Dominated concentration

Andreas Maurer

Adalbertstr 55
D-80799 Munich
am@andreas-maurer.eu

Abstract

The concentration properties of one random variable may be governed by the values of another random variable which is concentrated and more easily analyzed. We present a general concentration inequality to handle such cases and apply it to the eigenvalues of the Gram matrix for a sample of independent vectors distributed in the unit ball of a Hilbert space. For large samples the deviation of the eigenvalues from their mean is shown to scale with the largest eigenvalue.

Key words: Concentration inequalities, random matrices

1. Introduction

For all of the following we assume that $\mathbf{\Omega} = \prod_1^n \Omega_i$ is some product space with product probability $\mu = \otimes_1^n \mu_k$ and that $F : \mathbf{\Omega} \to \mathbb{R}$ is some bounded measurable function. We write $\mathbb{E}F = \int F d\mu$. If $\mathbf{x} \in \mathbf{\Omega}$, $k \in \{1, ..., n\}$ and $y \in \Omega_k$ we use $\mathbf{x}_{y,k}$ to denote the vector obtained from $\mathbf{x}$ by replacing the k-th component with y, and we define a function $DF : \mathbf{\Omega} \to \mathbb{R}$ by

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

The function DF is a local measure of the sensitivity of F to modifications of its individual arguments. It is shown in (Maurer, 2006) that uniform bounds on DF lead to exponential tail inequalities for F, and that the upwards deviation bounds so obtained improve over the results obtained from Talagrand's convex distance inequality in many cases. If the function DF is bounded by a constant multiple of F itself other concentration properties can be deduced, as in the following result taken from Maurer (2006).

Theorem 1. *Suppose that $a > 0$ and that*

$$DF \leq aF. \tag{1}$$

Then for $t > 0$

$$\Pr\left\{F - \mathbb{E}\left[F\right] \geq t\right\} \leq \exp\left(\frac{-t^2}{2a\mathbb{E}\left[F\right] + at}\right),$$

and, if $a \geq 1$ *and* $F - \inf_k F \leq 1\,\forall k$*, then*

$$\Pr\left\{\mathbb{E}\left[F\right] - F \geq t\right\} \leq \exp\left(\frac{-t^2}{2a\mathbb{E}\left[F\right]}\right).$$

These results were derived from the entropy method, a technique which has been developed and refined by Ledoux, Bobkov, Massart, Boucheron, Lugosi, Rio, Bousquet and others (see Ledoux (1996), Massart (2000), Boucheron et al (2003), etc). The entropy method is rooted in the tensorization property of the entropy and seems to be evolving into a general toolbox to derive concentration inequalities. Recently Boucheron et al (2009) demonstrated that Theorem 1 above can be used to derive a version of Talagrand's convex distance inequality. The authors also weakened the condition $a \geq 1$ to $a \geq 1/3$ in the lower tail bound above. Following them we will call a function F *weakly self-bounded*, if it satisfies condition (1) above.

In some situations it is not possible to prove weak self-boundedness of F, but there is another function G which is weakly-self bounded, and DF is bounded by a constant multiple of G. In this situation one may use the following result, which is the principal contribution of this paper.

Theorem 2. *Suppose that* $F, G : \mathbf{\Omega} \to \mathbb{R}$ *and* $a \geq 1$ *such that*
(i) $0 \leq F \leq G$
(ii) $DF \leq aG$
(iii) $DG \leq aG$
Then, for $t > 0$*,*

$$\Pr\left\{F - \mathbb{E}F > t\right\} \leq \exp\left(\frac{-t^2}{4a\mathbb{E}G + 3at/2}\right).$$

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

$$\Pr\left\{\mathbb{E}F - F > t\right\} \leq \exp\left(\frac{-t^2}{4a\mathbb{E}G + at}\right).$$

The proof of Theorem 2 also uses the entropy method and the tools developed by Boucheron et al (2003) and Maurer (2006).

In many applications in learning theory concentration inequalities are used to estimate the expectation of a random variable from the observation of a single sample vector, when the underlying distribution is unknown. In such cases one might use the following corollary, which results from combining Theorem 1 with Theorem 2.

Corollary 3. *Under the conditions of Theorem 2, if* $G(\mathbf{x}) - G(\mathbf{x}_{y,k}) \leq 1$ *for all* $k, y \in \Omega_k$*, we have for* $\delta \in (0,1)$*:*

$$\Pr\left\{F - \mathbb{E}F \leq \sqrt{4aG \ln 2/\delta} + 3a \ln 2/\delta\right\} \geq 1 - \delta,$$

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

$$\Pr\left\{\mathbb{E}F - F \leq \sqrt{4aG \ln 2/\delta} + \frac{5}{2} a \ln 2/\delta\right\} \geq 1 - \delta.$$

To exemplify the utility of these results, 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 $\mathbb{H}$, let $A(\mathbf{X})$ be the Gramian $A(\mathbf{X})_{ij} = \langle X_i, X_j \rangle$ and $\lambda_d = \lambda_d(\mathbf{X})$ the d-th eigenvalue of $A(\mathbf{X})$ in descending order, counting eigenvalues according to their multiplicity. We will prove the following concentration property of the random variable λ_d.

Theorem 4. *For* $t > 0$

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

and

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

Since X is distributed in the unit ball, the trace of $A(\mathbf{X})$ 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 (McDiarmid, 1998) is applied to the eigenvalues of the Gramian, as done by Shawe-Taylor et al (2005).

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

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

$\hat{C}$ is sometimes called the (non-centered) empirical covariance operator. It 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 $\mu_d = \lambda_d / n$, as will be shown in Lemma 10 below. As with the more general Theorem 2, we obtain the following, purely empirical bound:

Corollary 5. *Let* $\delta \in (0,1)$*. Then*

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

and

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

For large n the size of the confidence interval for our estimation of $\mathbb{E}\mu_d$ by μ_d scales with the observed value of $\sqrt{\mu_{\max}}$, or, equivalently, with the largest singular value of the data-matrix $\mathbf{X}$.

2. Proofs

We first introduce some additional notation and state some useful auxiliary results. Then we prove Theorem 2 and Corollary 3, and finally we apply these results to the concentration of eigenvalues. Questions of measurability will be ignored throughout.

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

$$H_F(\beta) = \frac{1}{\beta}\ln \mathbb{E}e^{\beta F}.$$

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

$$\mathbb{E}_{\beta F} f = \frac{\mathbb{E}fe^{\beta F}}{\mathbb{E}e^{\beta F}} \text{ for } f : \mathbf{\Omega} \to \mathbb{R}.$$

To lighten notation we will not explicitely denote the dependence of H_F and $\mathbb{E}_{\beta F}$ on the underlying measure μ. We will also make repeated use of the real function g defined by

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

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

$$\frac{ag(t)}{1-atg(t)} \leq \frac{\max\{1,a\}}{2}. \tag{3}$$

The following lemma is proved in (Maurer, 2006, Lemma 11).

Lemma 6. *For $\beta > 0$ and any $F : \Omega \to \mathbb{R}$*

(i)

$$\ln \mathbb{E}\left[e^{\beta(F-\mathbb{E}[F])}\right] \leq \frac{\beta}{2}\int_0^\beta \mathbb{E}_{\gamma F}[DF]\,d\gamma. \tag{4}$$

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

$$\ln \mathbb{E}\left[e^{\beta(\mathbb{E}F-F)}\right] \leq \beta g(-\beta)\int_0^\beta \mathbb{E}_{-\gamma F}[DF]\,d\gamma. \tag{5}$$

Our proofs rely on the following decoupling technique: If μ and ν are two probability measures and ν is absolutely continuous w.r.t. μ then it is easy to show that

$$\mathbb{E}_\nu f \leq KL\left(d\nu, d\mu\right) + \ln \mathbb{E}_\mu e^f,$$

where $KL\left(.,.\right)$ is the Kullback-Leibler divergence or relative entropy $KL\left(d\nu, d\mu\right) = \mathbb{E}_\nu \ln\left(d\nu/d\mu\right)$. A straightforward computation gives

$$KL\left(\frac{e^{\beta F}d\mu}{\mathbb{E}_\mu e^{\beta F}}, d\mu\right) = \beta\mathbb{E}_{\beta F}F - \ln \mathbb{E}e^{\beta F} = \beta^2 H_F'\left(\beta\right),$$

so we obtain the following

Lemma 7. *We have for any* $f : \Omega \rightarrow \mathbb{R}$

$$\mathbb{E}_{\beta F}\left[f\right] \leq \beta^2 H_F'\left(\beta\right) + \ln \mathbb{E}\left[e^f\right]. \tag{6}$$

We also need two technical optimization inequalities.

Lemma 8. *For* $t \geq 0$ *we have*

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

Proof. Consider the polynomial

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

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

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

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

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

□

Lemma 9. *Let* C *and* b *denote two positive real numbers,* $t > 0$*. Then*

$$\inf_{\beta\in[0,1/b)} \left(-\beta t + \frac{C\beta^2}{1-b\beta}\right) \leq \frac{-t^2}{2\left(2C+bt\right)}. \tag{7}$$

The proof of this lemma can be found in (Maurer, 2006, Lemma 12).

PROOF OF THEOREM 2. We first claim that for $\beta \in (0, 2/a)$

$$\ln \mathbb{E}\left[e^{\beta G}\right] \leq \frac{\beta \mathbb{E} G}{1 - a\beta/2}, \tag{8}$$

a fact which we will need for both tailbounds. Using Lemma 6 (i) and the weak self-boundedness assumption (iii) we have for $\beta > 0$ that

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

where the last identity follows from the fact that $\mathbb{E}_{\gamma G}\left[G\right] = (d/d\gamma) \ln \mathbb{E} e^{\gamma G}$. Thus

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

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 7 for any function $f : \mathbf{\Omega} \to \mathbb{R}$,

$$\begin{aligned}
\int_0^\beta \mathbb{E}_{\gamma F}\left[f\right] d\gamma &\leq \int_0^\beta \gamma^2 H_F'(\gamma)\, d\gamma + \beta \ln \mathbb{E}\left[e^f\right] \\
&= \beta \ln \mathbb{E}\left[e^{\beta F}\right] - 2 \int_0^\beta \ln \mathbb{E}\left[e^{\gamma F}\right] d\gamma + \beta \ln \mathbb{E}\left[e^f\right] \\
&\leq \beta \ln \mathbb{E}\left[e^{\beta F}\right] + \beta \ln \mathbb{E}\left[e^f\right] \\
&= \beta \ln \mathbb{E}\left[e^{\beta(F - \mathbb{E}[F])}\right] + \beta^2 \mathbb{E}\left[F\right] + \beta \ln \mathbb{E}\left[e^f\right].
\end{aligned}$$

In the second line we used integration by parts and in the third line the fact that $\ln \mathbb{E}\left[e^{\gamma F}\right] \geq 0$ if $\gamma \geq 0$, since $F \geq 0$. So, replacing f by βG we get by Lemma 6 (i) and $DF \leq aG$

$$\begin{aligned}
\ln \mathbb{E}\left[e^{\beta(F - \mathbb{E}[F])}\right] &\leq \frac{a}{2} \int_0^\beta \mathbb{E}_{\gamma F}\left[\beta G\right] d\gamma \\
&\leq \frac{a\beta}{2} \ln \mathbb{E}\left[e^{\beta(F - \mathbb{E}[F])}\right] + \frac{a\beta^2}{2} \mathbb{E}\left[F\right] + \frac{a\beta}{2} \ln \mathbb{E}\left[e^{\beta G}\right].
\end{aligned}$$

Substitution of (8) and subtracting $(a\beta/2) \ln \mathbb{E}\left[e^{\beta(F - \mathbb{E}[F])}\right]$ gives

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

where we used $\mathbb{E}F \leq \mathbb{E}G$ for the second inequality. Dividing by $1 - a\beta/2$ we obtain

$$\ln \mathbb{E}\left[e^{\beta(F-\mathbb{E}[F])}\right] \leq \frac{a}{2}\mathbb{E}[G]\,\frac{\beta^2(2-a\beta/2)}{(1-a\beta/2)^2}.$$

Now we make use of Lemma 8 for $t > 0$

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

From Markov's inequality we now conclude that for $t > 0$

$$\Pr\{F - \mathbb{E}F > t\} \leq \inf_{\beta\in(0,2/a)} \mathbb{E}e^{\beta(F-\mathbb{E}F)-\beta t} \leq \exp\left(\frac{-t^2}{4a\mathbb{E}[G] + 3at/2}\right).$$

To prove the lower tailbound let again $\beta \in (0, 2/a)$. Using Lemma 6 (ii) and $DF \leq aG$ we get

$$\ln \mathbb{E}e^{\beta(\mathbb{E}F-F)} \leq \beta g(-\beta)\int_0^\beta \mathbb{E}_{-\gamma F}[DF]\,d\gamma \leq ag(-\beta)\int_0^\beta \mathbb{E}_{-\gamma F}[\beta G]\,d\gamma. \quad (9)$$

Since F is nonnegative, $\ln \mathbb{E}e^{-\gamma F}$ is nonincreasing and $\int_0^\beta \ln \mathbb{E}e^{-\gamma F}d\gamma \geq \beta \ln \mathbb{E}e^{-\beta F}$. From integration by parts (using $H_F'(-\gamma) = -(d/d\gamma)H_F(-\gamma)$) we therefore find that

$$\int_0^\beta \gamma^2 H_F'(-\gamma)\,d\gamma = \beta \ln \mathbb{E}e^{-\beta F} - 2\int_0^\beta \ln \mathbb{E}e^{-\gamma F}d\gamma \leq -\beta \ln \mathbb{E}e^{-\beta F},$$

By the decoupling lemma 7 it follows that

$$\int_0^\beta \mathbb{E}_{-\gamma F}[\beta G]\,d\gamma \leq \int_0^\beta \left(\gamma^2 H_F'(-\gamma) + \ln \mathbb{E}e^{\beta G}\right)d\gamma \leq -\beta \ln \mathbb{E}e^{-\beta F} + \beta \ln \mathbb{E}e^{\beta G}.$$

Resubstitution of this result in (9) gives

$$\begin{aligned}
\ln \mathbb{E}e^{\beta(\mathbb{E}F-F)} &\leq ag(-\beta)\left(-\beta \ln \mathbb{E}e^{-\beta F} + \beta \ln \mathbb{E}e^{\beta G}\right) \\
&= -a\beta g(-\beta)\ln \mathbb{E}e^{\beta(\mathbb{E}F-F)} + ag(-\beta)\left(\beta^2\mathbb{E}F + \beta \ln \mathbb{E}e^{\beta G}\right).
\end{aligned}$$

Now add $a\beta g(-\beta)\ln \mathbb{E}e^{\beta(\mathbb{E}F-F)}$ to both sides, factor out $\ln \mathbb{E}e^{\beta(\mathbb{E}F-F)}$ and rearrange to get

$$\ln \mathbb{E}e^{\beta(\mathbb{E}F-F)} \leq \frac{ag(-\beta)}{1 + a\beta g(-\beta)}\left(\beta^2\mathbb{E}F + \beta \ln \mathbb{E}e^{\beta G}\right) \leq \frac{a}{2}\left(\beta^2\mathbb{E}F + \beta \ln \mathbb{E}e^{\beta G}\right),$$

where we used (3). But for $\beta \in (0, 2/a)$ we can substitute inequality (8) and use assumption (i) to get

$$\begin{aligned}\ln \mathbb{E} e^{\beta(\mathbb{E}F - F)} &\leq \frac{a}{2}\left(\beta^2 \mathbb{E}F + \frac{\beta^2 \mathbb{E}[G]}{1 - a\beta/2}\right) \leq \frac{a\mathbb{E}[G]}{2}\left(\frac{2\beta^2 - a\beta^3/2}{1 - a\beta/2}\right) \\ &\leq a\mathbb{E}[G] \frac{\beta^2}{1 - a\beta/2}.\end{aligned}$$

Now Lemma 9 gives us

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

Conclude with Markov's inequality as before □

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

$$\Pr\left\{F - \mathbb{E}F > 2\sqrt{\mathbb{E}G}\sqrt{a \ln 2/\delta} + \frac{3a \ln 2/\delta}{2}\right\} < \delta/2, \tag{10}$$

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

$$\Pr\left\{\mathbb{E}F - F > 2\sqrt{\mathbb{E}G}\sqrt{a \ln 2/\delta} + a \ln 2/\delta\right\} < \delta/2. \tag{11}$$

It follows from Theorem 1 that under the conditions of the corollary also

$$\Pr\left\{\mathbb{E}G - G > \sqrt{2a\mathbb{E}G \ln 2/\delta}\right\} < \delta/2,$$

from which we derive

$$\Pr\left\{\sqrt{\mathbb{E}G} > \sqrt{G} + \sqrt{2a \ln 2/\delta}\right\} < \delta/2.$$

If we use a union bound to substitute this inequality in (10) and (11) and observe that $\sqrt{2} < 3/2$, we obtain the conclusions □

To apply our result to the eigenvalues of Gramian matrices and the related inertial operators we first further clarify the relationship between these objects. Suppose that $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{H}^n$ is some configuration of vectors. The Gramian matrix $A(\mathbf{x})$ and the inertial operator $\hat{C}(\mathbf{x})$ have already been introduced. We define an operator $T(\mathbf{x}) \in \mathcal{L}(\mathbb{H})$ by the formula

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

so that $\hat{C}(\mathbf{x}) = (1/n) T(\mathbf{x})$. Notice that $T(\mathbf{x})$ has rank at most n. If S is a finite-rank operator (or matrix) acting on a Hilbert space we use $\sigma(S)$ to denote the set of eigenvalues of S.

Lemma 10. $\sigma(A(\mathbf{x})) \subseteq \sigma(T(\mathbf{x}))$ *and* $\sigma(T(\mathbf{x})) \setminus \{0\} \subseteq \sigma(A(\mathbf{x}))$.

Proof. If λ is any eigenvalue of $A(\mathbf{x})$ let $\boldsymbol{\gamma} \in \mathbb{R}^n$ be a corresponding eigenvector and set $e = \sum \gamma_i x_i$ to obtain

$$T(\mathbf{x})\, e = \sum_i \sum_j \gamma_j \langle x_j, x_i \rangle x_i = \sum_i (A(\mathbf{x}) \boldsymbol{\gamma})_i x_i = \lambda \sum_i \gamma_i x_i = \lambda e,$$

so λ is an eigenvalue of $T(\mathbf{x})$, which shows that $\sigma(A(\mathbf{x})) \subseteq \sigma(T(\mathbf{x}))$.

Observe, that we also showed that if $\boldsymbol{\gamma} \in \ker(A(\mathbf{x}))$, then $\sum \gamma_i x_i \in \ker(T(\mathbf{x}))$. Now if λ is a nonzero eigenvalue of $T(\mathbf{x})$ let e be a corresponding eigenvector. Since λ is nonzero we must have $e \in [\mathbf{x}]$, so $e = \sum \gamma_i x_i$, with $\gamma \neq 0$, and, by the above, also $\boldsymbol{\gamma} \notin \ker(A(\mathbf{x}))$. We have

$$\left(A(\mathbf{x})^2 \gamma\right)_j = \sum_i (A(\mathbf{x}) \boldsymbol{\gamma})_i \langle x_i, x_j \rangle = \langle T(\mathbf{x})\, e, x_j \rangle = \lambda \langle e, x_j \rangle = \lambda A(\mathbf{x})\, \gamma,$$

so λ is also an eigenvalue of $A(\mathbf{x})$ with nonzero eigenvector $A(\mathbf{x})\, \gamma$ □

Note that, whenever $\dim(\mathbb{H}) > n$ and the x_i are independent, zero is an eigenvalue of $T(\mathbf{x})$, but not of $A(\mathbf{x})$ so that $\sigma(A(\mathbf{x})) \subset \sigma(T(\mathbf{x}))$ is a proper inclusion. To prove Theorem 4 and Corollary 5 we will use the following technical result:

Proposition 11. *Let $\mathbb{B}$ be the unit ball in some separable real Hilbert-space $\mathbb{H}$. For $\mathbf{x} \in \mathbb{B}^n$ define $\lambda_d(\mathbf{x})$ to be the d-th eigenvalue (in descending order) of the Gramian $A_{ij}(\mathbf{x}) = \langle x_i, x_j \rangle$. Then $\forall \mathbf{x} \in \mathbb{B}^n$, $k \in \{1, ..., n\}$ we have*

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

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. By Lemma 10 $\lambda_d(\mathbf{x})$ is also the d-th eigenvalue of $T(\mathbf{x})$. Now let $y \in \mathbb{B}$ be arbitrary and let Q_y be the operator defined by $Q_y v = \langle v, y \rangle y$, for $v \in \mathbb{H}$. Then

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

By Weyls monotonicity theorem (see Horn and Johnson, 1985, Corollary 4.3.3) the d-th eigenvalue of $T(\mathbf{x}_{0,k})$ can only increase by adding the positive operator Q_y. Since the nonzero eigenvalues of $T(\mathbf{x})$ are the same as those of $A(\mathbf{x})$ 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 $A(\mathbf{x})$, and let W be the span of the $d-1$ dominant eigenvectors of $A(\mathbf{x}_{0,k})$. 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 Horn and Johnson (1985)): Since $u \in V$ we have $\lambda_d(\mathbf{x}) \leq \langle A(\mathbf{x}) u, u \rangle$, and since $u \in W^{\perp}$ we have $\langle A(\mathbf{x}_{0,k}) u, u \rangle \leq \lambda_d(\mathbf{x}_{0,k})$. Thus, using the definition of the Gramian, polarization and Cauchy-Schwarz,

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

which implies the first conclusion. The second conclusion is obtained by squaring and summing over k □

Proof of Theorem 4 and Corollary 5. Set $F = \lambda_d(\mathbf{X})/2$, $G = \lambda_{\max}(\mathbf{X})/2$. Clearly $0 \leq F \leq G$. By the previous proposition $F(\mathbf{x}) - \inf_y F(\mathbf{x}_{y,k}) \leq 1$, $DF \leq \lambda_{\max} = 2G$ and $DG \leq 2G$, so that Theorem 2 and Corollary 3 can be applied with $a = 2$. Theorem 4 and Corollary 5 follow □

The author wishes to thank the anonymous referee for many valuable comments and suggestions.

References

S. Boucheron, G. Lugosi, P. Massart, 2003. Concentration Inequalities using the entropy method, *Annals of Probability* Vol. 31, No. 3, pp. 1583-1614.

S. Boucheron, G. Lugosi, P. Massart, 2009. On concentration of self-bounding functions, *Electronic Journal of Probability*, Vol.14, Paper no. 64, pages 1884–1899.

R. A. Horn and C. R. Johnson, 1985. *Matrix Analysis*. Cambridge University Press.

M. Ledoux, 1996. On Talagrand's deviation inequalities for product measures. *ESAIM: Probability and Statistics*, 1:63–87.

M. Ledoux, 2001. *The Concentration of Measure Phenomenon,* AMS Surveys and Monographs 89.

P. Massart, 2000. Some applications of concentration inequalities to statistics, *Annales de la Faculté des Sciences de Toulouse* IX: 245–303.

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

C. McDiarmid, 1998. Concentration, in *Probabilistic Methods of Algorithmic Discrete Mathematics*, p. 195–248. Springer, Berlin.

J. Shawe-Taylor, C. K. I. Williams, N. Cristianini, J. S. Kandola, 2004. On the eigenspectrum of the gram matrix and the generalization error of kernel-PCA. *IEEE Transactions on Information Theory* 51(7): 2510–2522.