

CONCENTRATION INEQUALITIES FOR THE SPECTRAL MEASURE OF RANDOM MATRICES

BERNARD DELYON

IRMAR, Université Rennes 1, Campus de Beaulieu, F-35042 Rennes, France

email: bernard.delyon@univ-rennes1.fr

Submitted September 3, 2010, accepted in final form October 24, 2010

AMS 2000 Subject classification: Primary 15A52; Secondary 60F10.

Keywords: Spectral measure, random matrices

Abstract

We give new exponential inequalities for the spectral measure of random Wishart matrices. These results give in particular useful bounds when these matrices have the form $M = YY^T$, in the case where Y is a $p \times n$ random matrix with independent entries (weaker conditions are also proposed), and p and n are large.

1 Introduction

Recent literature has focused much interest on Wishart random matrices, that is matrices of the form:

$$M = \frac{1}{n} YY^T$$

where $Y \in \mathbb{R}^{p \times n}$ is a rectangular $p \times n$ matrix with random centered entries, and both n and $p \leq n$ tend to infinity: typically $p = p(n)$, and $p(n)/n$ tends to some limit.

M can be seen as the empirical covariance matrix of a random vector of dimension p sampled n times, each sample being a column of Y . It is common in applications to have a number of variables with a comparable order of magnitude with the number of observation samples; the analysis of the asymptotic spectral distribution for such matrices appears now to be an important issue, in particular for implementing significance test for eigenvalues (e.g. for principal component analysis with p variables and n observations).

Our objective is to give new exponential inequalities concerning symmetric functions of the eigenvalues, i.e. the variables of the form

$$Z = g(\lambda_1, \dots, \lambda_p)$$

where $(\lambda_i)_{1 \leq i \leq p}$ is the set of eigenvalues of M . These inequalities will be upper bounds on

$$E \left[e^{Z - E[Z]} \right]$$

and lead naturally to concentration bounds, i.e. bounds on $P(|Z - E[Z]| \geq x)$ for large values of x .

Our contribution is to improve on the existent literature on the following points: (i) the function g is a once or twice differentiable general function (i.e. not necessarily of the form (2) below), (ii) the entries of Y are possibly dependent, (iii) they are not necessarily identically distributed and (iv) the bound is instrumental when both p and n are large.

The columnwise dependence is usually taken into account in essentially two different ways: either one assumes that the columns of Y are i.i.d. but each column is a dependent vector, or that M can be written, for some mixture matrix $Q \in \mathbb{R}^{p \times p}$

$$M = \frac{1}{n} Q X X^T Q^T \quad (1)$$

where $X \in \mathbb{R}^{p \times n}$, $p < n$ with independent entries; this means that the j -th observation vector Y_j has the specific form $Q X_{\cdot j}$. This second way is thus more restrictive; but it is natural from a theoretical point of view since one obtains essentially the same results as in the case $Q = Id$ (cf. [1] Chapter 8). We shall consider both cases (Theorem 3 and Theorem 2), the first one is much easier, but leads to much larger bounds when p is large.

We shall give now a brief evaluation of results in this context. From now on in this section we assume for simplicity that $Q = Id$. In order to be more specific, and to be able to compare with existing results, we shall stick in this introduction to the case where

$$g(\lambda_1, \dots, \lambda_p) = \sum_{k=1}^p \varphi(\lambda_k) \quad (2)$$

although any V-statistics of $(\lambda_1, \dots, \lambda_p)$ can be considered. According to distributional limit-theorems (e.g. Theorem 9.10 p.247 of [1]), the asymptotic variance of Z as n tends to infinity and p/n converges to some limit $y > 0$ is a rather complicated function of y ; this suggests that the best one could hope in the independent case, and when p and n tend to infinity, would be something like

$$P(|Z - E[Z]| \geq x) \leq \exp(-f(p/n)x^2) \quad (3)$$

for some function f .

Guionnet and Zeitouni obtained in [3] for the independent case (Corollary 1.8¹ (a), with $Y = X$, $N = p$, $M = n$)

$$P(|Z - E[Z]| \geq x) \leq 4 \exp(-x^2/g^2) \quad (4)$$

for some reasonable constant g with order of magnitude $g \sim \xi^2 \|x\varphi'(x^2)\|_\infty$ where

$$\xi = \sup_{ij} \|X_{ij}\|_\infty$$

and under the assumption of convexity of the function $x \mapsto \varphi(x^2)$. A similar result is also given without this convexity assumption (Corollary 1.8 (b)), but a logarithmic Sobolev inequality is required for the densities of the X_{ij} s.

¹There is a misprint: Z has to be replaced by $Z/(N + M)$ in the statement of the Corollary.

We shall obtain here, in the independent case, the following particular consequence of Theorem 2 below (still with $Q = Id$):

$$E [\exp(Z - E[Z])] \leq \exp \left(\frac{64p}{n} \xi^4 \left(\|\varphi'\|_\infty + \frac{p}{n} \xi^2 \|\varphi''\|_\infty \right)^2 \right). \tag{5}$$

This leads to

$$P(|Z - E[Z]| \geq x) \leq 2 \exp \left(-\frac{n x^2}{p s^2} \right), \quad s = 16\xi^2 \left(\|\varphi'\|_\infty + \frac{p}{n} \xi^2 \|\varphi''\|_\infty \right). \tag{6}$$

In the case where n tends to infinity and $p = cn$, (4) gives, when it applies, a better bound than (6) because it does not involve second derivatives of φ ; when $p \ll n$ Equation (6) is always sharper. The amazing point is that this inequality is essentially a consequence of an extension by Boucheron, Lugosi and Massart of the McDiarmid inequality, Theorem 1 below.

Equation (5) will be applied in a forthcoming article with Jian-Feng Yao and Jia-Qi Chen where φ is an estimate of the limit spectral density of $E[M]$; it will be a key step to prove the validity of a cross-validation procedure.

In the case of independent columns only, recently Guntuboyina and Leeb obtained the following concentration inequality (Theorem 1 (ii) of [4]):

$$P(|Z - E[Z]| \geq x) \leq 2 \exp \left(-\frac{2x^2}{nV^2} \right) \tag{7}$$

where V is the total variation of the function φ on \mathbb{R} and it is assumed that $\|Y_{ij}\|_\infty \leq 1$ for all i, j . This bound is never close to (3) but is essentially optimal in the context of independent columns if p and n have the same order of magnitude (Example 3 in [4]). We provide here an estimate which can be written as

$$E [\exp(Z - E[Z])] \leq \exp \left(\frac{16p^2 \xi^4 \|\varphi'\|_\infty^2}{n} \right). \tag{8}$$

This is a simple consequence of Theorem 3 below with $a^2 = p\xi^2$ (since the entries are independent). In the independent case, this improves on Corollary 1.8 of [3] only if p^2 is significantly smaller than n ; it also improves on (7) if $p \ll n$ since in this case $p^2/n \ll n$. But it is hardly useful for applications where p and n have the same order of magnitude.

2 Results

The following Theorem is a bounded difference inequality due to Boucheron-Lugosi-Massart [2] (Theorem 2 with $\lambda = 1, \theta \rightarrow 0$); it looks like the McDiarmid inequality but the important difference here is that the infinity norm in (9) is outside the sum:

Theorem 1. *Let $Y = (Y_1, \dots, Y_n)$ be a zero-mean sequence of independent variables with values in some measured space E . Let f be a measurable function on E^n with real values. Let Y' be an independent copy of Y and set*

$$Y^k = (Y_1, \dots, Y_{k-1}, Y'_k, Y_{k+1}, \dots, Y_n).$$

Then, with the notation $x_+^2 = x^2 \mathbf{1}_{x>0}$,

$$E[\exp(f(Y) - E[f(Y)])] \leq \exp\left(\|E[\sum_k (f(Y) - f(Y^k))_+^2 | Y]\|_\infty\right).$$

We shall use actually the weaker inequality:

$$E[\exp(f(Y) - E[f(Y)])] \leq \exp\left(\|\sum_k (f(Y) - f(Y^k))_+^2\|_\infty\right). \quad (9)$$

Notations. In what follows $\|x\|$ will stand for the Euclidean norm of the vector x and $\|M\|$ will stand for the matrix norm of $M \in \mathbb{R}^{d \times d}$:

$$\|M\| = \sup_{\|x\|=1} \|Mx\|.$$

$\|X\|_p$ is the usual L_p -norm of the real random variable X . e_i will stand for the i -th vector of the canonical basis, and we set

$$M_{.j} = M e_j = (M_{ij})_{1 \leq i \leq d} \in \mathbb{R}^d$$

And similarly, M_i is the i -th row of M , a row vector, a $1 \times d$ matrix. If $u, v \in \mathbb{R}^d$, they will be understood also as $d \times 1$ matrix, in particular $u^T v = \langle u, v \rangle$, and uv^T is a $d \times d$ matrix.

Theorem 2. Let Q be a $p \times p$ deterministic matrix, $X = (X_{ij})$, $1 \leq i \leq p$, $1 \leq j \leq n$ be a matrix of random independent entries, and set

$$M = \frac{1}{n} Q X X^T Q^T.$$

Let $\lambda \mapsto g(\lambda)$ be a twice differentiable symmetric function on \mathbb{R}^p and define the random variable

$$Z = g(\lambda_1, \dots, \lambda_p)$$

where $(\lambda_1, \dots, \lambda_p)$ is the vector of the eigenvalues of M . Then

$$E[e^{Z - E[Z]}] \leq \exp\left(\frac{64p}{n} \xi^4 \left(\gamma_1 + \frac{p}{n} \xi^2 \gamma_2\right)^2\right). \quad (10)$$

where

$$\xi = \|Q\| \sup_{ij} \|X_{ij}\|_\infty \quad (11)$$

$$\gamma_1 = \sup_{k, \lambda} \left| \frac{\partial g}{\partial \lambda_k}(\lambda) \right| \quad (12)$$

$$\gamma_2 = \sup_{\lambda} \|\nabla^2 g(\lambda)\| \quad (\text{matrix norm}). \quad (13)$$

In particular, for any $x > 0$

$$P(|Z - E[Z]| > x) \leq 2 \exp\left(-\frac{n}{256p} x^2 \xi^{-4} \left(\gamma_1 + \frac{p}{n} \xi^2 \gamma_2\right)^{-2}\right). \quad (14)$$

The next theorem deals with matrices X with independent columns; since QX has again independent columns, we assume here that $Q = Id$.

Theorem 3. *Let $X = (X_{ij})$, $1 \leq i \leq p$, $1 \leq j \leq n$ be a matrix of random independent columns, and set*

$$M = \frac{1}{n}XX^T$$

Let $\lambda \mapsto g(\lambda)$ be a differentiable symmetric function on \mathbb{R}^p and define the random variable

$$Z = g(\lambda_1, \dots, \lambda_p) \tag{15}$$

where $(\lambda_1, \dots, \lambda_p)$ is the vector of the eigenvalues of M . Then

$$E \left[e^{Z-E[Z]} \right] \leq \exp \left(\frac{16a^4\gamma_1^2}{n} \right) \tag{16}$$

where

$$\begin{aligned} \gamma_1 &= \sup_{\lambda} \left| \frac{\partial g}{\partial \lambda_1}(\lambda) \right| \\ a &= \sup_j \left\| \sum_i X_{ij}^2 \right\|_{\infty}^{1/2}. \end{aligned}$$

In particular, for any $x > 0$

$$P(|Z - E[Z]| > x) \leq 2 \exp \left(-\frac{nx^2}{64a^4\gamma_1^2} \right). \tag{17}$$

3 Proof of theorem 2

Define the function on $\mathbb{R}^{p \times n}$

$$f(X) = g(\lambda_1, \dots, \lambda_p).$$

We have to estimate

$$\sum_{ij} (f(X) - f(X^{ij}))_+^2$$

with the notations of Theorem 1. If we set

$$\alpha = \sup_{ij} \|X_{ij}\|_{\infty},$$

the Taylor formula implies

$$|f(X) - f(X^{ij})| \leq 2\alpha \left\| \frac{\partial f}{\partial X_{ij}}(X) \right\| + 2\alpha^2 \sup_{k,l,Y} \left\| \frac{\partial^2 f}{\partial X_{kl}^2}(Y) \right\|_{\infty} = A_{ij} + B.$$

We shall estimate these quantities with the help of Lemma 4 below, specifically Equations (24) and (25), which give bounds for the derivatives of a function of the eigenvalues; the variable t is here X_{ij} . We have for any matrix P

$$\text{Tr} \left(P \frac{\partial M}{\partial X_{ij}} \right) = \frac{1}{n} \frac{\partial}{\partial X_{ij}} \text{Tr}(X^T Q^T P Q X) = \frac{1}{n} (Q^T (P + P^T) Q X)_{ij}. \quad (18)$$

Equation (24) can be rewritten as

$$\frac{\partial f(X)}{\partial X_{ij}} = \frac{2}{n} \sum_{k=1}^p d_k^{-1} \frac{\partial g(\lambda)}{\partial \lambda_k} (Q^T P_{\lambda_k} Q X)_{ij}$$

where P_{λ_k} is the orthogonal projection on the eigenspace of M corresponding to λ_k , and d_k is the dimension of this space. It happens that we can get a good bound for a partial Euclidean norm of the gradient of f : set (j is fixed)

$$\begin{aligned} g_k &= \frac{\partial g(\lambda)}{\partial \lambda_k} \\ \xi_k &= P_{\lambda_k} Q X_{\cdot j} = (P_{\lambda_k} Q X)_{\cdot j} \end{aligned}$$

where $A_{\cdot j}$ stands for the j th column of the matrix A . Notice that in case of eigenvalues of higher multiplicities, some ξ_k s are repeated, but if $\xi_j = \xi_k$, $\lambda_j = \lambda_k$, one has also $g_j = g_k$ (by symmetry of g); hence, denoting by K a set of indices such that $\sum_{k \in K} P_{\lambda_k} = Id$,

$$\begin{aligned} \frac{n^2}{4} \sum_i \left(\frac{\partial f(X)}{\partial X_{ij}} \right)^2 &= \sum_i \left(\sum_{k=1}^p g_k d_k^{-1} (Q^T \xi_k)_i \right)^2 \\ &= \sum_i \left(\sum_{k \in K} g_k (Q^T \xi_k)_i \right)^2 \\ &= \left\| Q^T \left(\sum_{k \in K} g_k \xi_k \right) \right\|^2 \\ &\leq \|Q\|^2 \left\| \sum_{k \in K} g_k \xi_k \right\|^2 \\ &= \|Q\|^2 \sum_{k \in K} g_k^2 \|\xi_k\|^2 \quad (\text{orthogonality}) \\ &\leq \|Q\|^2 \gamma_1^2 \sum_{k \in K} \|\xi_k\|^2 \\ &= \gamma_1^2 \|Q\|^2 \sum_{k \in K} \|\xi_k\|^2 \quad (\text{orthogonality}) \\ &= \gamma_1^2 \|Q\|^2 \|Q X_{\cdot j}\|^2 \\ &\leq \gamma_1^2 \|Q\|^4 \|X_{\cdot j}\|^2 \\ &\leq \gamma_1^2 \|Q\|^4 \alpha^2 p. \end{aligned}$$

For the estimation of the second derivatives we shall use again Lemma 4. Equation (25) can be rewritten as

$$\left| \frac{\partial^2 f}{\partial X_{ij}^2} \right| \leq \gamma_2 \text{Tr}(\dot{M}^2) + \gamma_1 \sum_k |v_k|.$$

where v_k is the k -th eigenvalue of $\frac{\partial^2 M}{\partial X_{ij}^2}$. Using now Equation (18) with $P = \dot{M} = P^T$

$$\begin{aligned} n \text{Tr}(\dot{M}^2) &= 2(Q^T \dot{M} Q X)_{ij} \\ &= 2 \langle Q e_i, \dot{M} Q X e_j \rangle \\ &\leq 2 \|Q\|^2 \|X e_j\| \|\dot{M}\| \\ &\leq 2 \|Q\|^2 (p \alpha^2)^{1/2} \text{Tr}(\dot{M}^2)^{1/2} \end{aligned}$$

hence

$$\text{Tr}(\dot{M}^2) \leq 4 p n^{-2} \|Q\|^4 \alpha^2$$

and since

$$\frac{\partial^2 M}{\partial X_{ij}^2} = \frac{2}{n} Q_{\cdot i} Q_{\cdot i}^T$$

the v_k are all zero except for one $2n^{-1} \|Q_{\cdot i}\|^2 \leq 2n^{-1} \|Q\|^2$ and finally we get

$$\left| \frac{\partial^2 f}{\partial X_{ij}^2} \right| \leq \frac{4p \|Q\|^4 \alpha^2}{n^2} \gamma_2 + \frac{2 \|Q\|^2}{n} \gamma_1.$$

Hence, for any j

$$\begin{aligned} \sum_i (f(X) - f(X^{ij}))^2 &\leq \sum_i A_{ij}^2 + 2B \sum_i A_{ij} + pB^2 \\ &\leq \sum_i A_{ij}^2 + 2B \sqrt{p} (\sum_i A_{ij}^2)^{1/2} + pB^2 \\ &= ((\sum_i A_{ij}^2)^{1/2} + B \sqrt{p})^2 \\ &\leq \left(\frac{4\alpha^2 \|Q\|^2 \sqrt{p}}{n} \gamma_1 + \frac{2\alpha^2}{n} \left(\frac{4p \|Q\|^4 \alpha^2 \gamma_2}{n} + 2 \|Q\|^2 \gamma_1 \right) \sqrt{p} \right)^2 \\ &= \left(\frac{8\alpha^2 \|Q\|^2 \sqrt{p}}{n} \gamma_1 + \frac{8p\alpha^4 \|Q\|^4 \gamma_2}{n^2} \sqrt{p} \right)^2 \\ &= \frac{64p\alpha^4 \|Q\|^4}{n^2} \left(\gamma_1 + \frac{p}{n} \alpha^2 \|Q\|^2 \gamma_2 \right)^2 \end{aligned}$$

and Equation (10) follows.

Equation (14) is easily deduced from (10): let us recall that a random variable X such that for any $t > 0$, and some $a > 0$, $E[e^{tX}] \leq e^{at^2}$, satisfies for any x , $P(X > x) \leq e^{-x^2/4a}$. In particular,

application of (10) with the function $tg(x)$ instead of $g(x)$ leads to the required bound on the Laplace transform and

$$\begin{aligned} P(|Z - E[Z]| > x) &\leq P(Z - E[Z] > x) + P(-Z - E[-Z] > x) \\ &\leq 2 \exp\left(-x^2 \frac{n}{256p} \xi^{-4} \left(\gamma_1 + \frac{p}{n} \xi^2 \gamma_2\right)^{-2}\right). \end{aligned} \quad (19)$$

4 Proof of theorem 3

Define the function on $\mathbb{R}^{p \times n}$

$$f(X) = g(\lambda) = g(\lambda_1, \dots, \lambda_p). \quad (20)$$

We have to estimate

$$\sum_j (f(X) - f(X^j))_+^2$$

with the notations of Theorem 1. Then the Taylor formula implies

$$|f(X) - f(X^j)| \leq \sup_{0 \leq t \leq 1} \left| \frac{d}{dt} f(X + t\Delta) \right|, \quad \Delta = X^j - X$$

The matrix $\Delta = X^j - X$ vanishes except on its j -th column which is the vector δ , $\delta_i = X_{ij}^j - X_{ij}$. We shall estimate these quantities with the help of Lemma 4 of the appendix; with $M(t) = \frac{1}{n}(X + t\Delta)(X + t\Delta)^T$. We have

$$\dot{M} = \frac{1}{n}(X\Delta^T + \Delta X^T + 2t\Delta\Delta^T) = \frac{1}{n}(\delta X_{\cdot j}^T + X_{\cdot j} \delta^T + 2t\delta\delta^T)$$

where $X_{\cdot j}$ is the j th column of X . We set $\dot{f} = \frac{d}{dt} f(X + t\Delta)$. Equation (24) can be rewritten as

$$\dot{f} = \frac{2}{n} \sum_{k=1}^p \frac{\partial g(\lambda)}{\partial \lambda_k} d_k^{-1} \langle P_{\lambda_k} \delta, X_{\cdot j} + t\delta \rangle$$

where $\lambda = \lambda(t)$ is the set of eigenvalues of $M(t)$; hence

$$\begin{aligned} |\dot{f}| &\leq \frac{2}{n} \gamma_1 \sum_{\lambda} \|P_{\lambda} \delta\| \|P_{\lambda}(X_{\cdot j} + t\delta)\| \\ &\leq \frac{2}{n} \gamma_1 \left(\sum_{\lambda} \|P_{\lambda} \delta\|^2 \right)^{1/2} \left(\sum_{\lambda} \|P_{\lambda}(X_{\cdot j} + t\delta)\|^2 \right)^{1/2} \\ &= \frac{2}{n} \gamma_1 \|\delta\| \|X_{\cdot j} + t\delta\| \\ &= \frac{4}{n} \gamma_1 a^2 \end{aligned}$$

hence, for any j

$$|f(X) - f(X^j)|^2 \leq \frac{16}{n^2} \gamma_1^2 a^4$$

and the result follows.

Equation (17) is easily deduced from (16) as before (cf. the arguments before Equation (19)).

A Smoothness of eigenvalues of a parameter dependent symmetric matrix

The following lemma has been a key instrument for proving Theorems 2 and 3:

Lemma 4. *Let $t \mapsto M(t)$ be a function of class C^{3p} from some open interval I with values in the set of real $p \times p$ symmetric matrices; there exists a twice differentiable parametrization of the eigenvalues $t \mapsto \Lambda(t) = (\lambda_1(t), \dots, \lambda_p(t))$, and the following equations are satisfied for $1 \leq k \leq p$ and $t \in I$:*

$$\sum_{j:\lambda_j=\lambda_k} \dot{\lambda}_j = \text{Tr}(\dot{M}P_{\lambda_k}) \quad (21)$$

$$\sum_{j:\lambda_j=\lambda_k} \ddot{\lambda}_j = \text{Tr}(\ddot{M}P_{\lambda_k}) + 2 \sum_{j:\lambda_j \neq \lambda_k} \frac{d_j^{-1}}{\lambda_k - \lambda_j} \text{Tr}(\dot{M}P_{\lambda_j} \dot{M}P_{\lambda_k}) \quad (22)$$

$$\sum_{j:\lambda_j=\lambda_k} \dot{\lambda}_j^2 = \text{Tr}(\dot{M}P_{\lambda_k} \dot{M}P_{\lambda_k}) \quad (23)$$

where both sides are functions of the implicit variable t , P_λ is the orthogonal projection on the eigenspace E_λ of the eigenvalue λ , d_k is the dimension of E_{λ_k} .

Let g be a twice differentiable symmetric function defined on some cube $Q = (a, b)^p$ containing the eigenvalues of the matrices $M(t)$, $t \in I$; then the function

$$\varphi(t) = g(\Lambda(t))$$

is twice differentiable on I and for any $t \in I$

$$\dot{\varphi} = \sum_k d_k^{-1} \frac{\partial g}{\partial \lambda_k}(\Lambda) \text{Tr}(\dot{M}P_{\lambda_k}) \quad (24)$$

$$|\dot{\varphi}(t)| \leq \gamma \text{Tr}(\dot{M}^2(t)) + \sup_k \left| \frac{\partial g}{\partial \lambda_k}(\Lambda(t)) \right| \sum_i |v_i(t)| \quad (25)$$

where $\gamma = \sup_{\Lambda \in Q} \|\nabla^2 g(\Lambda)\|$ (matrix norm), and $v_i(t)$ is the i -th eigenvalue of $\ddot{M}(t)$.

Proof. The smoothness comes from the possibility of finding a twice differentiable parametrization of the eigenvalues $t \mapsto (\lambda_1(t), \dots, \lambda_p(t))$ which is simply a consequence of the theorem of [5]; let us recall the third statement of that theorem:

Consider a continuous curve of polynomials

$$P(t)(x) = x^p + a_1(t)x^{p-1} + \dots + a_p(t), t \in \mathbb{R}.$$

If all a_i are of class C^{3p} then there is a twice differentiable parametrization $x = (x_1, \dots, x_p) : \mathbb{R} \rightarrow \mathbb{R}$ of the roots of P

It remains to prove Equations (21) to (25). We shall use below the notation \sum_λ for denoting a sum over distinct eigenvalues, hence

$$\sum_\lambda P_\lambda = Id$$

and for any function f :

$$\sum_{\lambda} f(\lambda) = \sum_k d_k^{-1} f(\lambda_k).$$

For any polynomial $x \mapsto f(x)$ one has

$$\frac{d}{dt} \text{Tr}(f(M)) = \text{Tr}(\dot{M}f'(M)).$$

This comes simply from the fact that this is true for $f(x) = x^n$, $n > 0$, since

$$\frac{d}{dt} M^n = \sum_{k=1}^n M^{k-1} \dot{M} M^{n-k}. \quad (26)$$

Hence for any polynomial f , one has

$$\frac{d}{dt} \text{Tr}(f(M)) = \text{Tr}(\dot{M}f'(M)) = \sum_{\lambda} \text{Tr}(\dot{M}P_{\lambda})f'(\lambda) \quad (27)$$

and on the other hand

$$\frac{d}{dt} \text{Tr}(f(M)) = \frac{d}{dt} \sum_k f(\lambda_k) = \sum_k \dot{\lambda}_k f'(\lambda_k). \quad (28)$$

By choosing a polynomial f such that $f'(\mu)$ is zero for any eigenvalue μ different from a specific eigenvalue λ_k identification of (27) and (28) gives

$$\sum_{j:\lambda_j=\lambda_k} \dot{\lambda}_j = \text{Tr}(\dot{M}P_{\lambda_k}).$$

This is (21). For proving Equation (22), we shall first prove the following formula valid for any polynomial $x \mapsto f(x)$:

$$\frac{d}{dt} f(M) = \sum_{\lambda} P_{\lambda} \dot{M} P_{\lambda} f'(\lambda) + \sum_{\lambda \neq \mu} \frac{f(\mu) - f(\lambda)}{\mu - \lambda} P_{\mu} \dot{M} P_{\lambda} \quad (29)$$

(we recall that in these sums, eigenvalues of higher multiplicities of $M = M(t)$ are counted only once). If indeed $f(x) = x^n$, Equation (26) implies

$$\frac{d}{dt} f(M) = \sum_{k=1}^n \sum_{\lambda} \sum_{\mu} \lambda^{k-1} \mu^{n-k} P_{\lambda} \dot{M} P_{\mu}.$$

Now (29) is just a consequence of the following formula for $\lambda \neq \mu$

$$\sum_{k=1}^n \lambda^{k-1} \mu^{n-k} = \frac{\lambda^n - \mu^n}{\lambda - \mu}.$$

Similarly we have on the one hand

$$\frac{d^2}{dt^2} \text{Tr}(f(M)) = \frac{d^2}{dt^2} \sum_k f(\lambda_k) = \sum_k \ddot{\lambda}_k f'(\lambda_k) + (\dot{\lambda}_k)^2 f''(\lambda_k) \quad (30)$$

and on the other hand

$$\frac{d^2}{dt^2} \text{Tr}(f(M)) = \frac{d}{dt} \text{Tr}(\dot{M}f'(M)) = \text{Tr}(\ddot{M}f'(M)) + \text{Tr}(\dot{M} \frac{d}{dt} f'(M)).$$

The first term is

$$\text{Tr}(\ddot{M}f'(M)) = \sum_{\lambda} \text{Tr}(\ddot{M}P_{\lambda})f'(\lambda) \tag{31}$$

and the second term may be computed via the Equation (29)

$$\text{Tr}(\dot{M} \frac{d}{dt} f'(M)) = \sum_{\lambda} \text{Tr}(\dot{M}P_{\lambda}\dot{M}P_{\lambda})f''(\lambda) + \sum_{\lambda \neq \mu} \frac{f'(\mu) - f'(\lambda)}{\mu - \lambda} \text{Tr}(\dot{M}P_{\mu}\dot{M}P_{\lambda}) \tag{32}$$

identifying the r.h.s. of (30) as the sum of the r.h.s. of (31) and (32) with a polynomial f such that $f''(\lambda)$ is zero for all eigenvalue λ of $M(t)$ and $f'(\lambda)$ is zero for any eigenvalue λ with the exception of a specific one λ_k for which $f'(\lambda_k) = 1$, we get $f'(M) = P_{\lambda_k}$ and

$$\sum_{j:\lambda_j=\lambda_k} \ddot{\lambda}_j = \text{Tr}(\ddot{M}P_{\lambda_k}) + 2 \sum_{\mu} \frac{1_{\mu \neq \lambda_k}}{\lambda_k - \mu} \text{Tr}(\dot{M}P_{\mu}\dot{M}P_{\lambda_k})$$

and (22) is proved. Equation (23) is proved similarly by considering a polynomial f such that $f'(\lambda)$ is zero for all eigenvalue λ of $M(t)$ and $f''(\lambda)$ is zero for any eigenvalue λ with the exception of a specific one λ_k for which $f''(\lambda_k) = 1$.

We now prove Equation (24). For any smooth symmetric function $f(x, y)$ of two real variables, we have, if we denote by f_1 and f_2 the partial derivatives:

$$f_1(x, y) = f_2(y, x) \tag{33}$$

$$f_{11}(x, y) = f_{22}(y, x) \tag{34}$$

$$f_{12}(x, y) = f_{12}(y, x) \tag{35}$$

which implies that on any point such that $x = y$, one has $f_1 = f_2$, and $f_{11} = f_{22}$. Hence using (21), the symmetry of g , and (33)

$$\begin{aligned} \dot{\varphi} &= \sum_k \frac{\partial g}{\partial \lambda_k}(\Lambda) \dot{\lambda}_k \\ &= \sum_k \frac{\partial g}{\partial \lambda_k}(\Lambda) d_k^{-1} \sum_{j:\lambda_j=\lambda_k} \dot{\lambda}_j \\ &= \sum_k d_k^{-1} \frac{\partial g}{\partial \lambda_k}(\Lambda) \text{Tr}(\dot{M}P_{\lambda_k}). \end{aligned}$$

It remains to prove (25). Setting $g_k = \frac{\partial g}{\partial \lambda_k}$ and $g_{kl} = \frac{\partial^2 g}{\partial \lambda_k \partial \lambda_l}$:

$$\begin{aligned} \ddot{\varphi}(t) &= \sum_k \ddot{\lambda}_k g_k(\Lambda) + \sum_{j,k} \dot{\lambda}_j \dot{\lambda}_k g_{jk}(\Lambda) \\ &= \sum_k d_k^{-1} \text{Tr}(\ddot{M}P_{\lambda_k}) g_k(\Lambda) + 2 \sum_{j,k:\lambda_j \neq \lambda_k} d_k^{-1} d_j^{-1} \frac{g_k(\Lambda)}{\lambda_k - \lambda_j} \text{Tr}(\dot{M}P_{\lambda_j}\dot{M}P_{\lambda_k}) + \sum_{j,k} \dot{\lambda}_j \dot{\lambda}_k g_{jk}(\Lambda) \\ &= T_1 + T_2 + T_3. \end{aligned}$$

For the estimation of the first term, we shall use the identity $\dot{M} = \sum_i v_i w_i w_i^T$ where (w_1, \dots, w_p) is an orthonormal basis of eigenvectors associated to (v_1, \dots, v_p) :

$$\begin{aligned} |T_1| &\leq \sup_{k,\Lambda} |g_k(\Lambda)| \sum_{\lambda} |\text{Tr}(\dot{M}P_{\lambda})| \\ &\leq \sup_{k,\Lambda} |g_k(\Lambda)| \sum_{\lambda} \left| \sum_i v_i \langle w_i, P_{\lambda} w_i \rangle \right| \\ &\leq \sup_{k,\Lambda} |g_k(\Lambda)| \sum_i \sum_{\lambda} |v_i| \|P_{\lambda} w_i\|^2 \\ &= \sup_{k,\Lambda} |g_k(\Lambda)| \sum_i |v_i|. \end{aligned}$$

The second term is by symmetry

$$T_2 = \sum_{j,k:\lambda_j \neq \lambda_k} d_k^{-1} d_j^{-1} \frac{g_k(\Lambda) - g_j(\Lambda)}{\lambda_k - \lambda_j} \text{Tr}(\dot{M}P_{\lambda_j} \dot{M}P_{\lambda_k})$$

We need to notice the following: the relation (33) implies that the symmetric function f satisfies

$$f_1(x, y) - f_2(x, y) = \frac{1}{2}(f_1(x, y) - f_2(x, y) - f_1(y, x) + f_2(y, x)) \quad (36)$$

and considering the function

$$u(t) = f_1(x + t\delta, y - t\delta) - f_2(x + t\delta, y - t\delta), \quad \delta = y - x$$

Equation (36) implies

$$2|f_1(x, y) - f_2(x, y)| = |u(0) - u(1)| \leq \sup_{0 \leq t \leq 1} |u'(t)| \leq |y - x| \|f_{11} - 2f_{12} + f_{22}\|_{\infty}.$$

Hence we have

$$\left| \frac{g_k - g_j}{\lambda_k - \lambda_j} \right| \leq \frac{1}{2} \|g_{jj} + g_{kk} - 2g_{jk}\|_{\infty} = \frac{1}{2} \sup_{\Lambda \in Q} |\langle e_j - e_k, \nabla^2 g(\Lambda)(e_j - e_k) \rangle| \leq \gamma.$$

Hence

$$\begin{aligned} |T_2| &\leq \gamma \sum_{j,k:\lambda_j \neq \lambda_k} d_k^{-1} d_j^{-1} |\text{Tr}(\dot{M}P_{\lambda_j} \dot{M}P_{\lambda_k})| \\ &= \gamma \sum_{\lambda \neq \mu} |\text{Tr}(\dot{M}P_{\lambda} \dot{M}P_{\mu})| \\ &= \gamma \sum_{\lambda \neq \mu} \text{Tr}((P_{\lambda} \dot{M}P_{\mu})^T P_{\lambda} \dot{M}P_{\mu}) \\ &= \gamma \sum_{\lambda \neq \mu} \text{Tr}(\dot{M}P_{\lambda} \dot{M}P_{\mu}) \\ &= \gamma \text{Tr}(\dot{M}^2) - \gamma \sum_{\lambda} \text{Tr}(\dot{M}P_{\lambda} \dot{M}P_{\lambda}). \end{aligned}$$

For T_3 notice that using (23)

$$|T_3| \leq \gamma \sum_k \dot{\lambda}_k^2 = \gamma \sum_{\lambda} \text{Tr}(\dot{M}P_{\lambda} \dot{M}P_{\lambda})$$

hence $|T_2| + |T_3| \leq \gamma \text{Tr}(\dot{M}^2)$ and Equation (25) is proved. \square

References

- [1] Zhidong Bai and Jack W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, first edition, 2006. MR2567175
- [2] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. Concentration inequalities using the entropy method. *Ann. Probab.*, 31(3):1583–1614, 2003. MR1989444
- [3] A. Guionnet and O. Zeitouni. Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.*, 5:119–136 (electronic), 2000. MR1781846
- [4] Adityanand Guntuboyina and Hannes Leeb. Concentration of the spectral measure of large Wishart matrices with dependent entries. *Electron. Commun. Probab.*, 14:334–342, 2009. MR2535081
- [5] Andreas Kriegl, Mark Losik, and Peter W. Michor. Choosing roots of polynomials smoothly. II. *Israel J. Math.*, 139:183–188, 2004. MR2041790
- [6] Colin McDiarmid. On the method of bounded differences. In *Surveys in combinatorics, 1989 (Norwich, 1989)*, volume 141 of *London Math. Soc. Lecture Note Ser.*, pages 148–188. Cambridge Univ. Press, Cambridge, 1989. MR1036755