A tail inequality for quadratic forms of subgaussian random vectors

Daniel Hsu[1], Sham M. Kakade[1,2], and Tong Zhang[3]

[1]Microsoft Research New England
[2]Department of Statistics, Wharton School, University of Pennsylvania
[3]Department of Statistics, Rutgers University

October 14, 2011

Abstract

We prove an exponential probability tail inequality for positive semidefinite quadratic forms in a subgaussian random vector. The bound is analogous to one that holds when the vector has independent Gaussian entries.

1 Introduction

Suppose that $x = (x_1, \ldots, x_n)$ is a random vector. Let $A \in \mathbb{R}^{m \times n}$ be a fixed matrix. A natural quantity that arises in many settings is the quadratic form $\|Ax\|^2 = x^\top (A^\top A) x$. Throughout $\|v\|$ denotes the Euclidean norm of a vector v, and $\|M\|$ denotes the spectral (operator) norm of a matrix M. We are interested in how close $\|Ax\|^2$ is to its expectation.

Consider the special case where $x_1, \ldots, x_n$ are independent standard Gaussian random variables. The following proposition provides an (upper) tail bound for $\|Ax\|^2$.

Proposition 1. *Let $A \in \mathbb{R}^{m \times n}$ be a matrix, and let $\Sigma := A^\top A$. Let $x = (x_1, \ldots, x_n)$ be an isotropic multivariate Gaussian random vector with mean zero. For all $t > 0$,*

$$\Pr\left[\|Ax\|^2 > \mathrm{tr}(\Sigma) + 2\sqrt{\mathrm{tr}(\Sigma^2)t} + 2\|\Sigma\|t\right] \leq e^{-t}.$$

The proof, given in Appendix A.2, is straightforward given the rotational invariance of the multivariate Gaussian distribution, together with a tail bound for linear combinations of χ^2 random variables due to Laurent and Massart (2000). We note that a slightly weaker form of Proposition 1 can be proved directly using Gaussian concentration (Pisier, 1989).

In this note, we consider the case where $x = (x_1, \ldots, x_n)$ is a *subgaussian* random vector. By this, we mean that there exists a $\sigma \geq 0$, such that for all $\alpha \in \mathbb{R}^n$,

$$\mathbb{E}\left[\exp\left(\alpha^\top x\right)\right] \leq \exp\left(\|\alpha\|^2\sigma^2/2\right).$$

We provide a sharp upper tail bound for this case analogous to one that holds in the Gaussian case (indeed, the same as Proposition 1 when $\sigma = 1$).

E-mail: dahsu@microsoft.com, skakade@wharton.upenn.edu, tzhang@stat.rutgers.edu

Tail inequalities for sums of random vectors

One motivation for our main result comes from the following observations about sums of random vectors. Let $a_1, \ldots, a_n$ be vectors in a Euclidean space, and let $A = [a_1|\cdots|a_n]$ be the matrix with a_i as its ith column. Consider the squared norm of the random sum

$$\|Ax\|^2 = \left\|\sum_{i=1}^n a_i x_i\right\|^2 \tag{1}$$

where $x := (x_1, \ldots, x_n)$ is a martingale difference sequence with $\mathbb{E}[x_i \mid x_1, \ldots, x_{i-1}] = 0$ and $\mathbb{E}[x_i^2 \mid x_1, \ldots, x_{i-1}] = \sigma^2$. Under mild boundedness assumptions on the x_i, the probability that the squared norm in (1) is much larger than its expectation

$$\mathbb{E}[\|Ax\|^2] = \sigma^2 \sum_{i=1}^n \|a_i\|^2 = \sigma^2 \operatorname{tr}(A^\top A)$$

falls off exponentially fast. This can be shown, for instance, using the following lemma by taking $u_i = a_i x_i$ (the proof is standard, but we give it for completeness in Appendix A.1).

Proposition 2. *Let $u_1, \ldots, u_n$ be a martingale difference vector sequence (*i.e.*, $\mathbb{E}[u_i|u_1, \ldots, u_{i-1}] = 0$ for all $i = 1, \ldots, n$) such that*

$$\sum_{i=1}^n \mathbb{E}\big[\|u_i\|^2 \mid u_1, \ldots, u_{i-1}\big] \leq v \quad \textit{and} \quad \|u_i\| \leq b$$

for all $i = 1, \ldots, n$, almost surely. For all $t > 0$,

$$\Pr\left[\left\|\sum_{i=1}^n u_i\right\| > \sqrt{v} + \sqrt{8vt} + (4/3)bt\right] \leq e^{-t}.$$

After squaring the quantities in the stated probabilistic event, Proposition 2 gives the bound

$$\|Ax\|^2 \leq \sigma^2 \cdot \operatorname{tr}(A^\top A) + \sigma^2 \cdot O\left(\operatorname{tr}(A^\top A)(\sqrt{t} + t) + \sqrt{\operatorname{tr}(A^\top A)} \max_i \|a_i\|(t + t^{3/2}) + \max_i \|a_i\|^2 t^2\right)$$

with probability at least $1 - e^{-t}$ when the x_i are almost surely bounded by 1 (or any constant).

Unfortunately, this bound obtained from Proposition 2 can be suboptimal when the x_i are subgaussian. For instance, if the x_i are Rademacher random variables, so $\Pr[x_i = +1] = \Pr[x_i = -1] = 1/2$, then it is known that

$$\|Ax\|^2 \leq \operatorname{tr}(A^\top A) + O\left(\sqrt{\operatorname{tr}((A^\top A)^2) t} + \|A\|^2 t\right) \tag{2}$$

with probability at least $1 - e^{-t}$. A similar result holds for any subgaussian distribution on the x_i (Hanson and Wright, 1971). This is an improvement over the previous bound because the deviation terms (*i.e.*, those involving t) can be significantly smaller, especially for large t.

In this work, we give a simple proof of (2) with explicit constants that match the analogous bound when the x_i are independent standard Gaussian random variables.

2 Positive semidefinite quadratic forms

Our main theorem, given below, is a generalization of (2).

Theorem 1. *Let $A \in \mathbb{R}^{m \times n}$ be a matrix, and let $\Sigma := A^\top A$. Suppose that $x = (x_1, \ldots, x_n)$ is a random vector such that, for some $\mu \in \mathbb{R}^n$ and $\sigma \geq 0$,*

$$\mathbb{E}\left[\exp\left(\alpha^\top (x - \mu)\right)\right] \leq \exp\left(\|\alpha\|^2 \sigma^2 / 2\right) \tag{3}$$

for all $\alpha \in \mathbb{R}^n$. For all $t > 0$,

$$\Pr\left[\ \|Ax\|^2 > \sigma^2 \cdot \left(\operatorname{tr}(\Sigma) + 2\sqrt{\operatorname{tr}(\Sigma^2) t} + 2\|\Sigma\| t\right) + \|A\mu\|^2 \cdot \left(1 + 4\left(\frac{\|\Sigma\|^2}{\operatorname{tr}(\Sigma^2)} t\right)^{1/2} + \frac{4\|\Sigma\|^2}{\operatorname{tr}(\Sigma^2)} t\right)^{1/2}\ \right] \leq e^{-t}.$$

Remark 1. Note that when $\mu = 0$ and $\sigma = 1$ we have:

$$\Pr\left[\ \|Ax\|^2 > \operatorname{tr}(\Sigma) + 2\sqrt{\operatorname{tr}(\Sigma^2) t} + 2\|\Sigma\| t\ \right] \leq e^{-t}$$

which is the same as Proposition 1.

Remark 2. Our proof actually establishes the following upper bounds on the moment generating function of $\|Ax\|^2$ for $0 \leq \eta < 1/(2\sigma^2\|\Sigma\|)$:

$$\begin{aligned}\mathbb{E}\left[\exp\left(\eta\|Ax\|^2\right)\right] &\leq \mathbb{E}\left[\exp\left(\sigma^2\|A^\top z\|^2 \eta + \mu^\top A^\top z \sqrt{2\eta}\right)\right] \\ &\leq \exp\left(\sigma^2 \operatorname{tr}(\Sigma)\eta + \frac{\sigma^4 \operatorname{tr}(\Sigma^2)\eta^2 + \|A\mu\|^2 \eta}{1 - 2\sigma^2\|\Sigma\|\eta}\right)\end{aligned}$$

where z is a vector of m independent standard Gaussian random variables.

Proof of Theorem 1. Let z be a vector of m independent standard Gaussian random variables (sampled independently of x). For any $\alpha \in \mathbb{R}^m$,

$$\mathbb{E}\left[\exp\left(z^\top \alpha\right)\right] = \exp\left(\|\alpha\|^2/2\right).$$

Thus, for any $\lambda \in \mathbb{R}$ and $\varepsilon \geq 0$,

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda z^\top Ax\right)\right] &\geq \mathbb{E}\left[\exp\left(\lambda z^\top Ax\right) \,\middle|\, \|Ax\|^2 > \varepsilon\right] \cdot \Pr\left[\|Ax\|^2 > \varepsilon\right] \\ &\geq \exp\left(\frac{\lambda^2 \varepsilon}{2}\right) \cdot \Pr\left[\|Ax\|^2 > \varepsilon\right].\end{aligned} \tag{4}$$

Moreover,

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda z^\top Ax\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[\exp\left(\lambda z^\top A(x - \mu)\right) \,\middle|\, z\right] \exp\left(\lambda z^\top A\mu\right)\right] \\ &\leq \mathbb{E}\left[\exp\left(\frac{\lambda^2\sigma^2}{2}\|A^\top z\|^2 + \lambda\mu^\top A^\top z\right)\right]\end{aligned} \tag{5}$$

Let $USV^\top$ be a singular value decomposition of A; where U and V are, respectively, matrices of orthonormal left and right singular vectors; and $S = \operatorname{diag}(\sqrt{\rho_1}, \dots, \sqrt{\rho_m})$ is the diagonal matrix of corresponding singular values. Note that

$$\|\rho\|_1 = \sum_{i=1}^m \rho_i = \operatorname{tr}(\Sigma), \quad \|\rho\|_2^2 = \sum_{i=1}^m \rho_i^2 = \operatorname{tr}(\Sigma^2), \quad \text{and} \quad \|\rho\|_\infty = \max_i \rho_i = \|\Sigma\|.$$

By rotational invariance, $y := U^\top z$ is an isotropic multivariate Gaussian random vector with mean zero. Therefore $\|A^\top z\|^2 = z^\top U S^2 U^\top z = \rho_1 y_1^2 + \dots + \rho_m y_m^2$ and $\mu^\top A^\top z = \nu^\top y = \nu_1 y_1 + \dots + \nu_m y_m$, where $\nu := SV^\top \mu$ (note that $\|\nu\|^2 = \|SV^\top\mu\|^2 = \|A\mu\|^2$). Let $\gamma := \lambda^2\sigma^2/2$. By Lemma 1,

$$\mathbb{E}\left[\exp\left(\gamma \sum_{i=1}^m \rho_i y_i^2 + \frac{\sqrt{2\gamma}}{\sigma}\sum_{i=1}^m \nu_i y_i\right)\right] \le \exp\left(\|\rho\|_1\gamma + \frac{\|\rho\|_2^2\gamma^2 + \|\nu\|^2\gamma/\sigma^2}{1 - 2\|\rho\|_\infty\gamma}\right) \tag{6}$$

for $0 \le \gamma < 1/(2\|\rho\|_\infty)$. Combining (4), (5), and (6) gives

$$\Pr\left[\|Ax\|^2 > \varepsilon\right] \le \exp\left(-\varepsilon\gamma/\sigma^2 + \|\rho\|_1\gamma + \frac{\|\rho\|_2^2\gamma^2 + \|\nu\|^2\gamma/\sigma^2}{1 - 2\|\rho\|_\infty\gamma}\right)$$

for $0 \le \gamma < 1/(2\|\rho\|_\infty)$ and $\varepsilon \ge 0$. Choosing

$$\varepsilon := \sigma^2(\|\rho\|_1 + \tau) + \|\nu\|^2\sqrt{1 + \frac{2\|\rho\|_\infty\tau}{\|\rho\|_2^2}} \quad \text{and} \quad \gamma := \frac{1}{2\|\rho\|_\infty}\left(1 - \sqrt{\frac{\|\rho\|_2^2}{\|\rho\|_2^2 + 2\|\rho\|_\infty\tau}}\right),$$

we have

$$\begin{aligned}\Pr\left[\|Ax\|^2 > \sigma^2(\|\rho\|_1 + \tau) + \|\nu\|^2\sqrt{1 + \frac{2\|\rho\|_\infty\tau}{\|\rho\|_2^2}}\right] &\le \exp\left(-\frac{\|\rho\|_2^2}{2\|\rho\|_\infty^2}\left(1 + \frac{\|\rho\|_\infty\tau}{\|\rho\|_2^2} - \sqrt{1 + \frac{2\|\rho\|_\infty\tau}{\|\rho\|_2^2}}\right)\right) \\ &= \exp\left(-\frac{\|\rho\|_2^2}{2\|\rho\|_\infty^2} h_1\left(\frac{\|\rho\|_\infty\tau}{\|\rho\|_2^2}\right)\right)\end{aligned}$$

where $h_1(a) := 1 + a - \sqrt{1 + 2a}$, which has the inverse function $h_1^{-1}(b) = \sqrt{2b} + b$. The result follows by setting $\tau := 2\sqrt{\|\rho\|_2^2 t} + 2\|\rho\|_\infty t = 2\sqrt{\operatorname{tr}(\Sigma^2)t} + 2\|\Sigma\|t$. □

The following lemma is a standard estimate of the logarithmic moment generating function of a quadratic form in standard Gaussian random variables, proved much along the lines of the estimate due to Laurent and Massart (2000).

Lemma 1. *Let z be a vector of m independent standard Gaussian random variables. Fix any non-negative vector $\alpha \in \mathbb{R}_+^m$ and any vector $\beta \in \mathbb{R}^m$. If $0 \le \lambda < 1/(2\|\alpha\|_\infty)$, then*

$$\log \mathbb{E}\left[\exp\left(\lambda\sum_{i=1}^m \alpha_i z_i^2 + \sum_{i=1}^m \beta_i z_i\right)\right] \le \|\alpha\|_1\lambda + \frac{\|\alpha\|_2^2\lambda^2 + \|\beta\|_2^2/2}{1 - 2\|\alpha\|_\infty\lambda}.$$

Proof. Fix $\lambda \in \mathbb{R}$ such that $0 \leq \lambda < 1/(2\|\alpha\|_\infty)$, and let $\eta_i := 1/\sqrt{1 - 2\alpha_i \lambda} > 0$ for $i = 1, \dots, m$. We have

$$\mathbb{E}\left[\exp\left(\lambda \alpha_i z_i^2 + \beta_i z_i\right)\right] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-z_i^2/2\right) \exp\left(\lambda \alpha_i z_i^2 + \beta_i z_i\right) dz_i$$
$$= \eta_i \exp\left(\frac{\beta_i^2 \eta_i^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\eta_i^2}} \exp\left(-\frac{1}{2\eta_i^2}\left(z_i - \beta_i \eta_i^2\right)^2\right) dz_i$$

so

$$\log \mathbb{E}\left[\exp\left(\lambda \sum_{i=1}^m \alpha_i z_i^2 + \sum_{i=1}^m \beta_i z_i\right)\right] = \frac{1}{2}\sum_{i=1}^m \beta_i^2 \eta_i^2 + \frac{1}{2}\sum_{i=1}^m \log \eta_i^2.$$

The right-hand side can be bounded using the inequalities

$$\frac{1}{2}\sum_{i=1}^m \log \eta_i^2 = -\frac{1}{2}\sum_{i=1}^m \log(1 - 2\alpha_i\lambda) = \frac{1}{2}\sum_{i=1}^m \sum_{j=1}^\infty \frac{(2\alpha_i\lambda)^j}{j} \leq \|\alpha\|_1 \lambda + \frac{\|\alpha\|_2^2 \lambda^2}{1 - 2\|\alpha\|_\infty \lambda}$$

and

$$\frac{1}{2}\sum_{i=1}^m \beta_i^2 \eta_i^2 \leq \frac{\|\beta\|_2^2/2}{1 - 2\|\alpha\|_\infty \lambda}.$$

□

Example: fixed-design regression with subgaussian noise

We give a simple application of Theorem 1 to fixed-design linear regression with the ordinary least squares estimator.

Let $x_1, \dots, x_n$ be fixed design vectors in $\mathbb{R}^d$. Let the responses $y_1, \dots, y_n$ be random variables for which there exists $\sigma > 0$ such that

$$\mathbb{E}\left[\exp\left(\sum_{i=1}^n \alpha_i (y_i - \mathbb{E}[y_i])\right)\right] \leq \exp\left(\sigma^2 \sum_{i=1}^n \alpha_i^2\right)$$

for any $\alpha_1, \dots, \alpha_n \in \mathbb{R}$. This condition is satisfied, for instance, if

$$y_i = \mathbb{E}[y_i] + \varepsilon_i$$

for independent subgaussian zero-mean noise variables $\varepsilon_1, \dots, \varepsilon_n$. Let $\Sigma := \sum_{i=1}^n x_i x_i^\top / n$, which we assume is invertible without loss of generality. Let

$$\beta := \Sigma^{-1}\left(\frac{1}{n}\sum_{i=1}^n x_i \mathbb{E}[y_i]\right)$$

be the coefficient vector of minimum expected squared error. The ordinary least squares estimator is given by

$$\hat{\beta} := \Sigma^{-1}\left(\frac{1}{n}\sum_{i=1}^n x_i y_i\right).$$

The excess loss $R(\hat{\beta})$ of $\hat{\beta}$ is the difference between the expected squared error of $\hat{\beta}$ and that of β:

$$R(\hat{\beta}) := \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i^\top \hat{\beta} - y_i)^2\right] - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i^\top \beta - y_i)^2\right].$$

It is easy to see that

$$R(\hat{\beta}) = \left\|\Sigma^{1/2}(\hat{\beta} - \beta)\right\|^2 = \left\|\sum_{i=1}^{n}\left(\Sigma^{-1/2}x_i\right)\left(y_i - \mathbb{E}[y_i]\right)\right\|^2.$$

By Theorem 1,

$$\Pr\left[R(\hat{\beta}) > \frac{\sigma^2\left(d + 2\sqrt{dt} + 2t\right)}{n}\right] \leq e^{-t}.$$

Note that in the case that $\mathbb{E}[(y_i - \mathbb{E}[y_i])^2] = \sigma^2$ for each i, then

$$\mathbb{E}[R(\hat{\beta})] = \frac{\sigma^2 d}{n};$$

so the tail inequality above is essentially tight when the y_i are independent Gaussian random variables.

References

D. L. Hanson and F. T. Wright. A bound on tail probabilities for quadratic forms in independent random variables. *The Annals of Mathematical Statistics*, 42(3):1079–1083, 1971.

B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28(5):1302–1338, 2000.

G. Pisier. *The volume of convex bodies and Banach space geometry.* Cambridge University Press, 1989.

A Standard tail inequalities

A.1 Martingale tail inequalities

The following is a standard form of Bernstein's inequality stated for martingale difference sequences.

Lemma 2 (Bernstein's inequality for martingales)**.** *Let $d_1, \ldots, d_n$ be a martingale difference sequence with respect to random variables $x_1, \ldots, x_n$ (*i.e., $\mathbb{E}[d_i | x_1, \ldots, x_{i-1}] = 0$ *for all $i = 1, \ldots, n$) such that $|d_i| \leq b$ and $\sum_{i=1}^{n} \mathbb{E}[d_i^2 | x_1, \ldots, x_{i-1}] \leq v$. For all $t > 0$,*

$$\Pr\left[\sum_{i=1}^{n} d_i > \sqrt{2vt} + (2/3)bt\right] \leq e^{-t}.$$

The proof of Proposition 2, which is entirely standard, is an immediate consequence of the following two lemmas together with Jensen's inequality.

Lemma 3. *Let $u_1, \ldots, u_n$ be random vectors such that*

$$\sum_{i=1}^{n} \mathbb{E}\left[\|u_i\|^2 \mid u_1, \ldots, u_{i-1}\right] \le v \quad \textit{and} \quad \|u_i\| \le b.$$

for all $i = 1, \ldots, n$, almost surely. For all $t > 0$,

$$\Pr\left[\ \left\|\sum_{i=1}^{n} u_i\right\| - \mathbb{E}\left[\left\|\sum_{i=1}^{n} u_i\right\|\right] > \sqrt{8vt} + (4/3)bt\ \right] \le e^{-t}.$$

Proof. Let $s_n := u_1 + \cdots + u_n$. Define the Doob martingale

$$d_i := \mathbb{E}[\|s_n\| \mid u_1, \ldots, u_i] - \mathbb{E}[\|s_n\| \mid u_1, \ldots, u_{i-1}]$$

for $i = 1, \ldots, n$, so $d_1 + \cdots + d_n = \|s_n\| - \mathbb{E}[\|s_n\|]$. First, clearly, $\mathbb{E}[d_i | u_1, \ldots, u_{i-1}] = 0$. Next, the triangle inequality implies

$$\begin{aligned} d_i &= \mathbb{E}\left[\|(s_n - u_i) + u_i\| \mid u_1, \ldots, u_i\right] - \mathbb{E}\left[\|(s_n - u_i) + u_i\| \mid u_1, \ldots, u_{i-1}\right] \\ &\le \mathbb{E}\left[\|s_n - u_i\| + \|u_i\| \mid u_1, \ldots, u_i\right] - \mathbb{E}\left[\|\|s_n - u_i\| - \|u_i\| \mid u_1, \ldots, u_{i-1}\right] \\ &= \|u_i\| + \mathbb{E}\left[\|u_i\| \mid u_1, \ldots, u_{i-1}\right], \end{aligned}$$

$$\text{and similarly,}\ d_i \ge -\|u_i\| - \mathbb{E}\left[\|u_i\| \mid u_1, \ldots, u_{i-1}\right].$$

Therefore,

$$|d_i| \le \|u_i\| + \mathbb{E}\left[\|u_i\| \mid u_1, \ldots, u_{i-1}\right] \le 2b \quad \text{almost surely.}$$

Moreover,

$$\begin{aligned} \mathbb{E}\left[d_i^2 \mid u_1, \ldots, u_{i-1}\right] &\le \mathbb{E}\Big[\|u_i\|^2 + 2 \cdot \|u_i\| \cdot \mathbb{E}\left[\|u_i\| \mid u_1, \ldots, u_{i-1}\right] \\ &\qquad + \mathbb{E}\left[\|u_i\| \mid u_1, \ldots, u_{i-1}\right]^2 \mid u_1, \ldots, u_{i-1}\Big] \\ &= \mathbb{E}\left[\|u_i\|^2 \mid u_1, \ldots, u_{i-1}\right] + 3 \cdot \mathbb{E}\left[\|u_i\| \mid u_1, \ldots, u_{i-1}\right]^2 \\ &\le 4 \cdot \mathbb{E}\left[\|u_i\|^2 \mid u_1, \ldots, u_{i-1}\right], \end{aligned}$$

$$\text{so} \quad \sum_{i=1}^{n} \mathbb{E}\left[d_i^2 \mid u_1, \ldots, u_{i-1}\right] \le 4v \quad \text{almost surely.}$$

The claim now follows from Bernstein's inequality (Lemma 2). $\square$

Lemma 4. *If $u_1, \ldots, u_n$ is a martingale difference vector sequence (*i.e., $\mathbb{E}[u_i | u_1, \ldots, u_{i-1}] = 0$ *for all $i = 1, \ldots, n$), then*

$$\mathbb{E}\left[\left\|\sum_{i=1}^{n} u_i\right\|^2\right] = \sum_{i=1}^{n} \mathbb{E}\left[\|u_i\|^2\right].$$

Proof. Let $s_i := u_1 + \cdots + u_i$ for $i = 1, \ldots, n$; we have

$$\begin{aligned} \mathbb{E}\left[\|s_n\|^2\right] &= \mathbb{E}\left[\mathbb{E}\left[\|u_n + s_{n-1}\|^2 \mid u_1, \ldots, u_{n-1}\right]\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\|u_n\|^2 + 2u_n^\top s_{n-1} + \|s_{n-1}\|^2 \mid u_1, \ldots, u_{n-1}\right]\right] \\ &= \mathbb{E}\left[\|u_n\|^2\right] + \mathbb{E}\left[\|s_{n-1}\|^2\right] \end{aligned}$$

so the claim follows by induction. $\square$

A.2 Gaussian quadratic forms and χ^2 tail inequalities

It is well-known that if $z \sim \mathcal{N}(0,1)$ is a standard Gaussian random variable, then z^2 follows a χ^2 distribution with one degree of freedom. The following inequality due to Laurent and Massart (2000) gives a bound on linear combinations of χ^2 random variables.

Lemma 5 (χ^2 tail inequality; Laurent and Massart, 2000). *Let $q_1, \ldots, q_n$ be independent χ^2 random variables, each with one degree of freedom. For any vector $\gamma = (\gamma_1, \ldots, \gamma_n) \in \mathbb{R}_+^n$ with non-negative entries, and any $t > 0$,*

$$\Pr\left[\sum_{i=1}^{n} \gamma_i q_i > \|\gamma\|_1 + 2\sqrt{\|\gamma\|_2^2 t} + 2\|\gamma\|_\infty t\right] \leq e^{-t}.$$

Proof of Proposition 1. Let $V\Lambda V^\top$ be an eigen-decomposition of $A^\top A$, where V is a matrix of orthonormal eigenvectors, and $\Lambda := \operatorname{diag}(\rho_1, \ldots, \rho_n)$ is the diagonal matrix of corresponding eigenvalues $\rho_1, \ldots, \rho_n$. By the rotational invariance of the distribution, $z := V^\top x$ is an isotropic multivariate Gaussian random vector with mean zero. Thus, $\|Ax\|^2 = z^\top \Lambda z = \rho_1 z_1^2 + \cdots + \rho_n z_n^2$, and the z_i^2 are independent χ^2 random variables, each with one degree of freedom. The claim now follows from a tail bound for χ^2 random variables (Lemma 5, due to Laurent and Massart, 2000). $\square$