefits. For example, it lends itself readily optimization. In any case, DEs like this are the only known deterministic formulations of the channel capacity, given a variance profile.

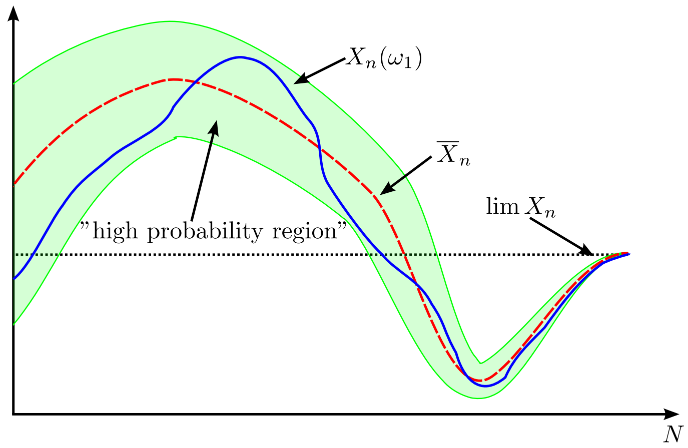

FIGURE *1. Qualitative comparison of the DE with classical limit calculus and single realization.*

5.2. Accuracy Considerations

Most publications using large dimensional techniques, take a rather pragmatic approach to the questions of accuracy and reliability of large dimensional results in systems of practical sizes. They usually simply provide a few simulations, which compare the obtained closed form results with few corresponding data points obtained by exhaustive Monte-Carlo analyses for finite dimensions. The regions between the verified points are then assumed to follow the observed trend. This approach usually produces acceptable outcomes as the almost sure convergence of DEs offer large advantages with respect to more classical limit analysis. In Figure 1 we show the implications of both approaches. It illustrates a typical realization of a sequence of random variables $X_n(\omega_1)$, which represents some system performance indicator (for example, random with respect to the channel realisations) that also depends on the generic system size N. Taking the classic limit w.r.t. the system size one could only obtain $\lim_{\to\infty} X_n$, which gives an arbitrarily accurate provable result for an infinitely large system. However, the usefulness of such a result is constraint to only the infinitely large system. The deterministic equivalent approach on the other hand gives us more information. Intuitively, one can remark that $\overline{X}_n$ is still "contains" the factor n, even as $n \to \infty$. In fact, the DE gives us an approximation for each value of N, which becomes more precise for increasing N. The realizations of the random variable, almost surely $(a.s.)$ fall within a increasingly narrow bound around the DE; see the "high probability region" in the figure. Furthermore, the DE approach also allows for approximations of random sequences that do not even converge at all (unlike the one chosen for illustrative purposes in Figure 1), which is completely impossible using classic limits. Thus, DEs tend to be much more accurate for finite (and even small) system dimensions than the classical limits. In general, one observes good agreement of DE and MC results for n in the tens, for first order statistics.

Assume we analyse a random quantity involving a random matrix $\mathbf{X} \in \mathbb{C}^{M\times K}$. RMT handles cases in which both dimensions (M, K) grow large, while classic limit approximations (e.g., the strong law of large numbers) can only treat the case where M grows large. As a consequence, RMT results exploit more degrees of freedom than classical approaches and, thus, usually far outperform them w.r.t. convergence speed. For example, even an 8×8 matrix offers already up to 64 degrees of freedom, which mostly leads to quite acceptable convergence. In general, RMT achieves impressive convergence rates for linear functionals of eigenvalues, e.g., for central limit theorems in $1/M$ (i.e., $M(X_M - \overline{X}_M) \to \mathcal{N}(0,1)$) (Hachem, Loubaton, Najim, 2008, Remark 3.1)(Pastur, Shcherbina, 2011) and for expectations in $1/M^2$ (i.e., $\mathbb{E}[X_M] = \overline{X}_M + \mathcal{O}(1/M^2)$) (Hachem, Khorunzhy *et al.*, 2008, Proof of Theorem 1)(Pastur, Shcherbina, 2011)(Vallet *et al.*, 2010, Proposition 4), when the random quantities are complex Gaussian distributed. Quadratic forms are usually slower, for example central limit theorems in $1/\sqrt{M}$ (Kammoun *et al.*, 2009, Theorem 3)(Rubio *et al.*, 2012, Theorems 1&2) and expectations in $1/M^{\frac{3}{2}}$ (Rubio *et al.*, 2012, Lemma 1).

5.3. DEs and Communications Problems

We have already seen the connection between Stieltjes transforms and traces of resolvents in (2). Now we want to take a quick, but detailed, look at how the trace of a resolvent is often found in communications problems, especially in questions pertaining to SINRs, and how it relates to DE theorems. The following example is based on (Couillet, Debbah, 2009) and (Couillet, Debbah, 2011), where a CDMA system using MMSE detection was treated.

Assume an uplink MU-MIMO system with K single antenna users simultaneously transmitting to a single base station with M antennas, which utilizes linear MMSE detection. The small scale fading part of the channel from user k to the BS is modelled as $\mathbf{h}_k \sim \mathcal{CN}(0, \frac{1}{M}\mathbf{I}_M) \in \mathbb{C}^M$, thus $\mathbb{E}\|\mathbf{h}_k\|_2 = 1$. We include the large scale fading effects, e.g., shadowing and pathloss effects, for user k as a variable d_k. Using Gaussian signalling for the transmitted symbols $s_k \sim \mathcal{CN}(0,1)$ and taking the receiver noise $\mathbf{n}$ to be additive Gaussian with zero mean and variance σ^2 leads to the following transmission model at any one given symbol time instance

$$\mathbf{y} = \sum_{k=1}^{K} d_k \mathbf{h}_k s_k + \mathbf{n} = \mathbf{HDs} + \mathbf{n}$$

where $\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_K] \in \mathbb{C}^{M\times K}$, $\mathbf{s} = [s_1, \ldots, s_K] \in \mathbb{C}^K$ and $\mathbf{D} = \text{diag}(d_1, \ldots, d_K) \in \mathbb{C}^{K\times K}$. The linear MMSE detector for each user is given as

$$\mathbf{r}_k^{\text{H}} = \mathbf{h}_k^{\text{H}} \left(\mathbf{HD}^2\mathbf{H}^{\text{H}} + \sigma^2 \mathbf{I}_M\right)^{-1} .$$

Hence, the signal to interference and noise ratio (SINR) pertaining to user k is defined as

$$\begin{aligned}\mathrm{SINR}_k &= \frac{\mathbb{E}_{\mathbf{s}}|\mathbf{r}_k^{\mathrm{H}} d_k \mathbf{h}_k s_k|^2}{\mathbb{E}_{\mathbf{s},\mathbf{n}}|\sum_{j\neq k} \mathbf{r}_k^{\mathrm{H}} d_j \mathbf{h}_j s_j + \mathbf{r}_k^{\mathrm{H}}\mathbf{n}|^2} \\
&= \frac{|d_k|^2 \mathbf{r}_k^{\mathrm{H}} \mathbf{h}_k \mathbf{h}_k^{\mathrm{H}} \mathbf{r}_k}{\sum_{j\neq k} |d_j|^2 \mathbf{r}_k^{\mathrm{H}} \mathbf{h}_j \mathbf{h}_j^{\mathrm{H}} \mathbf{r}_k + \mathbf{r}_k^{\mathrm{H}} \sigma^2 \mathbf{r}_k} \\
&= \frac{|d_k|^2 \mathbf{r}_k^{\mathrm{H}} \mathbf{h}_k \mathbf{h}_k^{\mathrm{H}} \mathbf{r}_k}{\mathbf{r}_k^{\mathrm{H}} \left(\mathbf{H}\mathbf{D}^2\mathbf{H}^{\mathrm{H}} - |d_k|^2 \mathbf{h}_k \mathbf{h}_k^{\mathrm{H}}\right) \mathbf{r}_k + \mathbf{r}_k^{\mathrm{H}} \sigma^2 \mathbf{r}_k} \\
&= \frac{|d_k|^2 \mathbf{r}_k^{\mathrm{H}} \mathbf{h}_k \mathbf{h}_k^{\mathrm{H}} \mathbf{r}_k}{\mathbf{r}_k^{\mathrm{H}} \left(\mathbf{H}\mathbf{D}^2\mathbf{H}^{\mathrm{H}} + \sigma^2 \mathbf{I}_M - \sigma^2 \mathbf{I}_M - |d_k|^2 \mathbf{h}_k \mathbf{h}_k^{\mathrm{H}}\right) \mathbf{r}_k + \mathbf{r}_k^{\mathrm{H}} \sigma^2 \mathbf{r}_k} \\
&= \frac{|d_k|^2 \mathbf{r}_k^{\mathrm{H}} \mathbf{h}_k \mathbf{h}_k^{\mathrm{H}} \mathbf{r}_k}{\mathbf{r}_k^{\mathrm{H}} \left(\mathbf{H}\mathbf{D}^2\mathbf{H}^{\mathrm{H}} + \sigma^2 \mathbf{I}_M\right) \mathbf{r}_K - \mathbf{r}_k^{\mathrm{H}} \sigma^2 \mathbf{r}_k - |d_k|^2 \mathbf{r}_k^{\mathrm{H}} \mathbf{h}_k \mathbf{h}_k^{\mathrm{H}} \mathbf{r}_k + \mathbf{r}_k^{\mathrm{H}} \sigma^2 \mathbf{r}_k} \,.\end{aligned}$$

Taking into account the cancelling terms, also those present within the definition of $\mathbf{r}_k^{\mathrm{H}}$, we have:

$$\mathrm{SINR}_k = \frac{|d_k|^2 \mathbf{r}_k^{\mathrm{H}} \mathbf{h}_k \mathbf{h}_k^{\mathrm{H}} \mathbf{r}_k}{\mathbf{x}^{\mathrm{H}} \mathbf{r}_K - |d_k|^2 \mathbf{r}_k^{\mathrm{H}} \mathbf{h}_k \mathbf{h}_k^{\mathrm{H}} \mathbf{r}_k} = \frac{|d_k|^2 \mathbf{r}_k^{\mathrm{H}} \mathbf{h}_k}{1 - |d_k|^2 \mathbf{r}_k^{\mathrm{H}} \mathbf{h}_k}$$

finally, re-introducing the definition of $\mathbf{r}_k^{\mathrm{H}}$ everywhere and applying Lemma 20, we have

$$\mathrm{SINR}_k = |d_k|^2 \mathbf{h}^{\mathrm{H}} \left(\mathbf{H}\mathbf{D}^2\mathbf{H}^{\mathrm{H}} - |d_k|^2 \mathbf{h}_k \mathbf{h}_k^{\mathrm{H}} + \sigma^2 \mathbf{I}_M\right)^{-1} \mathbf{h}_k \,. \tag{14}$$

Until now, this derivation did not use any concepts from RMT. This changes now, as one can simplify the SINR equation in (14) even further in the large system regime $M \to \infty$, where $0 < M/K = c < \infty$. We begin by calling first upon Lemma 17 and then Lemma 22 to arrive at

$$\mathrm{SINR}_k - \frac{1}{M} |d_k|^2 \mathrm{tr}\, \left(\mathbf{H}\mathbf{D}^2\mathbf{H}^{\mathrm{H}} + \sigma^2 \mathbf{I}_M\right)^{-1} \xrightarrow[M\to+\infty]{a.s.} 0 \,. \tag{15}$$

We then remember that the definition of the Stieltjes transform (Definition 3) together with the normalised counting measure of the eigenvalues of the matrix $\mathbf{H}\mathbf{D}^2\mathbf{H}^{\mathrm{H}}$, i.e., $m_{\mathbf{H}\mathbf{D}^2\mathbf{H}^{\mathrm{H}}}(z) = \mathrm{tr}\, \left(\mathbf{H}\mathbf{D}^2\mathbf{H}^{\mathrm{H}} - z\mathbf{I}_M\right)^{-1}$ (see Definition 5). This allows us to rewrite (15) as

$$\mathrm{SINR}_k - |h_k|^2 m_{\mathbf{X}\mathbf{D}^2\mathbf{X}^{\mathrm{H}}}(-\sigma^2) \xrightarrow[M\to+\infty]{a.s.} 0 \,.$$

We remember, that the Stieltjes transform $m_{\mathbf{X}\mathbf{D}^2\mathbf{X}^{\mathrm{H}}}(-\sigma^2)$ still represents a random quantity. It is possible to find a DE of this expression, which we will demonstrate in Subsection 5.4, and leads to Theorem 24. Direct application leads to

$$m_{\mathbf{H}\mathbf{D}^2\mathbf{H}^{\mathrm{H}}}(-\sigma^2) - \overline{m}_{\mathbf{H}\mathbf{D}^2\mathbf{H}^{\mathrm{H}}}(-\sigma^2) \xrightarrow[M\to+\infty]{a.s.} 0 \,.$$

with

$$\overline{m}_{\mathbf{HD}^2\mathbf{H}^{\mathrm{H}}}(-\sigma^2) = \left(\sigma^2 + c\sum_{i=1}^{K}\frac{|d_i|^2}{1+|d_i|^2\overline{m}_{\mathbf{HD}^2\mathbf{H}^{\mathrm{H}}}(-\sigma^2)}\right)^{-1}$$

This formulation is deterministic w.r.t. the entries of $\mathbf{H}$, but conditionally on the entries of $\mathbf{D}$, i.e., the large scale fading coefficients d_k.

5.4. Derivation of a DE

We continue this chapter by applying the introduced tools and concepts in an example derivation of a DE. Here we tried to include every step of the derivation; even those that might seem obvious to many. This example also serves to illustrate one approach to finding a DE in the first place [7]. However, the main focus here is to give an interesting application case for the previously introduced tools and lemmas. This example only tries to give an intuitive understanding of how one could believably take on the derivation of a new DE. The following is a simplification of the work in (Couillet *et al.*, 2011). Aspects of the work that were deemed too technical or not helpful for understanding have been left out. For all technical details, we invite the reader to refer to the original paper (Couillet *et al.*, 2011), or a less reduced version in (Couillet, 2010).

THÉORÈME 24. — *Let* $\mathbf{T} \in \mathbb{C}^{M\times K}$ *be a non negative definite diagonal matrix and* $\mathbf{R} \in \mathbb{C}^{M\times M}$ *be a non negative definite matrix, both having bounded spectral norm, i.e.,* $\limsup_{K\to\infty}\|\mathbf{T}\| = \limsup_{K\to\infty}\lambda_{\max}(\mathbf{T}) < \infty$ *and* $\limsup_{K\to\infty}\|\mathbf{R}\| < \infty$. *Let* $\mathbf{X} \in \mathbb{C}^{M\times K}$ *be a matrix, whose elements are distributed as* $\mathcal{CN}(0, \frac{1}{K})$. *Define also* $\mathbf{B} = \mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{H}\mathbf{R}^{\frac{1}{2}}$.

Then, as $M, K \to \infty$, *such that* $M/K \to c$, *where* c *is some bounded constant, i.e.,* $0 < c < \infty$. *The following result holds*

$$\frac{1}{M}\mathrm{tr}\left[(\mathbf{B} - z\mathbf{I}_M)^{-1}\right] - \overline{m}_M(z) \xrightarrow[M,K\to+\infty]{a.s.} 0$$

where $z \in \mathbb{C}\setminus\mathbb{R}^+$ *and to* $\overline{m}_M$ *is given by*

$$\overline{m}_M = \frac{1}{M}\mathrm{tr}\left(\mathbf{R}\,e(z) - z\mathbf{I}_M\right)^{-1}$$

which includes finding the unique positive solution to the fixed-point equation

$$e(z) = \frac{1}{K}\sum_{i=1}^{K}\frac{t_i}{1 + t_i c\frac{1}{M}\mathrm{tr}\,\mathbf{R}\left(\mathbf{R}\,e(z) - z\mathbf{I}_M\right)^{-1}}\,.$$

7. The method shown in following is often referred to as the “Bai-Silverstein approach”, after the steps outlined for example in (Silverstein, Bai, 1995). There are many other proof techniques, e.g. the “Pastur approach”, which relies on “Gaussian methods” (Pastur, 1999 ; Dupuy, Loubaton, 2010) and is generally considered more powerful, but also less evident.

It is not immediately obvious how one could arrive at such a theorem. Following the Bai-Silverstein approach we start by making an educated guess of the general form of the result (see Remark 25 later on, to motivate this choice):

$$\frac{1}{M}\mathrm{tr}\,(\mathbf{B}-z\mathbf{I}_M)^{-1} - \qquad ??? \qquad \xrightarrow[M,K\to+\infty]{a.s.} 0$$

$$\xrightarrow[\text{guess}]{\text{educated}} \quad \frac{1}{M}\mathrm{tr}\,(\mathbf{B}-z\mathbf{I}_M)^{-1} - \frac{1}{M}\mathrm{tr}\,(\mathbf{R}\,e(z)-z\mathbf{I}_M)^{-1} \xrightarrow[M,K\to+\infty]{a.s.} 0\,.$$

The main goal is now to find a formulation for $e(z)$, that does not depend on the random quantities and adheres to the almost sure convergence. Using the resolvent identity (5), one quickly finds

$$\begin{aligned}
&\frac{1}{M}\mathrm{tr}\,(\mathbf{B}-z\mathbf{I}_M)^{-1} - \frac{1}{M}\mathrm{tr}\,(\mathbf{R}\,e(z)-z\mathbf{I}_M)^{-1}\\
&\overset{(5)}{=} \frac{1}{M}\mathrm{tr}\,\left[(\mathbf{B}-z\mathbf{I}_M)^{-1}\,(e(z)\mathbf{R}-\mathbf{B}-z\mathbf{I}_M+z\mathbf{I}_M)\,(\mathbf{R}\,e(z)-z\mathbf{I}_M)^{-1}\right]\\
&= \frac{1}{M}\mathrm{tr}\,\left[(\mathbf{B}-z\mathbf{I}_M)^{-1}\left(e(z)\mathbf{R}-\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}}\right)(\mathbf{R}\,e(z)-z\mathbf{I}_M)^{-1}\right]\\
&= \frac{1}{M}\mathrm{tr}\,\left[(\mathbf{B}-z\mathbf{I}_M)^{-1}\,\mathbf{R}^{\frac{1}{2}}\,(e(z)-\mathbf{X}\mathbf{T}\mathbf{X}^{\mathrm{H}})\,\mathbf{R}^{\frac{1}{2}}\,(\mathbf{R}\,e(z)-z\mathbf{I}_M)^{-1}\right]\\
&= \frac{1}{M}\mathrm{tr}\,\left[(\mathbf{B}-z\mathbf{I}_M)^{-1}\,\mathbf{R}\,(\mathbf{R}\,e(z)-z\mathbf{I}_M)^{-1}\right]e(z)\\
&\qquad - \frac{1}{M}\mathrm{tr}\left[\,(\mathbf{B}-z\mathbf{I}_M)^{-1}\,\mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}}\,(\mathbf{R}\,e(z)-z\mathbf{I}_M)^{-1}\,\right].
\end{aligned}$$

Remembering that for $\mathbf{X}=[\mathbf{x}_1,\dots\mathbf{x}_K]$ and $\mathbf{T}=\mathrm{diag}\,(t_1,\dots,t_K)$, we have $\mathbf{X}\mathbf{T}\mathbf{X}^{\mathrm{H}}=\sum_{i=1}^{K}t_i\mathbf{x}_i\mathbf{x}_i^{\mathrm{H}}$. Hence we can pull this sum outside.

$$\begin{aligned}
&= \frac{1}{M}\mathrm{tr}\,\left[(\mathbf{B}-z\mathbf{I}_M)^{-1}\,\mathbf{R}\,(\mathbf{R}\,e(z)-z\mathbf{I}_M)^{-1}\right]e(z)\\
&\qquad - \frac{1}{M}\sum_{i=1}^{K}t_i\mathrm{tr}\,\left[\underbrace{(\mathbf{B}-z\mathbf{I}_M)^{-1}\,\mathbf{R}^{\frac{1}{2}}\mathbf{x}_i}_{\tilde{\mathbf{x}}_i}\;\underbrace{\mathbf{x}_i^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}}\,(\mathbf{R}\,e(z)-z\mathbf{I}_M)^{-1}}_{\tilde{\mathbf{x}}_i^{\mathrm{H}}}\right].
\end{aligned}$$

Since the argument of the trace operators is a scalar, it is possible to remove the operator, obtaining:

$$\begin{aligned}
&= \frac{1}{M}\mathrm{tr}\,\left[(\mathbf{B}-z\mathbf{I}_M)^{-1}\,\mathbf{R}\,(\mathbf{R}\,e(z)-z\mathbf{I}_M)^{-1}\right]e(z)\\
&\qquad - \frac{1}{M}\sum_{i=1}^{K}t_i\mathbf{x}_i^{\mathrm{H}}\underbrace{\mathbf{R}^{\frac{1}{2}}\,(\mathbf{R}\,e(z)-z\mathbf{I}_M)^{-1}\,(\mathbf{B}-z\mathbf{I}_M)^{-1}\,\mathbf{R}^{\frac{1}{2}}}_{\triangleq\tilde{\mathbf{A}}}\mathbf{x}_i\,.
\end{aligned}$$

One might be tempted to apply the trace lemma (Lemma 17) to the form $\mathbf{x}_i^{\mathrm{H}}\tilde{\mathbf{A}}\mathbf{x}_i$ directly at this point, but it is a good idea to verify the prerequisites. In particular, we need to be sure that $\mathbf{x}_i$ is statistically independent of $\tilde{\mathbf{A}}$, which is only possible if $\mathbf{x}_i$ is statistically independent of $\mathbf{B}$. This is obviously not the case, as (in greatest possible detail):

$$\mathbf{B} = \mathbf{R}^{\frac{1}{2}}\mathbf{X}\mathbf{T}\mathbf{X}^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}} = \sum_{j=1}^{K} t_j \mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}} = \sum_{j\neq i}^{K} t_j \mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}} + t_i\mathbf{R}^{\frac{1}{2}}\mathbf{x}_i\mathbf{x}_i^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}} .$$

Hence, we need to apply Lemma 20 first, in order to "remove" the dependent part. So, analogously to what has been done above, it is possible to split the equation as:

$$(\mathbf{B} - z\mathbf{I}_M)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_i = \Bigg(\underbrace{\sum_{j\neq i}^{K} t_j \mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}} - z\mathbf{I}_M}_{\tilde{\mathbf{A}}} + \underbrace{t_i}_{\tilde{c}}\underbrace{\mathbf{R}^{\frac{1}{2}}\mathbf{x}_i}_{\tilde{x}_i}\underbrace{\mathbf{x}_i^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}}}_{\tilde{x}_i^{\mathrm{H}}}\Bigg)^{-1}\underbrace{\mathbf{R}^{\frac{1}{2}}\mathbf{x}_i}_{\tilde{x}_i}$$

and apply the matrix inversion lemma to arrive at

$$\begin{aligned}
&\frac{1}{M}\mathrm{tr}\,(\mathbf{B} - z\mathbf{I}_M)^{-1} - \frac{1}{M}\mathrm{tr}\,(\mathbf{R}\,e(z) - z\mathbf{I}_M)^{-1} \\
&\quad = \frac{1}{M}\mathrm{tr}\left[(\mathbf{B} - z\mathbf{I}_M)^{-1}\mathbf{R}\,(\mathbf{R}\,e(z) - z\mathbf{I}_M)^{-1}\right]e(z) \\
&\qquad - \frac{1}{M}\sum_{i=1}^{K} t_i \frac{\mathbf{x}_i^{\mathrm{H}}\overbrace{\mathbf{R}^{\frac{1}{2}}\,(\mathbf{R}\,e(z) - z\mathbf{I}_M)^{-1}\left(\sum_{j\neq i}^{K} t_j \mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}} - z\mathbf{I}_M\right)^{-1}\mathbf{R}^{\frac{1}{2}}}^{\tilde{\mathbf{A}}}\mathbf{x}_i}{1 + t_i\mathbf{x}_i^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}}\left(\sum_{j\neq i}^{K} t_j \mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}} - z\mathbf{I}_M\right)^{-1}\mathbf{R}^{\frac{1}{2}}\mathbf{x}_i} .
\end{aligned}$$

Now, we see that $\tilde{\mathbf{A}}$ is statistically independent of $\mathbf{x}_i$ and thus we can finally apply the trace Lemma (Lemma 17) in the numerator and denominator. Thus giving us the convergenge $\mathbf{x}_i^{\mathrm{H}}\tilde{\mathbf{A}}\mathbf{x}_i - \frac{1}{K}\,\mathrm{tr}(\tilde{\mathbf{A}}) \xrightarrow[K\to+\infty]{a.s.} 0$. We also remark that the following steps are only valid in the almost sure sense and only for the defined large matrix regime. We will slightly abuse the notation "$\approx$" in the following to mark this restriction, when needed.

$$\begin{aligned}
&\frac{1}{M}\mathrm{tr}\,(\mathbf{B} - z\mathbf{I}_M)^{-1} - \frac{1}{M}\mathrm{tr}\,(\mathbf{R}\,e(z) - z\mathbf{I}_M)^{-1} \\
&\quad \approx \frac{1}{M}\mathrm{tr}\left[(\mathbf{B} - z\mathbf{I}_M)^{-1}\mathbf{R}\,(\mathbf{R}\,e(z) - z\mathbf{I}_M)^{-1}\right]e(z) \\
&\qquad - \frac{1}{M}\sum_{i=1}^{K} t_i \frac{\frac{1}{K}\mathrm{tr}\,\mathbf{R}^{\frac{1}{2}}\,(\mathbf{R}\,e(z) - z\mathbf{I}_M)^{-1}\left(\sum_{j\neq i}^{K} t_j \mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}} - z\mathbf{I}_M\right)^{-1}\mathbf{R}^{\frac{1}{2}}}{1 + t_i\frac{1}{K}\mathrm{tr}\,\mathbf{R}^{\frac{1}{2}}\left(\sum_{j\neq i}^{K} t_j \mathbf{R}^{\frac{1}{2}}\mathbf{x}_j\mathbf{x}_j^{\mathrm{H}}\mathbf{R}^{\frac{1}{2}} - z\mathbf{I}_M\right)^{-1}\mathbf{R}^{\frac{1}{2}}} .
\end{aligned}$$

From the Rank-one-Pertubation lemma (Lemma 22), we know that

$$\operatorname{tr} \mathbf{R}^{\frac{1}{2}} \left(\sum_{j \neq i}^{K} t_j \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j \mathbf{x}_j^{\mathrm{H}} \mathbf{R}^{\frac{1}{2}} - z \mathbf{I}_M \right)^{-1} \mathbf{R}^{\frac{1}{2}}$$

converges (almost surely) to

$$\operatorname{tr} \mathbf{R}^{\frac{1}{2}} \left(\sum_{j=1}^{K} t_j \mathbf{R}^{\frac{1}{2}} \mathbf{x}_j \mathbf{x}_j^{\mathrm{H}} \mathbf{R}^{\frac{1}{2}} - z \mathbf{I}_M \right)^{-1} \mathbf{R}^{\frac{1}{2}} = \operatorname{tr} \mathbf{R}^{\frac{1}{2}} \left(\mathbf{B} - z \mathbf{I}_M \right)^{-1} \mathbf{R}^{\frac{1}{2}} .$$

Therefore, it is possible to write

$$\begin{aligned}
&\frac{1}{M} \operatorname{tr} \left(\mathbf{B} - z \mathbf{I}_M \right)^{-1} - \frac{1}{M} \operatorname{tr} \left(\mathbf{R}\, e(z) - z \mathbf{I}_M \right)^{-1} \\
&\quad \approx \frac{1}{M} \operatorname{tr} \left[\left(\mathbf{B} - z \mathbf{I}_M \right)^{-1} \mathbf{R} \left(\mathbf{R}\, e(z) - z \mathbf{I}_M \right)^{-1} \right] e(z) \\
&\qquad - \frac{1}{M} \sum_{i=1}^{K} t_i \frac{\frac{1}{K} \operatorname{tr} \mathbf{R}^{\frac{1}{2}} \left(\mathbf{R}\, e(z) - z \mathbf{I}_M \right)^{-1} \left(\mathbf{B} - z \mathbf{I}_M \right)^{-1} \mathbf{R}^{\frac{1}{2}}}{1 + t_i \frac{1}{K} \operatorname{tr} \mathbf{R}^{\frac{1}{2}} \left(\mathbf{B} - z \mathbf{I}_M \right)^{-1} \mathbf{R}^{\frac{1}{2}}} \\
&\quad = \frac{1}{M} \operatorname{tr} \left[\left(\mathbf{B} - z \mathbf{I}_M \right)^{-1} \mathbf{R} \left(\mathbf{R}\, e(z) - z \mathbf{I}_M \right)^{-1} \right] e(z) \\
&\qquad - \frac{1}{K} \sum_{i=1}^{K} t_i \frac{\frac{1}{M} \operatorname{tr} \left(\mathbf{B} - z \mathbf{I}_M \right)^{-1} \mathbf{R} \left(\mathbf{R}\, e(z) - z \mathbf{I}_M \right)^{-1}}{1 + t_i \frac{M}{K} \frac{1}{M} \operatorname{tr} \mathbf{R}^{\frac{1}{2}} \left(\mathbf{B} - z \mathbf{I}_M \right)^{-1} \mathbf{R}^{\frac{1}{2}}}
\end{aligned}$$

REMARQUE 25 (Educated Guess). — It might only be at this point where one conclusively sees that our educated guess was advantageous. This choice has resulted in a form tr $[\mathbf{A}]e(z)$ − tr $[\mathbf{A}]x(z)$, where $x(z)$ is a candidate for the wanted DE. Finding DE with the educated guess approach usually relies on much trial and error. □

Collecting the common terms, we finally find

$$\begin{aligned}
&\frac{1}{M} \operatorname{tr} \left(\mathbf{B} - z \mathbf{I}_M \right)^{-1} - \frac{1}{M} \operatorname{tr} \left(\mathbf{R}\, e(z) - z \mathbf{I}_M \right)^{-1} \\
&\quad = \overbrace{\frac{1}{M} \operatorname{tr} \left[\left(\mathbf{B} - z \mathbf{I}_M \right)^{-1} \mathbf{R} \left(\mathbf{R}\, e(z) - z \mathbf{I}_M \right)^{-1} \right]}^{\text{Bounded, as } \mathbf{R} \text{ is bounded.}} \times \\
&\qquad \underbrace{\left[e(z) - \frac{1}{K} \sum_{i=1}^{K} \frac{t_i}{1 + t_i c \frac{1}{M} \operatorname{tr} \mathbf{R} \left(\mathbf{B} - z \mathbf{I}_M \right)^{-1}} \right]}_{\stackrel{!}{\to} 0} .
\end{aligned}$$

Thus, one realizes that choosing $e(z)$ such that the right multiplicative term becomes 0 could give us the wanted result. However, such a result would still contain randomness.

Moreover, the expression in the denominator $\frac{1}{M}\mathrm{tr}\,\mathbf{R}\left(\mathbf{B}-z\mathbf{I}_M\right)^{-1}$ differs from the desired result. If it was $\frac{1}{M}\mathrm{tr}\,\left(\mathbf{B}-z\mathbf{I}_M\right)^{-1}$, we could have closed a loop and could have found a deterministic expression for our original problem. Instead we created a new term, which needs to be evaluated. This will be done in the following.

To solve this problem, we need to restart from the beginning. Yet, this time we begin with the complementary problem $\frac{1}{M}\mathrm{tr}\,\mathbf{R}\left(\mathbf{B}-z\mathbf{I}_M\right)^{-1}$ and "guess" the complementary solution $\frac{1}{M}\mathrm{tr}\,\mathbf{R}\left(\mathbf{R}\,e(z)-z\mathbf{I}_M\right)^{-1}$. This will give us a complementary solution that, as well shall see, combined with the first result will finally admit a closed form solution. Following the (exact) same steps as before:

$$
\begin{aligned}
&\frac{1}{M}\mathrm{tr}\,\mathbf{R}\left(\mathbf{B}-z\mathbf{I}_M\right)^{-1}-\frac{1}{M}\mathrm{tr}\,\mathbf{R}\left(\mathbf{R}\,e(z)-z\mathbf{I}_M\right)^{-1}\\
&\ \vdots\\
&\approx\frac{1}{M}\mathrm{tr}\,\left[\mathbf{R}\left(\mathbf{B}-z\mathbf{I}_M\right)^{-1}\mathbf{R}\left(\mathbf{R}\,e(z)-z\mathbf{I}_M\right)^{-1}\right]\\
&\qquad\left[e(z)-\frac{1}{K}\sum_{i=1}^{K}\frac{t_i}{1+t_i c\frac{1}{M}\mathrm{tr}\,\mathbf{R}\left(\mathbf{B}-z\mathbf{I}_M\right)^{-1}}\right].
\end{aligned}
$$

Now, we finally chose

$$
e(z)=\frac{1}{K}\sum_{i=1}^{K}\frac{t_i}{1+t_i c\frac{1}{M}\mathrm{tr}\,\mathbf{R}\left(\mathbf{B}-z\mathbf{I}_M\right)^{-1}}
$$

which simultaneously solves this and also our previous "guess". From the second "guess" we also see that (for this particular choice of $e(z)$), we have $\frac{1}{M}\mathrm{tr}\,\mathbf{R}\left(\mathbf{B}-z\mathbf{I}_M\right)^{-1}-\frac{1}{M}\mathrm{tr}\,\mathbf{R}\left(\mathbf{R}\,e(z)-z\mathbf{I}_M\right)^{-1}\xrightarrow{\text{a.s.}}0$. Based on this observation, another choice for $e(z)$, which is fully deterministic will be

$$
e(z)=\frac{1}{K}\sum_{i=1}^{K}\frac{t_i}{1+t_i c\frac{1}{M}\mathrm{tr}\,\mathbf{R}\left(\mathbf{R}\,e(z)-z\mathbf{I}_M\right)^{-1}}
$$

which is an iteratively solvable fixed-point equation [8]. As a next step, one would need to show that the fixed-pint equation has a unique solution and that it converges in the first place. This could be achieved relatively easily by using the *standard interference functions framework* from (Yates, 1995), as shown in (Hoydis, 2012, Theorems 22, 23, 24).

6. Conclusion

In this paper we provided an introduction to the practical application of RMT in general (and DEs in particular) for the analysis of communication problems. To this

8. We recognize that the last step in this example is somewhat more intuitive than rigorous.

end, we first established some necessary basic theoretical concepts, lemmas and tools from the field of RMT. This knowledge was then applied and illustrated in several step by step derivations, showcasing the approach in communication problems for simple and slightly more advanced systems and models. While RMT is often of tremendous use, one should also keep its limitations in mind. For example, one should be mindful of the sometimes deteriorating performance given "badly conditioned" resolvents and the possibly relatively slow convergence of the DEs to their respective random quantity. Also the tightness of the DE is not guaranteed to be the same for each choice of system variables, thus a common sense approach to interpreting obtained results and occasional verification by classical Monte-Carlo techniques is advised. Still, as is seen in many published works, RMT is a very robust and useful approach for the abstraction of complex problems involving potentially large systems, that relies on few system parameters and also holds for relatively small system sizes.

Biographie (AR_Inter2)

Les revues Traitement du signal (TS) et ... demandent à leurs auteurs une biographie courte : 5 lignes maximum, avec leurs domaines d?étude et de recherche :

Robert Dupont est chercheur en ... Il est actuellement en charge de au sein de l?équipe du laboratoire. Etc.

Remerciements

bla bla blub

Bibliographie

Bai Z. D., Silverstein J. W. (1998a). No Eigenvalues outside the Support of the Limiting Spectral Distribution of Large-dimensional Random Matrices. *The Annals of Applied Probability*, vol. 26, p. 316-345.

Bai Z. D., Silverstein J. W. (1998b, janvier). No Eigenvalues Outside the Support of the Limiting Spectral Distribution of Large Dimensional Sample Covariance Matrices. *Annals of Probability*, vol. 26, n° 1, p. 316-345.

Bai Z. D., Silverstein J. W. (2007). On the Signal-to-Interference Ratio of CDMA Systems in Wireless Communications. *The Annals of Applied Probability*, vol. 17, p. 81-101.

Bai Z. D., Silverstein J. W. (2009). *Spectral Analysis of Large Dimensional Random Matrices* (2nd edition éd.). Springer Series in Statistics.

Billingsley P. (1995). *Probability and Measure* (Third éd.). Hoboken, NJ, John Wiley & Sons, Inc.

Couillet R. (2010). *Application of Random Matrix Theory to Future Wireless Flexible Networks*. Thèse de doctorat non publiée, Supélec.

Couillet R., Debbah M. (2009). Uplink Capacity of Self-Organizing Clustered Orthogonal CDMA Networks in Flat Fading Channels. In *Information theory workshop, 2009. itw 2009. ieee.* Taormina, Sicily.

Couillet R., Debbah M. (2011). *Random Matrix Methods for Wireless Communications*. New York, NY, USA, Cambridge University Press. (First Edition)

Couillet R., Debbah M., Silverstein J. W. (2011). A Deterministic Equivalent for the Analysis of Correlated MIMO Multiple Access Channels. *IEEE Transactions on Information Theory*, vol. 57, n^{o} 6, p. 3493–3514.

Dohler M. (2003). *Virtual Antenna Arrays*. Thèse de doctorat non publiée, University of London.

Dupuy F., Loubaton P. (2010). On the Capacity Achieving Covariance Matrix for Frequency Selective MIMO Channels Using the Asymptotic Approach. *IEEE Transactions on Information Theory*. Consulté sur http://arxiv.org/abs/1001.3102

Hachem W., Khorunzhy O., Loubaton P., Najim J., Pastur L. A. (2008, Sept.). A New Approach for Capacity Analysis of Large Dimensional Multi-Antenna Channels. *IEEE Transactions on Information Theory*, vol. 54, n^{o} 9, p. 3987-4004.

Hachem W., Loubaton P., Najim J. (2007). Deterministic Equivalents for Certain Functionals of Large Random Matrices. *Annals of Applied Probability*, vol. 17, n^{o} 3, p. 875-930.

Hachem W., Loubaton P., Najim J. (2008, décembre). A CLT for Information Theoretic Statistics of Gram Random Matrices with a Given Variance Profile. *Annals of Probability*, vol. 18, n^{o} 6, p. 2071-2130.

Hoydis J. (2010). *Channel Modeling for Cooperative Multi-Cell Systems.* (Unpublished Lecture Notes)

Hoydis J. (2012). *Random Matrix Methods for Advanced Communication Systems*. Thèse de doctorat non publiée, Supélec, Alcatel-Lucent Chair on Flexible Radio.

Hoydis J., Kobayashi M., Debbah M. (2011). Optimal Channel Training in Uplink Network MIMO Systems. *IEEE Transactions on Signal Processing*, vol. 59, n^{o} 6, p. 2824-2833.

Kammoun A., Kharouf M., Hachem W., Najim J. (2009, Nov.). A Central Limit Theorem for the SINR at the LMMSE Estimator Output for Large-Dimensional Signals. *IEEE Transactions on Information Theory*, vol. 55, n^{o} 11, p. 5048–5063.

Krein M. K., Nudelman A. A. (1977). *The Markov Moment Problem and Extremal Problems*. Providence, RI, USA, American Mathematical Society.

Marčenko V. A., Pastur L. A. (1967, avril). Distributions of Eigenvalues for Some Sets of Random Matrices. *Math USSR-Sbornik*, vol. 1, n^{o} 4, p. 457-483.

Müller A. (2014). *Random matrix analysis of future multi cell mu-mimo networks*. Thèse de doctorat non publiée, SUPÉLEC.

Pastur L. A. (1999). A Simple Approach to Global Regime of Random Matrix Theory. In *Mathematical Results in Statistical Mechanics*, p. 429-454. World Scientific Publishing.

Pastur L. A., Shcherbina M. (2011). *Eigenvalue distribution of large random matrices* n^{o} 171. American Mathematical Soc.

Peacock M. J. M., Collings I. B., Honig M. L. (2008). Eigenvalue Distributions of Sums and Products of Large Random Matrices via Incremental Matrix Expansions. *IEEE Transactions on Information Theory*, vol. 54, n^{o} 5, p. 2123-2138.

Rubio F., Mestre X., Hachem W. (2012). A CLT on the SNR of Diagonally Loaded MVDR Filters. *Signal Processing, IEEE Transactions on*, vol. 60, n^{o} 8, p. 4178–4195.

Silverstein J. W., Bai Z. D. (1995). On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices. *Journal of Multivariate Analysis*, vol. 54, nº 2, p. 175-192.

Tao T. (2012). *Topics in Random Matrix Theory* (vol. 132). American Mathematical Soc.

Telatar E. (1999). Capacity of Multi-antenna Gaussian Channels. *European Transactions on Telecommunications*, vol. 10, nº 6, p. 585–595. (Also, 1995 in Technical Memorandum, Bell Laboratories, Lucent Technologies)

Tse D. N. C., Hanly S. V. (1999, février). Linear Multiuser Receivers: Effective Interference, Effective Bandwidth and User Capacity. *IEEE Transactions on Information Theory*, vol. 45, nº 2, p. 641-657.

Tulino A., Verdú S. (2004). Random Matrix Theory and Wireless Communications. *Foundations and Trends in Communications and Information Theory*, vol. 1, nº 1, p. 1-182.

Vallet P., Loubaton P., Mestre X. (2010). Improved Subspace Estimation for Multivariate Observations of High Dimension: The Deterministic Signals Case. *IEEE Transactions on Information Theory*. Consulté sur http://arxiv.org/abs/1002.3234

Van der Vaart A. W. (2000). *Asymptotic Statistics*. New York, Cambridge University Press.

Verdú S., Shamai S. (1999, février). Spectral Efficiency of CDMA with Random Spreading. *IEEE Transactions on Information Theory*, vol. 45, nº 2, p. 622-640.

Wagner S., Couillet R., Debbah M., Slock D. (2012, July). Large System Analysis of Linear Precoding in MISO Broadcast Channels with Limited Feedback. *IEEE Transactions on Information Theory*, vol. 58, nº 7, p. 4509-4537.

Wigner E. (1958, mars). On the Distribution of Roots of Certain Symmetric Matrices. *The Annals of Mathematics*, vol. 67, nº 2, p. 325-327.

Yates R. D. (1995). A Framework for Uplink Power Control in Cellular Radio Systems. *IEEE Journal on Selected Areas in Communications*, vol. 13, nº 7, p. 1341–1347.

Axel Müller. *Biography of the first author...*

Mérouane Debbah. *Biography of the second author...*