PROCEEDINGS OF THE YEREVAN STATE UNIVERSITY

Physical and Mathematical Sciences 2020, **54**(2), p. 96–100

M a t h e m a t i c s

ON RANDOM WEIGHTED SUM OF POSITIVE SEMI-DEFINITE MATRICES

T. V. GALSTYAN *, A. G. MINASYAN **

YerevaNN Research Lab, YSU, Armenia

Let $A_1,\dots,A_n$ be fixed positive semi-definite matrices, i.e. $A_i \in \mathbb{S}_p^+(\mathbf{R})$ $\forall i \in \{1,\dots,n\}$ and $u_1,\dots,u_n$ are i.i.d. with $u_i \sim \mathcal{N}(1,1)$. Then, the object of our interest is the following probability

$$\mathbb{P}\left(\sum_{i=1}^{n} u_i A_i \in \mathbb{S}_p^+(\mathbf{R})\right).$$

In this paper we examine this quantity for pairwise commutative matrices. Under some generic assumption about the matrices we prove that the weighted sum is also positive semi-definite with an overwhelming probability. This probability tends to 1 exponentially fast by the growth of number of matrices n and is a linear function with respect to the matrix dimension p.

https://doi.org/10.46991/PYSU:A/2020.54.2.096

MSC2010: 60A05; 65C50.

***Keywords*:** positive semi-definite matrices, random weighted sum, bootstrap.

Introduction. The problem arising from a quite unexpected setup of famous bootstrap technique first introduced by [1]. Bootstrap technique is a fundamental method in statistics and it is used in many estimation problems, when the data-sets are not large enough. It also has an advantage in the construction of confidence intervals, i.e. estimating the variance of the estimator, while other, more classical, methods don't. More on bootstrap technique, its variants and applications can be found in [2,3].

In the most generic setup, consider the following optimization

Problem 1. Maximize $\mathcal{L}(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(Y_i, \boldsymbol{\theta})$ subject to $\boldsymbol{\theta} \in \Theta$.

This is a typical setup for maximum likelihood estimation problem. Assuming the observations are coming from the exponential family [4, 5] and hence making the model from Generalized Linear Models (GLMs), it can be easily shown that the resulting likelihood function is concave (see, e.g., [6]), which, in turn, implies the unique solution of Problem 1. Moreover, in the context of optimization Problem 1 it can be shown that for $\forall\, i \in \{1,\dots,n\}$ the function $\ell_i(Y_i, \boldsymbol{\theta})$ is concave in $\boldsymbol{\theta}$, yielding concavity of the finite sum of concave functions $\mathcal{L}(\boldsymbol{\theta})$.

* E-mail: tigran@yerevann.com

** E-mail: minasyan@yerevann.com

The bootstrap counterpart of Problem 1 can be stated as the following

Problem 2. Maximize

$$\mathcal{L}^\flat(\boldsymbol{\theta}) = \sum_{i=1}^{n} \ell_i(Y_i, \boldsymbol{\theta}) w_i^\flat \tag{1}$$

subject to $\boldsymbol{\theta} \in \Theta$, where $w_i^\flat$ are known as bootstrap multipliers and in general need to satisfy the following properties:

$$\mathbb{E}^\flat w_i^\flat = 1, \quad \mathbb{V}ar^\flat w_i^\flat = 1, \quad \mathbb{E}^\flat \exp(w_i^\flat) < \infty. \tag{2}$$

It has been confirmed that for the proper choice of bootstrap weights $w_i^\flat$ (e.g. from the exponential distribution with parameter 1: $w_i^\flat \overset{\text{i.i.d.}}{\sim} \text{Exp}(1)$) the Problem 2 remains concave (see, e.g., [6]). In other words, the optimization methods used for solving Problem 1 yield the maximum point of (1) as well. As showed in [6] (page 13) the logarithms of probability density functions are of the following form $\ell_i(Y_i, \boldsymbol{\theta}) := Y_i \cdot \boldsymbol{\theta} - g(\boldsymbol{\theta})$. Let a matrix $H_{\boldsymbol{w}}$ be defined as follows

$$H_{\boldsymbol{w}}(\boldsymbol{\theta}) = \sum_{i=1}^{n} w_i^\flat g(\boldsymbol{\theta}), \tag{3}$$

hence,

$$\mathcal{L}^\flat(\boldsymbol{\theta}) = \left(\sum_{i=1}^{n} Y_i w_i^\flat \right) \boldsymbol{\theta} - H_{\boldsymbol{w}}(\boldsymbol{\theta}).$$

It is evident that for the choice $w_i^\flat \overset{\text{i.i.d.}}{\sim} \mathcal{N}(1,1)$ the matrix $H_{\boldsymbol{w}}$ is not necessarily positive semi-definite. However, the positive semi-definiteness of $H_{\boldsymbol{w}}$ guarantees that the algorithm will not be stuck in a local maximum. Moreover, the efficiency of numerical methods increases when $H_{\boldsymbol{w}}$ is positive semi-definite.

In this paper we study the conditions, under which the Problem 2 remains concave. The concavity of optimization Problem 2 is essential in proving the bootstrap validity. However, the problem as it was proposed in abstract is interesting by itself and is believed to be very hard for general matrices A_i and arbitrary random weights u_i satisfying (2).

Main Result. The main result concerns to the description of the conditions on non-negative definite symmetric matrices so that the probability mentioned in the abstract would be high. That means, if the weights $\{u_i\}_{i=1}^n$ are i.i.d. from $\mathcal{N}(1,1)$, then with high probability, converging to 1 exponentially with respect to n, the matrix

$$\mathbf{A} \overset{\text{def}}{=} \sum_{i=1}^{n} u_i A_i \in \mathbb{S}_p^+(\mathbf{R}),$$

i.e. it is non-negative definite.

The main theoretical finding of this paper is summarized in the following

Theorem. *Let $A_1, \dots A_n \in \mathbb{S}_p^+(\mathbf{R})$ such that for any $i, j \in [n]$ it holds $A_i A_j = A_j A_i$ and $u \sim \mathcal{N}(1,1)$. Assume $u_1, \dots, u_n$ are the i.i.d. copies of u. Then, assuming that all z_j ($j = 1, \dots, p$) defined in (5) satisfy $|z_j| \asymp \sqrt{n}$ for matrices A_i, then*

$$\mathbb{P}\left(\sum_{i=1}^{n} A_i u_i \in \mathbb{S}_p^+(\mathbf{R}) \right) \geq 1 - C \cdot \frac{p e^{-\frac{n}{2}}}{\sqrt{2\pi n}} \tag{4}$$

for an absolute universal constant C.

Proof. Let us first formulate a simple lemma without proof. We omit the proof from the paper and refer the reader to [7], Theorem 5.1, for details.

Lemma. *If matrices A and B commute, i.e. AB = BA, and both A and B are diagonalizable, then they are diagonalizable in the same basis.*

Since it is assumed that matrices $A_1,\dots,A_n$ are commuting, hence can be diagonalized in the same basis, then it is sufficient to consider the case of diagonal matrices only.

Let $A_i = \operatorname{diag}\left(\lambda_i^{(1)},\ \lambda_i^{(2)},\dots,\lambda_i^{(p)}\right)$ for all $i = 1,\dots,n$. Then we have that the condition $\sum\limits_{i=1}^{n} A_i u_i \in \mathbb{S}_p^+(\mathbf{R})$ is equivalent to

$$\bigcap_{j=1}^{p}\left\{\sum_{i=1}^{n} u_i\lambda_i^{(j)} \geq 0\right\}.$$

First, we notice that for all $j = 1,\dots,p$ it holds

$$\sum_{i=1}^{n} u_i\lambda_i^{(j)} \sim \mathcal{N}\left(\sum_{i=1}^{n}\lambda_i^{(j)},\ \sum_{i=1}^{n}\left(\lambda_i^{(j)}\right)^2\right).$$

Then,

$$\mathbb{P}\left(\sum_{i=1}^{n} u_i\lambda_i^{(j)} \geq 0\right) = \mathbb{P}\left(Z_0 \geq z_j\right),$$

where $Z_0 \sim \mathcal{N}(0,1)$ and

$$z_j \stackrel{\text{def}}{=} -\frac{\sum\limits_{i=1}^{n}\lambda_i^{(j)}}{\sqrt{\sum\limits_{i=1}^{n}\left(\lambda_i^{(j)}\right)^2}}. \tag{5}$$

It is easy to see that in general $|z_j| \in [1,\sqrt{n}]$. At this point we use the assumption that $|z_j| \asymp \sqrt{n}$. Then, using the fact that

$$F_{Z_0}(x) = \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\,dt \asymp \frac{e^{-\frac{x^2}{2}}}{x\sqrt{2\pi}},$$

we get

$$\mathbb{P}\left(\sum_{i=1}^{n} u_i\lambda_i^{(j)} \geq 0\right) \geq 1 - C\frac{e^{-\frac{n}{2}}}{\sqrt{2\pi n}}$$

for some universal constant C.

Hence, we obtain the final lower bound, by using the union bound for all $j = 1,\dots,p$

$$\mathbb{P}\left(\sum_{i=1}^{n} A_i u_i \in \mathbb{S}_p^+(\mathbf{R})\right) = \mathbb{P}\left(\bigcap_{j=1}^{p}\left\{\sum_{i=1}^{n} u_i\lambda_i^{(j)} \geq 0\right\}\right) \geq 1 - C\frac{pe^{-\frac{n}{2}}}{\sqrt{2\pi n}},$$

as desired. □

By the proposition below we illustrate that the assumptions made in Theorem are indeed realistic and we provide an example of such class of matrices.

Proposition. *Let matrices $A_1,\dots,A_n$ be such that that for each $j=1,\dots,p$ the eigenvalues $\lambda^j(A_1),\dots,\lambda^j(A_n)$ are assumed to be i.i.d. from $\mathcal{U}[0,1]$. Then, for such matrices $A_1,\dots,A_n$ the quantity z defined in (5) is indeed of order $\sqrt{n}$. Moreover, the simulation studies show, that if the eigenvalues are sampled from $\mathcal{U}[0,1]$, then the proportionality constant is approximately* 0.8.

Remark 1. *It is worth noticing that the condition $|z_j| \asymp \sqrt{n}$ means that the corresponding eigenvalues of matrices $A_1,\dots,A_n$ of the same order are not separated much.*

Notice that in Theorem the matrices satisfy certain conditions, however we illustrate an example when the probability presented in (4) is constant.

Remark 2. *It is evident that the assumption $|z| \asymp \sqrt{n}$ will not always be fulfilled, however as noticed in Proposition there exist classes of matrices, for which this assumption holds. Below we show a particular choice of matrices A such that the probability in (4) is not converging to* 1. *Let*

$$A_1 = \operatorname{diag}(1,0,\dots,0) \quad \textit{and} \quad A_2 = \dots = A_n = \mathbf{0}_{p\times p},$$

then the probability from (4) boils down to $\mathbb{P}(\mathcal{N}(1,1) \geq 0) \approx 0.85$, which is a constant probability.

Received 27.02.2020
Reviewed 20.05.2020
Accepted 17.08.2020

REFERENCES

1. Efron B. Bootstrap Methods: Another Look at the Jackknife. *Ann. Statist.* **7** (1979), 1–26. https://doi.org/10.1214/aos/1176344552
2. Chernick M.R., González-Manteiga W., Crujeiras R.M., Barrios E.B. Bootstrap Methods. In: Lovric M. (ed.) *International Encyclopedia of Statistical Science.* Springer, Berlin, Heidelberg (2011), 169–174. https://doi.org/10.1007/978-3-642-04898-2_150
3. González-Manteiga W., Prada Sánchez J.M., Romo J. *The Bootstrap – A Review.* Working Paper. Universidad Carlos III de Madrid. Departamento de Economía (1992). Available at: http://hdl.handle.net/10016/2833 (Accessed: 25 February 2020).
4. Ibragimov I.A., Has'minskii R.S. *Statistical estimation. Asymptotic theory*. Springer, New-York (1981). https://doi.org/10.1007/978-1-4899-0027-2
5. Minasyan A. Alternating Least Squares in Generalized Linear Models. *J. Contemp. Mathemat. Anal.*, **54** (2019), 302–312. https://doi.org/10.3103/S1068362319050078
6. Minasyan A. 'Bootstrap Confidence Sets for Generalized Linear Models'. Master's Thesis, Skolkovo Institute Science and Technology (Skoltech), Moscow (2018).
7. Conrad K.T. *The Minimal Polynomial and Some Applications.* Tech. report. Dept. of Mathematics. Univ. of Connecticut. (2008).

Տ. Վ. ԳԱԼՍՏՅԱՆ, Ա. Գ. ՄԻՆԱՍՅԱՆ

ԴՐԱԿԱՆ ԿԻՍԱՎՈՐՈՇՅԱԼ ՄԱՏՐԻՑՆԵՐԻ ՊԱՏԱՀԱԿԱՆ ԿՇԻՌՆԵՐՈՎ ԳՈՒՄԱՐԻ ՎԵՐԱԲԵՐՅԱԼ

Դիցուք $A_1,\dots,A_n$-ը ֆիքսված դրական կիսաորոշյալ մատրիցներ են, այսինքն՝ $A_i \in \mathbb{S}_p^+(\mathbf{R})\ \forall i \in \{1,\dots,n\}$ և $u_1,\dots,u_n$-ը անկախ պատահական մեծություններ են և ունեն միևնույն բաշխումը՝ $u_i \sim \mathcal{N}(1,1)$: Մեր հետաքրքրությունների առարկան է հետևյալ հավանականությունը.

$$\mathbb{P}\left(\sum_{i=1}^{n} u_i A_i \in \mathbb{S}_p^+(\mathbf{R})\right):$$

Այս հոդվածում մենք հետազոտում ենք վերոնշյալ հավանականությունը զույգ-առ-զույգ փեղափոխելի մատրիցների համար: Մատրիցների վերաբերյալ մի քանի ընդհանուր ենթադրությունների տեղի ունենալու դեպքում մենք ապացուցում ենք, որ վերոնշյալ կշիռներով գումարը շատ մեծ հավանականությամբ նույնպես դրական կիսաորոշյալ է: Այս հավանականությունը մատրիցների n քանակից կախված էքսպոնենցյալ արագությամբ ձգտում է 1-ի և կախված չէ մատրիցների p չափողականությունից:

Т. В. ГАЛСТЯН, А. Г. МИНАСЯН

О СЛУЧАЙНОЙ ВЗВЕШЕННОЙ СУММЕ ПОЛОЖИТЕЛЬНО ПОЛУОПРЕДЕЛЕННЫХ МАТРИЦ

Пусть $A_1,\dots,A_n$ – фиксированные положительно полуопределенные матрицы, т.е. $A_i \in \mathbb{S}_p^+(\mathbf{R})\ \forall i \in \{1,\dots,n\}$, и $u_1,\dots,u_n$ – независимые одинаково определенные случайные величины, т.е. $u_i \sim \mathcal{N}(1,1)$. Нас будет интересовать следующая вероятность:

$$\mathbb{P}\left(\sum_{i=1}^{n} u_i A_i \in \mathbb{S}_p^+(\mathbf{R})\right).$$

В данной статье мы исследуем вышеупомянутую вероятность для попарно коммутирующих матриц. При достаточно общих условиях мы доказали, что взвешенная сумма данных матриц с очень большой вероятностью тоже будет положительно полуопределенной. Эта вероятность экспоненциально стремится к 1 в зависимости от количества матриц n и не зависит от размерности матриц p.