

On the Positivity of Hadamard Powers of Random Matrices

Tiefeng Jiang

School of Statistics

University of Minnesota

224 Church Street SE, MN 55455

Email: jiang040@umn.edu

Belhal Karimi and Ping Li

Cognitive Computing Lab

Baidu Research

10900 NE 8th St. Bellevue, WA 98004

Email: {belhal.karimi, pingli98}@gmail.com

Abstract—The¹ paper studies the relation between the positive definiteness of a random matrix and its Hadamard power. Given important applications of random matrix theory and their Hadamard power, of exponent α , in various fields such as kernel learning, graphs learning and more, we establish a collection of theoretical results ensuring the existence of an interval of non integer exponent ensuring that given a *random* positive definite matrix, its *Hadamard* power can be in high probability, not non-negative. Based on mild assumptions on the support of entries of any random matrices, our results give a better understanding and a counterintuitive property of the Hadamard powers of random matrices. Key ingredients in the proof include a novel consideration of the supports of each random entry of our considered matrices, coupled with their probabilistic interpretation.

Keywords: Hadamard matrix function, Hadamard power, non-negative definite matrix, positive definite matrix, random matrix.

I. INTRODUCTION

Random matrices and Hadamard product, the two main topics of this paper, are of utmost importance in various applications, such as physics, finance, telecommunication, computational biology and machine learning for the random matrices, and combinatorial analysis, number theory or regular graphs to name a few for the Hadamard product, and in information theory. The Hadamard product, also known as the entry-wise or Schur product, is a type of matrix multiplication that is commutative and plays an important role as put in [1]. Several noteworthy applications and studies of the Hadamard product can be found in the literature, as in [2]–[5].

In particular, in our contribution, we focus on the Hadamard power of random matrices. Given $f(x)$, a real-valued function defined on \mathbb{R} . Let $\mathbf{A} = (a_{ij})$ be an $n \times n$ matrix, where a_{ij} 's are real numbers. Define $f : \mathbf{A} \rightarrow f(\mathbf{A}) = (f(a_{ij}))$. We call $f(\mathbf{A})$ a

Hadamard function to distinguish it from the usual notion of matrix functions, as here the function is applied to each element of the matrix. In particular, if $\alpha > 0$ and $f(x) = x^\alpha$, then we call $\mathbf{A}^{(\alpha)} := f(\mathbf{A})$ the Hadamard power of α .

Main Motivations: An important application of Hadamard powers of matrices is the kernel learning problem where the exponent, used in for instance polynomial kernels or Gaussian Radial Basis Function (RBF), is a tunable parameter, see [6]–[8]. In particular, positive definite kernels and the set of methods they have inspired in the machine learning literature, namely kernel methods, are particularly sensitive to the choice of their exponent. After having chosen a family of kernel, the task of choosing the adequate hyperparameters for this kernel is one of the biggest challenges when using kernel methods, as described in Section 12.3.4 of [9]. For instance, for polynomial kernels or Gaussian kernels, the searches are limited to the exponent parameters, see [10]–[12] for more details on the tuning of such kernels.

Kernel methods in particular, see [13] for enhanced details, are an important part of modern machine learning. They can be coupled with hashing techniques when large-scale optimization task is required to improve the overall computational efficacy of the methods. For large-scale machine learning, particular appealing to most modern applied problems, hashing and kernel methods are drawing a growing interest in the community and can be. In particular, in [7], the generalized min-max (GMM) kernel is thoroughly studied in order to provide state-of-the-art results. The GMM kernel can be derived as various tunable kernels such as the Gamma GMM kernel that leverages an exponent as a tunable hyperparameter for enhanced performances. A thorough empirical study results in the choice of non-integers powers as best hyperparameter for a broad range of tasks, which again highlights the importance of our contribution.

Our Contributions: We provide in this contribution several important and intriguing results on the

¹The work of Tiefeng Jiang was conducted as consulting researcher at Baidu Research.

positivity of Hadamard powers of random matrices. While we show in Theorem 4 that for deterministic and square matrices with non-negative entries, their Hadamard powers remain positive definite, we also provide in Theorem 5 and Theorem 6, that even though Hadamard products preserve positive semidefiniteness, we are able to prove that for some non integer exponent values, the positive definite random matrices have their Hadamard powers non non-negative definite.

Section II is devoted to the main concepts, notations and existing results to lay the core foundation of our study. Several illustrative examples are also provided for the sake of clarity. Section III develops the main results of our paper and their proofs are in Section IV. Section V concludes our work.

II. NOTATIONS, EXISTING RESULTS AND EXAMPLES

We recall in this section some important results in the literature followed by our main theoretical contributions to the domain of random matrices. In the definition of the Hadamard power stated in the introduction, careful attention needs to be paid to the domain of the function $f(x) = x^\alpha$. If the power $\alpha > 0$ is an integer, the function is defined for every $x \in \mathbb{R}$. If $\alpha > 0$ is not an integer, the function $f(x) = x^\alpha$ is defined only on $[0, \infty)$. By the Schur product theorem, it is known that $\mathbf{A}^{(\alpha)}$ is a positive definite matrix if $\mathbf{A} = (a_{ij})$ is a positive definite matrix and $\alpha = 1, 2, \dots$; see, for example, Theorem 5.2.1 from [14]. Besides, for positive definite matrices $\mathbf{U} = (u_{ij})_{n \times n}$ and $\mathbf{V} = (v_{ij})_{n \times n}$, set $\mathbf{U} \circ \mathbf{V} = (u_{ij}v_{ij})_{n \times n}$. Then $\lambda_{\min}(\mathbf{U}) \cdot \min_{1 \leq j \leq n} v_{jj} \leq \lambda_i(\mathbf{U} \circ \mathbf{V}) \leq \lambda_{\max}(\mathbf{U}) \cdot \max_{1 \leq j \leq n} v_{jj}$ for each $1 \leq i \leq n$; see, for example, [15] or Theorem 5.3.4 from [14]. If α is a positive integer, then $\mathbf{A}^{(\alpha)} = \mathbf{A} \circ \dots \circ \mathbf{A}$ from which there are α times \mathbf{A} in the product. Thus, by induction, we have that $\lambda_{\min}(\mathbf{A}) > 0$ if \mathbf{A} is positive definite.

A. Existing Results

Let $a \in (0, \infty]$ and $f : (0, a) \rightarrow \mathbb{R}$. We say that $f(x)$ is *absolutely monotonic* on $(0, a)$ if $f^{(k)}(x) \geq 0$ for every $x \in (0, a)$ and $k \geq 0$. The following general conclusion can be found in several research articles including [16], [17] and [18]. Theorem 1 gives a formal result regarding the monotonicity of a value function $f(\cdot)$ and a non-negative definite matrix \mathbf{A} extracted from point (i) of Theorem 4.1 from [18].

THEOREM 1: (Theorem 4.1 from [18]) Assume $a \in (0, \infty]$ and $f(x)$ is a real function defined on $(-a, a)$. Then $f(\mathbf{A})$ is non-negative definite for every non-negative definite matrix \mathbf{A} with entries in

$(-a, a)$ if and only if $f(x)$ is analytic and absolutely monotonic on $(0, a)$.

THEOREM 2: (Theorem 6.3.7 from [14]) Let $f(\cdot)$ be an $(n-1)$ -times continuously differentiable real valued function on $(0, \infty)$, and suppose that the Hadamard function $f(\mathbf{A}) = (f(a_{ij}))$ is non-negative definite for every non-negative definite matrix \mathbf{A} that has positive entries. Then:

$$f^{(k)}(t) \geq 0$$

for all $t \in (0, \infty)$ and all $k \in [0, n-1]$.

An additional result can be derived from Theorem 2 stated above, and states:

COROLLARY 2.1: (Corollary 6.3.8 from [14]) Let $0 < \alpha < n-2$, where α is not an integer. There exists some $n \times n$ non-negative definite matrix noted \mathbf{A} with positive entries, such that the Hadamard power of \mathbf{A} noted $\mathbf{A}^{(\alpha)} = (a_{ij}^\alpha)$ is not non-negative definite.

This Corollary translates the result of the Theorems stated above in the particular case when the value function $f(\cdot)$ maps a non-negative definite matrix into its Hadamard power. The equivalence with a non-negative definite Hadamard power matrix is then a consequence of the above. Another important existing result worth noting, as an introduction to our novel results introduced in the sequel, is a quantification of an infimum value for the power α used in the Hadamard power operation. The following theorem states as follows:

THEOREM 3: (Theorem 6.3.9 from [14]) Let $\mathbf{A} = (a_{ij})$ be a non-negative definite matrix with non-negative entries. If $\alpha \geq n-2$, then the Hadamard power $\mathbf{A}^{(\alpha)}$ is non-negative definite.

B. Some Illustrative Examples

For the sake of clarity, we introduce the following notations, used throughout the following sections including the several examples we provide for illustrative purposes and the statements of our main theoretical results. For a matrix \mathbf{M} , we note $\mathbf{M} > 0$ if \mathbf{M} is positive definite; $\mathbf{M} \geq 0$ if \mathbf{M} is non-negative definite; $\mathbf{M} \not\geq 0$ if \mathbf{M} is not non-negative definite.

We now provide an example of such *not non-negative definite* matrix through the use of the Hadamard power operation.

EXAMPLE 2.1: Consider the following 3×3 matrix

$$\mathbf{M} = \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & \frac{1}{2} & 1 \end{pmatrix}.$$

Its Hadamard power of power α , where α is not an integer, is given by

$$\mathbf{M}^{(\alpha)} = \begin{pmatrix} 1 & \frac{1}{2^\alpha} & 0 \\ \frac{1}{2^\alpha} & 1 & \frac{1}{2^\alpha} \\ 0 & \frac{1}{2^\alpha} & 1 \end{pmatrix}.$$

Then, the computation of its determinant yields

$$\det(\mathbf{M}^{(\alpha)}) = 1 - \frac{2}{4^\alpha}.$$

Therefore, we have that $\mathbf{M} = \mathbf{M}^{(1)} > 0$. However, $\mathbf{M}^{(\alpha)} \not\geq 0$ if $\alpha \in (0, \frac{1}{2})$.

Now, for any integer $n \geq 3$, we define the matrix noted \mathbf{M}_n and defined as $\mathbf{M}_n = \mathbf{M}$ for $n = 3$ and

$$\mathbf{M}_n = \begin{cases} \mathbf{M}, & \text{if } n = 3 \\ \begin{pmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-3} \end{pmatrix}, & \text{if } n \geq 4 \end{cases}$$

where \mathbf{I}_{n-3} denotes the identity matrix of dimension $n - 3$. Then, the matrix \mathbf{M}_n is positive definite as $n \geq 3$. However, we note that its Hadamard power $\mathbf{M}_n^{(\alpha)} \not\geq 0$ is not non-negative definite when $\alpha \in (0, \frac{1}{2})$.

The example given above shows that the construction of a sequence of positive definite matrix, in dimension 3, while its Hadamard power is not non-negative, for a certain value of power α , is achievable. For better understanding, we provide the following additional example in dimension 4.

EXAMPLE 2.2: Consider a 4×4 matrix defined as

$$\mathbf{M} = \mathbf{a}\mathbf{a}' + \mathbf{b}\mathbf{b}' + 10^{-4}\mathbf{I}_4,$$

where $\mathbf{a}^T = (1, 1, 1, 1)$ and $\mathbf{b}^T = (0, 1, 2, 3)$. Note that the additive term $10^{-4}\mathbf{I}_4$ purely ensures the positivity of the resulting \mathbf{M} , thus being a positive definite one and falling into the framework of our study. The matrix $\mathbf{a}\mathbf{a}' + \mathbf{b}\mathbf{b}'$ is of rank 2. Obviously, $\mathbf{M} > 0$. It is also easy to check that

$$\det(\mathbf{M}^{1.1}) = -0.000118654.$$

Hence, $\mathbf{M}^{1.1} \not\geq 0$. Similarly to the previous example, define the sequence of matrices \mathbf{M}_n defined as follows:

$$\mathbf{M}_n = \begin{cases} \mathbf{M}, & \text{if } n = 4 \\ \begin{pmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-4} \end{pmatrix}, & \text{if } n \geq 5 \end{cases}$$

Then, we observe that for $n \geq 4$ we have that $\mathbf{M}_n > 0$. However, its Hadamard power matrix $\mathbf{M}_n^{1.1} \not\geq 0$ is not non-negative definite.

Those simple illustrative examples lay the context of the next section, where several new results are provided, along with their proofs.

III. MAIN RESULTS

We now introduce a collection of theorems aiming at improving the theoretical understanding of Hadamard powers of random matrices.

A. Non-negative matrices

Following the extensive literature presented in the introduction of our paper, we provide the following result for non-negative (and deterministic) definite matrices:

THEOREM 4: Let $\mathbf{A} = (a_{ij})$ be an $n \times n$ matrix of which the entries are non-negative. Assume $a_{ii} \geq \sum_{j \neq i} a_{ij}$ for each $1 \leq i \leq n$. Then $\mathbf{A} \geq 0$ and $\mathbf{A}^{(\alpha)} \geq 0$ for all $\alpha \geq 1$. Note that the conclusion still holds if all three “ \geq ” are replaced by “ $>$ ”, i.e., if $a_{ii} > \sum_{j \neq i} a_{ij}$ for each $1 \leq i \leq n$, then $\mathbf{A} > 0$ and $\mathbf{A}^{(\alpha)} > 0$ for all $\alpha > 1$.

B. Random Matrices

Here, we consider random matrices, in the sense that the entries of the matrices we study are independent random variables. Then, under some mild assumptions, that we rigorously give in each statement, we provide the existence, in probability, of positive definitive matrices while their Hadamard power, for some values of exponent α , is not non-negative definite. To begin with, considering large symmetric matrices of size $n \geq 4$, it is possible to derive the following theoretical result:

THEOREM 5: Assume $n \geq 4$. Let $\mathbf{M} = (\xi_{ij})$ be an $n \times n$ symmetric matrix, where $\{\xi_{ij}; 1 \leq i \leq n, 1 \leq j \leq n\}$ are independent random variables. Suppose all of the supports of ξ_{ij} ’s contain a common interval $[u, v]$ for some $v > u > 0$. Then there exists $\alpha \geq 1$ for which

$$P(\mathbf{M} \geq 0 \text{ and } \mathbf{M}^{(\alpha)} \not\geq 0) > 0.$$

On the same line of work, we also develop a similar result for non-square matrices. Then under the same mild assumptions on the support of the random variables constituting the matrices involved and the value of the non integer exponent, we state the following:

THEOREM 6: Assume $n \geq 4$. Let $\mathbf{X} = (x_{ij})$ be an $n \times p$ matrix, where $\{x_{ij}; 1 \leq i \leq n, 1 \leq j \leq p\}$ are independent random variables. Suppose all of the supports of ξ_{ij} ’s contain a common interval $[u, v]$ for some $v > u > 0$. Then there exists $\alpha \geq 1$ such that

$$P(\mathbf{X}^T \mathbf{X} > 0 \text{ and } (\mathbf{X}^T \mathbf{X})^{(\alpha)} \not\geq 0) > 0.$$

IV. PROOFS OF THE MAIN RESULTS

Using intermediary Lemmas, that we rigorously state and prove in this section, we now provide the proofs of the main results presented above.

A. Intermediary Theoretical Results

The following result guarantees the existence of a positive definite symmetric matrix, where the support of its entries are well defined, such that its Hadamard power, in a range of exponent, is not non-negative.

LEMMA 1: Let $\mathbf{A} = (a_{ij})$ be a given $n \times n$ matrix with non-negative entries. For any $n \geq 4$, there exist $\alpha = 1.1$, $\delta > 0$ and an $n \times n$ symmetric matrix $\mathbf{M} = (m_{ij})$ where for all $1 \leq i, j \leq n$ we have $m_{ij} \geq 0$, such that the following holds: (i) $\mathbf{M} = (m_{ij}) > 0$ for every $m_{ij} \in [a_{ij}, a_{ij} + \delta]$ and $1 \leq i, j \leq n$.

(ii) $\mathbf{M}^{(\alpha)} = (m_{ij}^{(\alpha)}) \not\geq 0$ for any $m_{ij} \in [a_{ij}, a_{ij} + \delta]$ and any $1 \leq i, j \leq n$.

Proof of Lemma 1. For any $n \times n$ symmetric matrix $\mathbf{M} = (m_{ij})$, we denote by $\|\mathbf{M}\|$ the spectral norm of \mathbf{M} . We use $\lambda_1(\mathbf{M}) \geq \lambda_2(\mathbf{M}) \geq \dots \geq \lambda_n(\mathbf{M})$ to denote the eigenvalues of \mathbf{M} . It is well established that $\|\mathbf{M}\| \leq (\sum_{1 \leq i, j \leq n} |m_{ij}|^2)^{1/2}$. Also denote by $\mathbf{M}_1 = (m_{ij})$ and $\mathbf{M}_2 = (\tilde{m}_{ij})$, two symmetric matrices of dimension $n \times n$. The Weyl's perturbation theorem [see, e.g., [19]] states that

$$\max_{1 \leq i \leq n} |\lambda_i(\mathbf{M}_1) - \lambda_i(\mathbf{M}_2)| \leq \|\mathbf{M}_1 - \mathbf{M}_2\|.$$

Therefore, we obtain the following

$$\begin{aligned} & \max_{1 \leq i \leq n} |\lambda_i(\mathbf{M}_1) - \lambda_i(\mathbf{M}_2)| \\ & \leq \left(\sum_{1 \leq i, j \leq n} |m_{ij} - \tilde{m}_{ij}|^2 \right)^{1/2}. \end{aligned} \quad (4.1)$$

The inequality above concludes that the eigenvalues of a matrix are continuous functions of its entries. This is particularly true for the smallest eigenvalues.

According to Example 2.2, there exists $\alpha \geq 1$ and an $n \times n$ symmetric matrix $\mathbf{A} = (a_{ij})$ such that $a_{ij} \geq 0$ for all $1 \leq i, j \leq n$, we have $\mathbf{A} > 0$ and its Hadamard power matrix $\mathbf{A}^{(\alpha)} \not\geq 0$. For any $n \times n$ symmetric matrix $\mathbf{M} = (m_{ij})$, let us define

$$f(\mathbf{M}) := \min \{ \lambda_n(\mathbf{M}), -\lambda_n(\mathbf{M}^{(\alpha)}) \}.$$

As explained above, $f(\mathbf{M})$ is a continuous function in the entries $\{m_{ij}; 1 \leq i \leq j \leq n\}$ of the matrix \mathbf{M} . Since $f(\mathbf{A}) > 0$, there exist $\{\delta_{ij} > 0; 1 \leq i, j \leq n\}$ with $\delta_{ij} = \delta_{ji}$ for all $1 \leq i, j \leq n$ such that $f(\mathbf{M}) > 0$ for any entry $m_{ij} \in [a_{ij}, a_{ij} + \delta_{ij}]$ where $1 \leq i, j \leq n$. Set $\delta := \min\{\delta_{ij}; 1 \leq i \leq j \leq n\}$. Then, $\delta > 0$. Also, $\lambda_n(\mathbf{M}) > 0$ and $\lambda_n(\mathbf{M}^{(\alpha)}) < 0$ for every $m_{ij} \in [a_{ij}, a_{ij} + \delta]$ and every $1 \leq i, j \leq n$. In summary, this boils down to observing that $\mathbf{M} > 0$ and $\mathbf{M}^{(\alpha)} \not\geq 0$ for any $m_{ij} \in [a_{ij}, a_{ij} + \delta]$ and any $1 \leq i, j \leq n$. which concludes the proof of our Lemma. \square

We now derive a similar result true for any non-square matrix whose entries supports are within some

specific range. Note that the following Lemma holds for any matrix $\mathbf{X} = (x_{ij})_{n \times p} \in \mathbb{R}^{n \times p}$.

LEMMA 2: For any n and p with $n \geq p \geq 4$, there exist $\alpha \geq 1$, $\delta > 0$ and $n \times p$ matrix $\mathbf{A} = (a_{ij})$ with $a_{ij} \geq 0$ for all $1 \leq i \leq n$ and $1 \leq j \leq p$ such that the following holds:

(i) The matrix $\mathbf{X}'\mathbf{X}$ is positive definite for every matrix \mathbf{X} with entries $x_{ij} \in [a_{ij}, a_{ij} + \delta]$ and $1 \leq i \leq n$ and $1 \leq j \leq p$.

(ii) The Hadamard power $(\mathbf{X}'\mathbf{X})^{(\alpha)} \not\geq 0$ for any $x_{ij} \in [a_{ij}, a_{ij} + \delta]$, $1 \leq i \leq n$ and $1 \leq j \leq p$.

Proof of Lemma 2. Let $\mathbf{a}^T = (1, 1, 1, 1)$ and $\mathbf{b}^T = (0, 1, 2, 3)$ be as in Example 2.2. It can be shown that the Hadamard power $(\mathbf{a}\mathbf{a}^T + \mathbf{b}\mathbf{b}^T)^{(1.1)}$ has determinant -1.1856×10^{-4} . Set

$$\mathbf{A}(\epsilon) = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & \epsilon & 0 \\ 0 & 0 & 0 & \epsilon \end{pmatrix}, \quad \epsilon > 0.$$

We can observe that $\lim_{\epsilon \rightarrow 0^+} [\mathbf{A}(\epsilon)^T \mathbf{A}(\epsilon)]^{(1.1)} = (\mathbf{a}\mathbf{a}^T + \mathbf{b}\mathbf{b}^T)^{(1.1)}$ with the entrywise convergence. By continuity of determinants, there exists $\epsilon_0 > 0$ such that the determinant of $[\mathbf{A}(\epsilon_0)^T \mathbf{A}(\epsilon_0)]^{(1.1)}$ is negative. That is, $\mathbf{A}(\epsilon_0)^T \mathbf{A}(\epsilon_0) > 0$ but the Hadamard power $[\mathbf{A}(\epsilon_0)^T \mathbf{A}(\epsilon_0)]^{(1.1)} \not\geq 0$. Now we define an $n \times p$ matrix \mathbf{A} such that

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}(\epsilon_0) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-4} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}_{n \times p},$$

where the size of each submatrix $\mathbf{0}$ of \mathbf{A} can be derived from those of $\mathbf{A}(\epsilon_0)$ and \mathbf{I}_{p-4} . In particular, the size of the “0” in the bottom-right of \mathbf{A} is $(n - p) \times (p - 4)$. In case $n = p$, there is no third row of submatrices in \mathbf{A} ; in case $p = 4$, there is no second row of submatrices of \mathbf{A} . Since we have

$$\mathbf{A}^T \mathbf{A} = \begin{pmatrix} \mathbf{A}(\epsilon_0)^T \mathbf{A}(\epsilon_0) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{p-4} \end{pmatrix}.$$

Hence, $\mathbf{A}^T \mathbf{A} > 0$, nevertheless, its Hadamard power, of exponent 1.1 satisfies $(\mathbf{A}^T \mathbf{A})^{(1.1)} \not\geq 0$. The inequality (4.1) shows that the smallest eigenvalue $\lambda_n(\mathbf{M})$ of $\mathbf{M} = \mathbf{X}\mathbf{X}^T$ is a continuous function of the entries of \mathbf{M} , which in turn are the continuous functions of the entries of \mathbf{X} . Let $\mathbf{X} = (x_{ij})_{n \times p}$. Hence, $\lambda_n(\mathbf{M})$ is a continuous function of the entries x_{ij} 's. Set

$$f(\mathbf{M}) := \min \{ \lambda_n(\mathbf{M}), -\lambda_n(\mathbf{M}^{(\alpha)}) \}.$$

Then $f(\mathbf{M})$ is a continuous function of x_{ij} 's and $f(\mathbf{A}\mathbf{A}^T) > 0$. Write $\mathbf{A} = (a_{ij})_{n \times p}$. Then there exist $\delta_{ij} > 0$ for all $1 \leq i \leq n$ and $1 \leq j \leq p$ such that $f(\mathbf{M}) > 0$ for all $x_{ij} \in [a_{ij}, a_{ij} + \delta_{ij}]$ with

$1 \leq i \leq n$ and $1 \leq j \leq p$. Denote $\delta = \min\{\delta_{ij}; 1 \leq i \leq n, 1 \leq j \leq p\}$. Then $\delta > 0$ and $f(\mathbf{X}\mathbf{X}^T) > 0$ for all $x_{ij} \in [a_{ij}, a_{ij} + \delta]$ with $1 \leq i \leq n$ and $1 \leq j \leq p$. Hence, under these restrictions of x_{ij} 's, we have $\lambda_n(\mathbf{X}\mathbf{X}^T) > 0$ and $\lambda_n((\mathbf{X}\mathbf{X}^T)^{(\alpha)}) < 0$. This yields the proofs for (i) and (ii) of Lemma 2. \square

B. Proofs of Theorems 4-5-6

The proof of Theorem 4, dealing with non-square deterministic matrices, reads as follows:

Proof of Theorem 4. By the Gershgorin Disk Theorem (see for instance [19]), all eigenvalues of \mathbf{A} are in the set

$$\bigcup_{1 \leq i \leq n} \left(a_{ii} - \sum_{j \neq i} a_{ij}, a_{ii} + \sum_{j \neq i} a_{ij} \right). \quad (4.2)$$

By assumption, all eigenvalues are non-negative, hence $\mathbf{A} \geq 0$. On the other hand,

$$a_{ii}^\alpha \geq \left(\sum_{j \neq i} a_{ij} \right)^\alpha \geq \sum_{j \neq i} a_{ij}^\alpha$$

for all $\alpha \geq 1$ by the given condition. Using the Gershgorin Disk theorem again, all of the eigenvalues of the Hadamard power matrix $\mathbf{A}^{(\alpha)}$ are non-negative. Therefore, it is set that $\mathbf{A}^{(\alpha)} \geq 0$. Evidently, if $a_{ii} > \sum_{j \neq i} a_{ij}$ for each $1 \leq i \leq n$ then all of the eigenvalues of \mathbf{A} and $\mathbf{A}^{(\alpha)}$ are positive by (4.2) with “ a_{ij} ” being replaced by a_{ij}^α for all i and j . Hence $\mathbf{A} > 0$ and $\mathbf{A}^{(\alpha)} > 0$ for all $\alpha \geq 1$. \square

Proof of Theorem 5. Based on the results obtained in Lemma 1, we now give a detailed proof of Theorem 5. Let $\delta > 0$ be the value defined in Lemma 1 as the length of the interval defining the supports of the entries of a matrix. Since $[a_{ij} + \frac{1}{2}\delta, a_{ij} + \delta] \subset [a_{ij}, a_{ij} + \delta]$ for each pair of (i, j) with $1 \leq i \leq j \leq n$, then Lemma 1 still holds if we strengthen its conclusion with the requirement of having $a_{ij} > 0$ for all $1 \leq i \leq j \leq n$. Therefore, we obtain

$$0 < \alpha < \beta,$$

where $\alpha := \min\{a_{ij}; 1 \leq i \leq j \leq n\}$ and $\beta := \max\{a_{ij}; 1 \leq i \leq j \leq n\} + \delta$. In the sequel, we denote by $\text{supp}(\xi)$, the support of a random variable ξ . In particular, we have that $P(a \leq \xi \leq b) > 0$ provided $[a, b] \subset \text{supp}(\xi)$. Notice that $\text{supp}(\lambda\xi_{ij}) = \lambda \cdot \text{supp}(\xi_{ij})$ for each i, j where $1 \leq i \leq j \leq n$. Choose $\lambda > 0$ such that $\lambda[u, v] \supset [\alpha, \beta]$. Then, it follows that

$$\bigcup_{1 \leq i \leq j \leq n} [a_{ij}, a_{ij} + \delta] \subset [\alpha, \beta] \quad (4.3)$$

$$\subset \bigcap_{1 \leq i \leq j \leq n} \text{supp}(\lambda\xi_{ij}).$$

Besides, observe that

$$\begin{aligned} & P(\mathbf{M} > 0 \text{ and } \mathbf{M}^{(\alpha)} \not\geq 0) \\ &= P(\lambda\mathbf{M} > 0 \text{ and } (\lambda\mathbf{M})^{(\alpha)} \not\geq 0). \end{aligned} \quad (4.4)$$

By Lemma 1 and the independence assumption, the last probability in (4.4) reads

$$\begin{aligned} & P(\lambda\xi_{ij} \in [a_{ij}, a_{ij} + \delta] \text{ for each } 1 \leq i \leq j \leq n) \\ &= \prod_{1 \leq i \leq j \leq n} P(\lambda\xi_{ij} \in [a_{ij}, a_{ij} + \delta]) > 0, \end{aligned}$$

where the last inequality comes from (4.3). The proof is thus complete. \square

Having established Lemma 2 leads to the following proof of Theorem 6:

Proof of Theorem 6. Let $\delta > 0$ be the strictly positive value defined in Lemma 2. Leveraging the same argument as in (4.3) (see Proof of Theorem 5), and without loss of generality, we assume that $a_{ij} > 0$ for all $1 \leq i \leq n$ and $1 \leq j \leq p$. Therefore,

$$0 < \alpha < \beta,$$

where $\alpha := \min\{a_{ij}; 1 \leq i \leq n, 1 \leq j \leq p\}$ and $\beta := \max\{a_{ij}; 1 \leq i \leq n, 1 \leq j \leq p\} + \delta$. By choosing $\lambda > 0$ such that $\lambda[u, v] \supset [\alpha, \beta]$, we then have

$$\bigcup [a_{ij}, a_{ij} + \delta] \subset [\alpha, \beta] \subset \bigcap \text{supp}(\lambda x_{ij}), \quad (4.5)$$

where the union and the intersection are taken over $1 \leq i \leq n$ and $1 \leq j \leq p$. Let $\mathbf{X} = (x_{ij})$ be an $n \times p$ matrix. By setting $\mathbf{Y} = \lambda\mathbf{X}$, we have

$$\begin{aligned} & P(\mathbf{X}^T \mathbf{X} > 0 \text{ and } (\mathbf{X}^T \mathbf{X})^{(\alpha)} \not\geq 0) \\ &= P(\mathbf{Y}^T \mathbf{Y} > 0 \text{ and } (\mathbf{Y}^T \mathbf{Y})^{(\alpha)} \not\geq 0). \end{aligned}$$

From Lemma 2, then for $1 \leq i \leq n$ and $1 \leq j \leq p$, the above is at least following the relationship below:

$$\begin{aligned} & P(\lambda x_{ij} \in [a_{ij}, a_{ij} + \delta]) \\ &= \prod_{1 \leq i \leq n, 1 \leq j \leq p} P(\lambda x_{ij} \in [a_{ij}, a_{ij} + \delta]) > 0 \end{aligned}$$

where the last step follows from (4.5) and the independence assumption of $\{\xi_{ij}; 1 \leq i \leq j \leq n\}$, hence concluding our proof. \square

V. CONCLUSION

We have studied in this paper the impact of the Hadamard product, and especially the Hadamard powers, on positive definite deterministic and random matrices. Our main results yield the existence of an interval of non integer values for the exponent of the Hadamard powers such that for any positive definite random matrix, up to a mild assumption on the support of its entries, has non non-negative Hadamard powers. Our initial results pave the way for a better

understanding of kernel learning problems where the exponent, of either polynomial kernel or Gaussian Radial Basis Function (RBF), is considered as a tuning parameter.

REFERENCES

- [1] K. J. Horadam, *Hadamard matrices and their applications*. Princeton university press, 2012.
- [2] A. Hedayat, W. D. Wallis *et al.*, “Hadamard matrices and their applications,” *The Annals of Statistics*, vol. 6, no. 6, pp. 1184–1238, 1978.
- [3] S. Agaian, *Hadamard matrices and their applications*. Springer, 2006, vol. 1168.
- [4] D. A. Bulutoglu and D. M. Kaziska, “A counterexample to beder’s conjectures about hadamard matrices,” *Journal of statistical planning and inference*, vol. 139, no. 9, pp. 3381–3383, 2009.
- [5] K. M. Audenaert, “Spectral radius of hadamard product versus conventional product for non-negative matrices,” *Linear algebra and its applications*, vol. 432, no. 1, pp. 366–368, 2010.
- [6] H. Jiang, W.-K. Ching, W.-S. Cheung, W. Hou, and H. Yin, “Hadamard kernel svm with applications for breast cancer outcome predictions,” *BMC systems biology*, vol. 11, no. 7, p. 138, 2017.
- [7] P. Li, “Several tunable gmm kernels,” *arXiv preprint arXiv:1805.02830*, 2018.
- [8] A. Belton, D. Guillot, A. Khare, and M. Putinar, “Simultaneous kernels of matrix hadamard powers,” *Linear Algebra and its Applications*, vol. 576, pp. 142–157, 2019.
- [9] T. Hastie, R. Tibshirani, and J. Friedman, “The elements of statistical learning: data mining, inference, and prediction, springer series in statistics,” *Springer New York*, 2009.
- [10] O. Bousquet and D. Herrmann, “On the complexity of learning the kernel matrix,” *Advances in neural information processing systems*, vol. 15, pp. 415–422, 2002.
- [11] H. Fröhlich, O. Chapelle, and B. Schölkopf, “Feature selection for support vector machines using genetic algorithms,” *International journal on artificial intelligence tools*, vol. 13, no. 04, pp. 791–800, 2004.
- [12] S. S. Keerthi, V. Sindhwani, and O. Chapelle, “An efficient method for gradient-based adaptation of hyperparameters in svm models,” in *Advances in neural information processing systems*, 2007, pp. 673–680.
- [13] B. Schölkopf, A. J. Smola, F. Bach *et al.*, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.
- [14] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge University Press, 1991.
- [15] J. Schur, “Bemerkungen zur theorie der beschränkten bilinearformen mit unendlich vielen veränderlichen,” *Journal für die reine und angewandte Mathematik (Crelles Journal)*, vol. 1911, no. 140, pp. 1–28, 1911.
- [16] I. Schoenberg, “Positive definite functions on spheres,” *Duke Math. J.*, vol. 1, p. 172, 1988.
- [17] H. Vasudeva, “Positive definite matrices and absolutely monotonic functions,” *Indian J. Pure Appl. Math.*, vol. 10, no. 7, pp. 854–858, 1979.
- [18] F. Hiai, “Monotonicity for entrywise functions of matrices,” *Linear algebra and its applications*, vol. 431, no. 8, pp. 1125–1146, 2009.
- [19] R. A. Horn and C. R. Johnson, Eds., *Matrix Analysis*. USA: Cambridge University Press, 1985.