Concentration properties of the eigenvalues of the Gram matrix

Andreas Maurer

May 14, 2009

Abstract

We consider the concentration of the eigenvalues of the Gram matrix for a sample of iid vectors distributed in the unit ball of a Hilbert space. The square-root term in the deviation bound is shown to scale with the largest eigenvalue, the remaining term decaying as n^{-1}. This result is the consequence of a general concentration inequality.

1 Introduction

Let $\mathbf{X} = (X_1, ..., X_n)$ be a vector of independent random variables with values in the unit ball $\mathbb{B}$ of some Hilbert space H, $G(\mathbf{X})$ the Gramian $G(\mathbf{X})_{ij} = \langle X_i, X_j \rangle$ and $\lambda_d = \lambda_d(\mathbf{X})$ the k-th eigenvalue of $G(\mathbf{X})$ in descending order, where each eigenvalue is repeated according to its multiplicity. We will prove the following concentration property of the random variable λ_d.

Theorem 1 *For* $t > 0$

$$\Pr\{\lambda_d - E\lambda_d > t\} \leq \exp\left(\frac{-t^2}{16E\lambda_{\max} + 6t}\right)$$

and

$$\Pr\{E\lambda_d - \lambda_d > t\} \leq \exp\left(\frac{-t^2}{16E\lambda_{\max} + 4t}\right)$$

Since X is distributed in the unit ball, the trace of $G(\mathbf{X})$ can be at most n, but $\lambda_{\max}$ can be much smaller, so the above bound can be considerably better than what we get if the bounded difference inequality (see [4]) is applied to the eigenvalues of the Gramian (see [5]).

Let $\hat{C}(\mathbf{X})$ be the random operator on H defined by

$$\left\langle \hat{C}(\mathbf{X})\, y, z \right\rangle = \frac{1}{n} \sum_{i=1}^{n} \langle y, X_i \rangle \langle X_i, z \rangle .$$

$\hat{C}$ describes the inertial moments of the empirical distribution $(1/n) \sum_{i=1}^{n} \delta_{X_i}$ about the origin. The nonzero eigenvalues μ_d of $\hat{C}$ satisfy $\mu_d = \lambda_d / n$. Our result can be converted into a purely empirical bound on the μ_d as in the following corollary:

Corollary 2 *Let $\delta \in (0,1)$. Then*

$$\Pr\left\{\mu_d - E\mu_d \leq \sqrt{\frac{16\mu_{\max}\ln 2/\delta}{n}} + \frac{12\ln 2/\delta}{n}\right\} \geq 1-\delta$$

and

$$\Pr\left\{E\mu_d - \mu_d \leq \sqrt{\frac{16\lambda_{\max}\ln 2/\delta}{n}} + \frac{10\ln 2/\delta}{n}\right\} \geq 1-\delta.$$

The proof of Theorem 1 relies on a general concentration result, which may be of independent interest. To state it we have to introduce some notation.

Suppose that $\Omega = \prod_1^n \Omega_i$ is some product space. If $\mathbf{x} \in \Omega$, $k \in \{1,...,n\}$ and $y \in \Omega_k$ we write $\mathbf{x}_{y,k}$ for the vector obtained from $\mathbf{x}$ by replacing the k-th component with y. Also, if $F : \Omega \to \mathbb{R}$ is bounded, we define a function $\Delta_F : \Omega \to \mathbb{R}$ by

$$\Delta_F(\mathbf{x}) = \sum_k \left(F(\mathbf{x}) - \inf_{y\in\Omega_k} F(\mathbf{x}_{y,k})\right)^2.$$

If $\mathbf{X}$ is a random vector distributed in Ω we write EF for the expectation of the random variable $F(\mathbf{X})$.

Theorem 3 *Let $\mathbf{X} = (X_1,...,X_n)$ be a vector of independent random variables with values in spaces $\Omega_1,...,\Omega_n$, $Z = Z(\mathbf{x})$ and $W = W(\mathbf{x})$ real functions on $\Omega = \prod_1^n \Omega_i$ and $a \geq 1$ such that*

(i) $0 \leq Z \leq W$

(ii) $\Delta_Z \leq aW$

(iii) $\Delta_W \leq aW$

Then

$$\Pr\{Z - EZ > t\} \leq \exp\left(\frac{-t^2}{4aEW + 3at/2}\right).$$

and, if in addition $Z(\mathbf{x}) - Z(\mathbf{x}_{y,k}) \leq 1$, for all k, $y \in \Omega_k$, then

$$\Pr\{EZ - Z > t\} \leq \exp\left(\frac{-t^2}{4aEW + at}\right).$$

The purely empirical bound of Corollary 2 is also valid in the more general setting of Theorem 3.

Corollary 4 *Under the conditions of Theorem 3, if $W(\mathbf{x}) - W(\mathbf{x}_{y,k}) \leq 1$ for all $k, y \in \Omega_k$, we have for $\delta \in (0,1)$:*

$$\Pr\left\{Z - EZ > \sqrt{4aW\ln 2/\delta} + 3a\ln 2/\delta\right\} < \delta,$$

and, if in addition $Z(\mathbf{x}) - Z(\mathbf{x}_{y,k}) \leq 1$ for all $k, y \in \Omega_k$, then

$$\Pr\left\{EZ - Z > \sqrt{4aW\ln 2/\delta} + \frac{5}{2}a\ln 2/\delta\right\} < \delta.$$

2 Proofs

We first introduce some additional notation and state some useful auxiliary results. Then we prove Theorem 3 and Corollary 4, and finally we apply these results to the concentration of eigenvalues.

Let Z be a bounded random variable, $\beta \in \mathbb{R} \backslash \{0\}$. The Helmholtz energy is the real number

$$H_Z(\beta) = \frac{1}{\beta} \ln E e^{\beta Z}.$$

By l'Hospital's rule the function H_Z is continuously extended to $\mathbb{R}$ by defining $H_Z(0) = EZ$. The thermal expectation at inverse temperature β is defined by

$$E_{\beta Z} W = \frac{E W e^{\beta Z}}{E e^{\beta Z}}.$$

We will also make repeated use of the real function g defined by

$$g(t) = \begin{cases} \left(e^{-t} + t - 1\right)/t^2 & \text{for} \quad t \neq 0 \\ 1/2 & \text{for} \quad t = 0 \end{cases} . \tag{1}$$

The function g is positive, nonincreasing, and for $t \leq 0$ and $a > 0$ we have

$$\frac{a g(t)}{1 - a t g(t)} \leq \frac{\max\{1, a\}}{2}. \tag{2}$$

The following lemma is proved in [3] (Lemma 11).

Lemma 5 *For $\beta > 0$ and any $Z : \Omega \to \mathbb{R}$*

(i)

$$\ln E\left[e^{\beta(Z - E[Z])}\right] \leq \frac{\beta}{2} \int_0^\beta E_{\gamma Z}\left[\Delta_Z\right] d\gamma. \tag{3}$$

(ii) If $Z - \inf_k Z \leq 1$ for all k, then

$$\ln E\left[e^{\beta(EZ - Z)}\right] \leq \beta g(-\beta) \int_0^\beta E_{-\gamma Z}\left[\Delta_Z\right] d\gamma. \tag{4}$$

A proof of the following decoupling lemma can be found in [1].

Lemma 6 *We have*

$$E_{\beta Z}[f] \leq \beta^2 H'(\beta) + \ln E_P\left[e^f\right]. \tag{5}$$

We also need two technical optimization inequalities.

Lemma 7 *For $t \geq 0$ we have*

$$\inf_{\beta \in [0,1)} -\beta t + \frac{\beta^2 (2 - \beta)}{(1 - \beta)^2} \leq \frac{-t^2}{8 + 3t}$$

Proof. Consider the polynomial

$$p(s) = 3s^2 - 3s - s^3 + 1.$$

Then $p(1) = 0$, $p'(1) = 0$ and $p''(s) \leq 0$ for all $s \geq 1$. It follows that $p(s) \leq 0$ for all $s \geq 1$. Now define

$$h(\beta, t) = \frac{\beta^2 (2-\beta)}{(1-\beta)^2} - \beta t + \frac{t^2}{8+3t}.$$

It suffices to show that $\inf_{\beta \in [0,1)} h(\beta, t) \leq 0$ for all $t \geq 0$. Write $s = \sqrt{1 + t/2}$, so that $s \geq 1$. Then

$$\begin{aligned} \inf_{\beta \in [0,1)} h(\beta, t) &= \inf_{\beta \in [0,1)} h\left(\beta, 2\left(s^2-1\right)\right) \leq h\left(1 - \frac{1}{s}, 2\left(s^2-1\right)\right) \\ &= \frac{\left(s^2-1\right)}{s\left(1+3s^2\right)} p(s) \leq 0. \end{aligned}$$

■

Lemma 8 *Let C and b denote two positive real numbers, $t > 0$. Then*

$$\inf_{\beta \in [0,1/b)} \left(-\beta t + \frac{C\beta^2}{1-b\beta}\right) \leq \frac{-t^2}{2(2C+bt)}. \tag{6}$$

The proof of this result can be found in [3] (Lemma 12).

Proof of Theorem 3. We first claim that for $\beta \in (0, 2/a)$

$$\ln E\left[e^{\beta W}\right] \leq \frac{\beta E W}{1 - a\beta/2}, \tag{7}$$

a fact, which we will need for both tailbounds. Using (3) and assumption (iii) we have for $\beta > 0$ that

$$\ln E\left[e^{\beta(W - E[W])}\right] \leq \frac{a\beta}{2} \int_0^\beta E_{\gamma W}[W]\, d\gamma = \frac{a\beta}{2} \ln E e^{\beta W},$$

where the last identity follows from the fact that $E_{\gamma W}[W] = (d/d\gamma) \ln E e^{\gamma W}$. Thus

$$\ln E\left[e^{\beta W}\right] \leq \frac{a\beta}{2} \ln E e^{\beta W} + \beta E W,$$

and rearranging this inequality for $\beta \in (0, 2/a)$ establishes the claim.

Now we prove the upwards deviation bound. For $\beta \in (0, 2/a)$ by Lemma 6 for any random variable W,

$$\begin{aligned}
\int_0^\beta E_{\gamma Z}[W]\, d\gamma &\leq \int_0^\beta \gamma^2 H'(\gamma)\, d\gamma + \beta \ln E\left[e^W\right] \\
&= \beta \ln E\left[e^{\beta Z}\right] - 2\int_0^\beta \ln E\left[e^{\gamma Z}\right] d\gamma + \beta \ln E\left[e^W\right] \quad (*) \\
&\leq \beta \ln E\left[e^{\beta Z}\right] + \beta \ln E\left[e^W\right] \quad (**) \\
&= \beta \ln E\left[e^{\beta Z - E[Z]}\right] + \beta^2 E[Z] + \beta \ln E\left[e^W\right].
\end{aligned}$$

In (*) we used integration by parts and in (**) the fact that $\ln E\left[e^{\gamma Z}\right] \geq 0$ if $\gamma \geq 0$, since $Z \geq 0$. So, replacing W by βW we get by Lemma 5 (ii) and $\Delta_Z \leq aW$

$$\begin{aligned}
\ln E\left[e^{\beta Z - E[Z]}\right] &\leq \frac{a}{2}\int_0^\beta E_{\gamma Z}[\beta W]\, d\gamma \\
&\leq \frac{a\beta}{2} \ln E\left[e^{\beta Z - E[Z]}\right] + \frac{a\beta^2}{2} E[Z] + \frac{a\beta}{2} \ln E\left[e^{\beta W}\right].
\end{aligned}$$

Substitution of (7) and subtracting $(a\beta/2) \ln E\left[e^{\beta Z - E[Z]}\right]$ gives

$$\begin{aligned}
\left(1 - \frac{a\beta}{2}\right) \ln E\left[e^{\beta Z - E[Z]}\right] &\leq \frac{a\beta^2}{2} E[Z] + \frac{a}{2} \frac{\beta^2 E[W]}{1 - a\beta/2} \\
&\leq \beta^2 \frac{a}{2} E[W] \left(1 + \frac{1}{1 - a\beta/2}\right),
\end{aligned}$$

where we used $EZ \leq EW$ for the second inequality. Dividing $1 - a\beta/2$ we obtain

$$\ln E\left[e^{\beta Z - E[Z]}\right] \leq \frac{a}{2} E[W] \frac{\beta^2 (2 - a\beta/2)}{(1 - a\beta/2)^2}.$$

Now we make use of Lemma 7

$$\begin{aligned}
&\inf_{\beta \in [0, 2/a)} \frac{a}{2} E[W] \frac{\beta^2 (2 - a\beta/2)}{(1 - a\beta/2)^2} - \beta t \\
= \;& \frac{2}{a} E[W] \inf_{\beta \in [0,1)} \left[\frac{\beta^2 (2 - \beta)}{(1 - \beta)^2} - \beta \left(\frac{t}{E[W]}\right)\right] \\
\leq \;& \frac{-t^2}{4aE[W] + 3at/2}.
\end{aligned}$$

Conclude with Markov's inequality.

To prove the lower tailbound let again $\beta \in (0, 2/a)$. Using Lemma 5 (ii) and $\Delta_Z \leq aW$ we get

$$\ln Ee^{\beta(EZ - Z)} \leq \beta g(-\beta) \int_0^\beta E_{-\gamma Z}[\Delta_Z]\, d\gamma \leq a g(-\beta) \int_0^\beta E_{-\gamma Z}[\beta W]\, d\gamma. \quad (8)$$

Since Z is nonnegative, $\ln Ee^{-\gamma Z}$ is nonincreasing and $\int_0^\beta \ln Ee^{-\gamma Z} d\gamma \geq \beta \ln Ee^{-\beta Z}$. From integration by parts we therefore find that

$$\int_0^\beta \gamma^2 H'(-\gamma)\, d\gamma = \beta \ln Ee^{-\beta Z} - 2\int_0^\beta \ln Ee^{-\gamma Z} d\gamma \leq -\beta \ln Ee^{-\beta Z},$$

By the decoupling lemma 6 it follows that

$$\int_0^\beta E_{-\gamma Z}[\beta W]\, d\gamma \leq \int_0^\beta \left(\gamma^2 H'(-\gamma) + \ln Ee^{\beta W}\right) d\gamma \leq -\beta \ln Ee^{-\beta Z} + \beta \ln Ee^{\beta W}.$$

Resubstitution of this result in (8) gives

$$\begin{aligned} \ln Ee^{\beta(EZ-Z)} &\leq ag(-\beta)\left(-\beta \ln Ee^{-\beta Z} + \beta \ln Ee^{\beta W}\right) \\ &= -a\beta g(-\beta) \ln Ee^{\beta(EZ-Z)} + ag(-\beta)\left(\beta^2 EZ + \beta \ln Ee^{\beta W}\right). \end{aligned}$$

Now add $a\beta g(-\beta) \ln Ee^{\beta(EZ-Z)}$ to both sides, factor out $\ln Ee^{\beta(EZ-Z)}$ and rearrange to get

$$\ln Ee^{\beta(EZ-Z)} \leq \frac{ag(-\beta)}{1 + a\beta g(-\beta)}\left(\beta^2 EZ + \beta \ln Ee^{\beta W}\right) \leq \frac{a}{2}\left(\beta^2 EZ + \beta \ln Ee^{\beta W}\right),$$

where we used (2). But for $\beta \in (0, 2/a)$ we can substitute inequality (7) and use assumption (i) to get

$$\begin{aligned} \ln Ee^{\beta(EZ-Z)} &\leq \frac{a}{2}\left(\beta^2 EZ + \frac{\beta^2 E[W]}{1 - a\beta/2}\right) \leq \frac{aE[W]}{2}\left(\frac{2\beta^2 - a\beta^3/2}{1 - a\beta/2}\right) \\ &\leq aE[W]\frac{\beta^2}{1 - a\beta/2}. \end{aligned}$$

Now Lemma 8 gives us

$$\inf_{\beta \in 0, 2/a}\left(-\beta t + aE[W]\frac{\beta^2}{1 - a\beta/2}\right) \leq \frac{-t^2}{4aE[W] + at}.$$

Conclude with Markovs inequality. ■

Proof of Corollary 4. Equating the two deviation probabilities in Theorem 3 to $\delta/2$ gives

$$\Pr\left\{Z - EZ > 2\sqrt{EW}\sqrt{a \ln 2/\delta} + \frac{3a \ln 2/\delta}{2}\right\} < \delta/2, \tag{9}$$

and, if $Z(\mathbf{x}) - Z(\mathbf{x}_{y,k})$ for all $k, y \in \Omega_k$, then

$$\Pr\left\{EZ - Z > 2\sqrt{EW}\sqrt{a \ln 2/\delta} + a \ln 2/\delta\right\} < \delta/2. \tag{10}$$

Theorem 13, 2nd conclusion in [3] shows that under the conditions of the corollary also

$$\Pr\left\{EW - W > \sqrt{2aEW \ln 2/\delta}\right\} < \delta/2,$$

from which we derive

$$\Pr\left\{\sqrt{EW} > \sqrt{W} + \sqrt{2a \ln 2/\delta}\right\} < \delta/2.$$

If we use a union bound to substitute this inequality in (9) and (10) and observe that $\sqrt{2} < 3/2$, we obtain the conclusions. ■

To prove Theorem 1 and Corollary 2 we use the following technical result on the eigenvalues of the Gramian:

Proposition 9 *Let $\mathbb{B}$ be the unit ball in some separable real Hilbert-space. For $\mathbf{x} \in \mathbb{B}^n$ define $\lambda_d(\mathbf{x})$ to be the d-th eigenvalue (in descending order) of the Gramian $\langle x_i, x_j \rangle$. Then $\forall \mathbf{x} \in \mathbb{B}^n$, $k \in \{1, ..., n\}$ we have*

$$\lambda_d(\mathbf{x}) - \inf_{y \in \mathbb{B}} \lambda_d(\mathbf{x}_{y,k}) \leq 2 \text{ and } \Delta_{\lambda_d}(\mathbf{x}) \leq 4\lambda_{\max}(\mathbf{x}).$$

Proof. Fix $\mathbf{x} \in \mathbb{B}^n$ and some integer $k \in \{1, ..., n\}$. We first claim that

$$\inf_{y \in \mathbb{B}} \lambda_d(\mathbf{x}_{y,k}) = \lambda_d(\mathbf{x}_{0,k}).$$

The l.h.s. is clearly less than or equal the r.h.s. so we just have to show the reverse inequality. It is easily verified that $\lambda_d(\mathbf{x})$ is also the d-th eigenvalue of the finite-rank operator $T(\mathbf{x}) \in L(H)$ defined by

$$T(\mathbf{x}) v = \sum_{i=1}^{n} \langle v, x_i \rangle x_i \text{ for } v \in H.$$

Now let $y \in \mathbb{B}$ be arbitrary and let Q_y be the operator $Q_y v = \langle v, y \rangle y$. Then

$$T(\mathbf{x}_{y,k}) = T(\mathbf{x}_{0,k}) + Q_y.$$

By Weyls monotonicity theorem (Corollary 4.3.3 in [2]) the d-th eigenvalue of $T(\mathbf{x}_{0,k})$ can only increase by adding the positive operator Q_y. Since the eigenvalues of $T(\mathbf{x})$ are the same as those of $G(\mathbf{x})$ we have $\lambda_d(\mathbf{x}_{0,k}) \leq \lambda_d(\mathbf{x}_{y,k})$, which proves the claim.

Now let V be the span of the d dominant eigenvectors $v_1, ..., v_d$ of $G(\mathbf{x})$, and let W be the span of the $d-1$ dominant eigenvectors of $G(\mathbf{x}_{0,k})$. Then $\dim W^{\perp} + \dim V = n + 1$, so $W^{\perp} \cap V \neq \{0\}$ and we can choose a unit vector $u \in W^{\perp} \cap V$. We now use the variational characterization of the eigenvalues (Theorem 4.2.11 in [2]): Since $u \in V$ we have $\lambda_d(\mathbf{x}) \leq \langle G(\mathbf{x}) u, u \rangle$, and since

$u \in W^{\perp}$ we have $\langle G(\mathbf{x}_{0,k}) u, u\rangle \leq \lambda_d(\mathbf{x}_{0,k})$. Thus, using the definition of the Gramian, polarization and Cauchy-Schwarz,

$$
\begin{aligned}
\lambda_d(\mathbf{x}) - \lambda_d(\mathbf{x}_{y,k}) &\leq \langle (G(\mathbf{x}) - G(\mathbf{x}_{0,k})) u, u\rangle \\
&\leq \|u_k(x_k - 0)\| \left\| \sum_i u_i \left(x_i + (\mathbf{x}_{0,k})_i\right) \right\| \\
&\leq |u_k| \left(\left\| \sum_i u_i x_i \right\| + \left\| \sum_i u_i (\mathbf{x}_{0,k})_i \right\| \right) \\
&= |u_k| \left(\langle G(\mathbf{x}) u, u\rangle^{1/2} + \langle G(\mathbf{x}_{0,k}) u, u\rangle^{1/2} \right) \\
&\leq 2|u_k| \langle G(\mathbf{x}) u, u\rangle^{1/2} \leq 2|u_k| \lambda_{\max}^{1/2}.
\end{aligned}
$$

The first conclusion follows from taking the infimum over $y \in \mathbb{B}$. The second conclusion is obtained by squaring and summing over k. ■

Proof of Theorem 1 and Corollary 2. Set $Z = \lambda_d(\mathbf{X})/2$, $W = \lambda_{\max}(\mathbf{X})/2$. Clearly $0 \leq Z \leq W$. By the previous proposition $Z(\mathbf{x}) - \inf_y Z(\mathbf{x}_{y,k}) \leq 1$, $\Delta_Z \leq \lambda_{\max} = 2W$ and $\Delta_W \leq 2W$, so that Theorem 3 and Corollary 4 can be applied with $a = 2$. Theorem 1 and Corollary 2 follow. ■

References

[1] S.Boucheron, G.Lugosi, P.Massart, Concentration Inequalities using the entropy method, *Annals of Probability* 31, Nr 3, 2003

[2] R. A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985

[3] A. Maurer. Concentration inequalities for functions of independent variables. *Random Structures and Algorithms* 29: 121–138, 2006

[4] C.McDiarmid, "Concentration", in *Probabilistic Methods of Algorithmic Discrete Mathematics*, p. 195-248. Springer, Berlin, 1998.

[5] J. Shawe-Taylor, C. K. I. Williams, N. Cristianini, J. S. Kandola: On the eigenspectrum of the gram matrix and the generalization error of kernel-PCA. *IEEE Transactions on Information Theory* 51(7): 2510-2522, 2005