Random Vectors and Matrices
High Dim Probability & Linear Algebra
for ML and Sig Proc

Namrata Vaswani

Iowa State University
2-norm of a subGaussian vector is close to $\sqrt{n}$ w.h.p.: 

**Theorem (Concen of norm of a subG vector)**

Let $X \in \mathbb{R}^n$ be a r. vector with independent entries $X_i$ with $\mathbb{E}[X_i^2] = 1$. Let $K = \max_i \|X_i\|_{\psi_2}$. Then $\|X\| - \sqrt{n}$ is a sub-G r.v. with sub-G norm at most $K^2$. Equivalently,

$$\text{Pr}(\|X\| - \sqrt{n} \geq t) \leq 2\exp(-ct^2/K^4)$$

**Proof:**

1. **For a subG r.v. with $E[Z^2] = 1$, $K_Z \geq 1$**

   *Reason: using $1 + x \leq e^x$, with $x = Z^2/K_Z^2$ $\mathbb{E}[1 + Z^2/K_Z^2] \leq \mathbb{E}[e^{Z^2/K_Z^2}]$ which implies $1 + 1/K_Z^2 \leq \mathbb{E}[e^{Z^2/K_Z^2}]$. By subG property, $\mathbb{E}[e^{Z^2/K_Z^2}] \leq 2$ and this gives $K_Z \geq 1$.**

2. **Consider $\frac{1}{n}\|X\|^2 - 1 = \frac{1}{n} \sum_i (X_i^2 - 1)$. By the properties from earlier, $X_i^2 - 1$ are independent, zero mean, sub-expo r.v.s with $K_{\text{expo}} \leq CK^2$. So we can apply the sub-expo Bernstein inequality to conclude that**

$$\text{Pr}(\frac{1}{n}\|X\|^2 - 1 \geq u) \leq 2\exp\left(-c \frac{n}{K^4} \min(u^2, u)\right)$$

*(the above also used $K \geq 1$).*
3 Use $|z - 1| \geq \delta$ implies $|z^2 - 1| \geq \max(\delta, \delta^2)$ and the fact that $A \Rightarrow B$ implies $\Pr(A) \leq \Pr(B)$ to conclude that

$$\Pr\left(\frac{1}{\sqrt{n}} \|X\| - 1 \geq \delta\right) \leq \Pr\left(\frac{1}{n} \|X\|^2 - 1 \geq \max(\delta, \delta^2)\right) \leq 2 \exp\left(-c \frac{n}{K^4} \delta^2\right)$$

(used: for $u = \max(\delta, \delta^2)$, $\min(u^2, u) = \delta^2$).

4 Set $\delta = t/\sqrt{n}$ to conclude that

$$\Pr\left(\|X\| - \sqrt{n} \geq t\right) \leq 2 \exp\left(-c \frac{1}{K^4} t^2\right)$$

2 When working with random vectors, we generally subtract mean first to get zero-mean random vectors.

3 Isotropic random vectors: $X \in \mathbb{R}^n$ is isotropic if

$$\mathbb{E}[XX^\top] = I_n$$

Properties of isotropic $X$

$\mathbb{E}[(a^\top X)^2] = \|a\|^2$ for all $a \in \mathbb{R}^n$ (this is equivalent to the definition)
High-dimensional random vectors III

- $\mathbb{E}[\|X\|^2] = n$

- $X, Y$ independent and isotropic, then $\mathbb{E}[(X'Y)^2] = n$
  - Implication of this and concentration of norm result (Remark 3.2.5): can argue that if $X, Y$ are indep., then $\frac{X}{\|X\|}, \frac{Y}{\|Y\|}$ are almost orthogonal, i.e. their inner product is of order $1/\sqrt{n}$.
  - TBD: quantify above claim, it is not quantified in the book.

- Examples of isotropic r. vectors:
  - i.i.d symmetric Bernoulli;
  - standard Gaussian vector;
  - any “product” distribution (coordinates of $X$ are independent) with zero mean and unit variance;
  - coordinate distribution ($X$ equally likely to be $\sqrt{n}e_i, i = 1, 2, \ldots, n$; recall $e_i$ is the $i$-th column of $I$)
  - $X \sim \text{Unif}(\sqrt{n}S^{n-1})$: this is isotropic but coordinates are not independent (proof is not obvious, TBD);
  - unif distrib on frames

4 Sub-Gaussian random vector
Definition: 
$X$ is a sub-G vector iff $a'X$ is sub-G for all $a \in \mathbb{R}^n$. Sub-G norm of $X$ is

$$\|X\|_{\psi_2} := \sup_{a \in S^{n-1}} \|a'X\|_{\psi_2}$$

Sub-G with independent coordinates $X = (X_1, X_2, \ldots, X_n)'$ with $X_i$’s independent sub-G: then

$$\|X\|_{\psi_2} \leq C \max_{i=1,2,\ldots,n} \|X_i\|_{\psi_2}$$

Spherical distribution is sub-Gaussian: $Z \sim Unif(\sqrt{n}S^{n-1})$ is sub-G with subG norm at most $C$. Proof:

1. Use the following property: For a standard Gaussian random vector, $X$, i.e., $X \sim \mathcal{N}(0, I)$

$$\theta := \frac{X}{\|X\|} \sim Unif(S^{n-1}),$$

Also, $\|X\|, \theta$ are independent.

2. Use this property to conclude that we can express $Z$ as

$$Z = \sqrt{n}G/\|G\|$$

where $G \sim \mathcal{N}(0, I)$. 
To prove that $Z$ is sub-G, we need to prove that $a' Z$ is sub-G for all $a \in \mathbb{R}^n$.

Rotation invariance property of $G$ implies that $a' G = e'_1 U'_a G = \tilde{G}_1$ where $\tilde{G} = U'_a G \sim \mathcal{N}(0, I)$ too and $\|\tilde{G}\| = \|G\|$. Here $U_a$ is an orthonormal matrix with first column $a/\|a\|$.

Thus, w.l.o.g., $a' Z = \sqrt{n}\tilde{G}_1/\|\tilde{G}_1\|$ and we need to bound $\Pr(\sqrt{n}\tilde{G}_1/\|\tilde{G}_1\| \geq u)$.

Apply concentration of norm result on $\|\tilde{G}\|$ with $t = \sqrt{n}/2$ to conclude that

$$\Pr(\|\tilde{G}\| \geq \sqrt{n}/2) \geq 1 - 2\exp(-cn)$$

(Ev)

(follows since $K$ for a standard Gaussian vector is a constant).

Using total probability with $Ev, Ev^c$,

$$\Pr(\sqrt{n}\tilde{G}_1/\|\tilde{G}_1\| \geq u) \leq \Pr(\sqrt{n}\tilde{G}_1/\|\tilde{G}_1\| \geq u \text{ and } Ev) + \Pr(Ev^c)$$

$$\leq \Pr(\tilde{G}_1 \geq u/2 \text{ and } Ev) + 2\exp(-cn)$$

$$\leq \Pr(\tilde{G}_1 \geq u/2) + 2\exp(-cn)$$

$$\leq 2\exp(-u^2/8) + 2\exp(-cn) \leq 4\exp(-u^2/8)$$

Reason for last bound:
If $u < \sqrt{n}$, then first term dominates and we can conclude that $Z$ is sub-G.
If $u \geq \sqrt{n}$, then $\Pr(\sqrt{n}\tilde{G}_1/\|\tilde{G}_1\| \geq u) = 0$ since $\tilde{G}_1 \leq \|\tilde{G}_1\|$.
Epsilon-net I

Epsilon net is a finite set of points that is used to “cover” a compact set in a metric space by using balls of radius \( \epsilon \). More precisely, it is a set of finite points so that any point on the compact set is within \( \epsilon \) distance of some point in the epsilon-net.

1. **Definition for** \( \mathcal{N}_\epsilon \) **that covers** \( S^{n-1} \) **in Euclidean distance:** \( \mathcal{N}_\epsilon \subset S^{n-1} \) is an \( \epsilon \) – **net** of \( S^{n-1} \) if for any \( x \in S^{n-1} \), there exists a \( \bar{x} \in \mathcal{N}_\epsilon \) s.t. \( \|x - \bar{x}\| \leq \epsilon \).

2. **Bound size of epsilon-net:** can use volume arguments to show that we can find an \( \epsilon \)-net that covers \( S^{n-1} \) with cardinality

\[
|\mathcal{N}_\epsilon| \leq (1 + 2/\epsilon)^n
\]

3. **Use to bound** \( \|A\| \) **by using** \( \|A\| = \max_{x \in S^{n-1}} \|Ax\| \):

Suppose \( x \) is the point on the sphere that achieves the above max. By definition, there exists an \( \bar{x}(x) \) in the net s.t. \( \|\bar{x} - x\| \leq \epsilon \). Thus

\[
\|A\| = \|Ax\| = \|A(\bar{x} + x - \bar{x})\| \leq \|A\bar{x}\| + \|A\|\|x - \bar{x}\| \leq \|A\bar{x}\| + \|A\|\epsilon
\]

So

\[
(1 - \epsilon)\|A\| \leq \|A\bar{x}\| \leq \max_{\bar{x} \in \mathcal{N}_\epsilon} \|A\bar{x}\|
\]

and hence

\[
\|A\| \leq \frac{1}{1 - \epsilon} \max_{\bar{x} \in \mathcal{N}_\epsilon} \|A\bar{x}\|
\]
4. Use to bound $\sigma_{\min}(A)$ by using $\sigma_{\min}(A) = \min_{x \in S^{n-1}} \|Ax\|$: proceed as above; this bound uses the bound on $\|A\|$ from above.

5. Use to bound $\|A\|$ by using $\|A\| = \max_{x \in S^{n-1}, y \in S^{m-1}} y'Ax$. In some proofs, the above norm definition is needed. One can show that

$$\|A\| \leq \frac{1}{1 - 2\epsilon} \max_{\bar{x} \in \mathcal{N}_\epsilon(S^{n-1}), \bar{y} \in \mathcal{N}_\epsilon(S^{m-1})} \bar{y}'A\bar{x}$$
Bound on min and max singular values of an $m \times n$ matrix with independent isotropic sub-Gaussian rows.

**Theorem (Sub-Gaussian rows matrix)**

Let $A$ be an $m \times n$ matrix whose rows, $A^i$, are independent, zero-mean, sub-G, isotropic r.vectors. Let $K = \max_i \| A^i \|_{\psi_2}$. Then, for a large enough numerical constant $C$,

$$\sqrt{m} - CK^2(\sqrt{n} + t) \leq s_i(A) \leq \sqrt{m} + CK^2(\sqrt{n} + t)$$

w.p. at least $1 - 2\exp(-t^2)$. Here $s_i(A)$ is the $i$-th singular value of $A$.

Claim: The bounds of the theorem will hold if we can instead prove that

$$\| \frac{1}{m} A' A - I \| \leq K^2 \max(\delta, \delta^2), \quad \delta = \frac{\sqrt{n} + t}{\sqrt{m}}$$

(1)

Bounding $\| \frac{1}{m} A' A - I \|$:
Approximation: use the following results for epsilon-nets: for a symmetric $M$,

$$\|M\| := \max_{x \in S^{n-1}} |x'Ax| \leq \frac{1}{1 - 2\epsilon} \max_{x \in N_\epsilon} |x'Ax|$$

where $N_\epsilon \subset S^{n-1}$ is an epsilon-net on $S^{n-1}$. By the covering number bound, we can find a $1/4$-net for which

$$|N_\epsilon| \leq (1 + 2/\epsilon)^n$$

Using these with $\epsilon = 1/4$ and simplifying,

$$\left\| \frac{1}{m} A' A - I \right\| \leq 2 \max_{x \in N_{1/4}} \left| \frac{1}{m} \|Ax\|^2 - 1 \right|$$

and

$$|N_{1/4}| \leq 9^n$$
Concentration: for a fixed $x \in \mathcal{N}_{1/4} \subset S^{n-1}$: Since the rows $A^i$ are isotropic (implies $\mathbb{E}[(x' A^i)^2] = 1$), sub-G, independent, with sub-G norm at most $K$, 

$$\frac{1}{m} \|Ax\|^2 - 1 = \frac{1}{m} \sum_{i=1}^{m} ((x' A^i)^2 - 1),$$

is a sum of $m$ independent, zero-mean, sub-expo r.v.s with sub-expo norm at most $CK^2/m$. We can apply sub-expo Bernstein ineq to conclude that

$$\Pr \left( \left| \frac{1}{m} \|Ax\|^2 - 1 \right| \geq \epsilon/2 \right) \leq 2 \exp \left( -cm \min(\epsilon^2/K^4, \epsilon/K^2) \right)$$

Use $\epsilon = K^2 \max(\delta, \delta^2)$ with $\delta = C(\sqrt{n} + \epsilon)/\sqrt{m}$ to get

$$\Pr \left( \left| \frac{1}{m} \|Ax\|^2 - 1 \right| \geq K^2 \max(\delta, \delta^2) \right) \leq 2 \exp(-cm\delta^2) \leq 2 \exp(-cC^2(n + t^2))$$
Union bound: over all $x \in \mathcal{N}_{1/4} \subset S^{n-1}$ gives:

$$\Pr\left( \max_{x \in \mathcal{N}_{1/4}} \left| \frac{1}{m} \|Ax\|^2 - 1 \right| \geq K^2 \max(\delta, \delta^2) \right) \leq 9^n 2 \exp(-cC^2(n + t^2)) \leq \exp(-t^2)$$

by choosing $C$ large enough.

By combining this with the Approximation step, (1) holds w.p. $\geq 1 - \exp(-t^2)$.

Implication of the theorem: if $m \geq CK^2 n$, then the min singular value of $A/\sqrt{m}$ is at least a constant $c < 1$ and the max singular value is at most a constant $C > 1$, thus the condition number is a constant.

Bound on expected value: using the above result and the integral identity applied to $Z = \|A^\top A - mI\|/(CK^2)$,

$$\mathbb{E}\left[ \left\| \frac{1}{m} A' A - I \right\| \right] \leq CK^2(\sqrt{n/m} + (n/m))$$
Proof: above result and \( \max(a, b) < a + b \) tells us that
\[
\Pr(Z < (\sqrt{mn} + n + \sqrt{mt} + t^2)) \geq 1 - \exp(-t^2).
\]
Let \( u_0 = (\sqrt{mn} + n) \). Thus, using integral identity applied to \( Z = \|A^\top A - mI\|/(CK^2) \),
\[
\mathbb{E}[Z] \leq u_0 + \int_{\tau = u_0}^{\infty} \Pr(Z > \tau) d\tau
\]
\[
= u_0 + \int_{t=0}^{\infty} \Pr(Z > u_0 + \sqrt{mt} + t^2)(\sqrt{m} + 2t) dt
\]
\[
\leq u_0 + \sqrt{m} \int_{t=0}^{\infty} \exp(-t^2) dt + \int_{t=0}^{\infty} \exp(-t^2) 2t dt
\]
\[
\leq u_0 + \sqrt{m} \frac{\sqrt{2\pi}}{2} + 2
\]
Second row used \( \tau = u_0 + \sqrt{mt} + t^2 \) so that \( d\tau = \sqrt{mdt} + 2tdt \); third row used Theorem conclusion; last row follows by using Gaussian pdf integral for second term and basic integration rules for last term.
Since \( u_0 = (\sqrt{mn} + n) \), for \( n \) large enough, \( u_0 + C\sqrt{m} + 2 < 1.1u_0 \). Thus, \( \mathbb{E}[Z] \leq 1.1u_0 \) and so \( \mathbb{E}[Z/m] \leq 1.1u_0/m \), i.e.,
\[
\mathbb{E}[\frac{1}{m} A' A - I] \leq CK^2(\sqrt{n/m} + (n/m))
\]
We can also use the Theorem and integral identity to show that
\[ \sqrt{m} - CK^2 \sqrt{n} \leq \mathbb{E}[s_n(A)], \text{ and } \mathbb{E}[s_1(A)] \leq \sqrt{m} + CK^2 \sqrt{n} \]

Can obtain an easy extension for the non-isotropic case as well.

Matrix Bernstein:

**Theorem (Matrix Bernstein)**

Let \( \mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m \) be independent, zero-mean, \( d_1 \times d_2 \) matrices with \( \| \mathbf{X}_i \| \leq L \) for all \( i = 1, 2, \ldots, m \). Define the “variance parameter” of the sum

\[ v := \max \left( \| \sum_i \mathbb{E}[\mathbf{X}_i \mathbf{X}_i^\top] \|, \| \sum_i \mathbb{E}[\mathbf{X}_i^\top \mathbf{X}_i] \| \right) \]

Then

\[ \Pr(\| \sum_{i=1}^m \mathbf{X}_i \| \geq t) \leq (d_1 + d_2) \exp \left( -c \frac{t^2}{v + Lt/3} \right) \leq 2 \exp(\log \max(d_1, d_2) - c \min(\frac{t^2}{v}, \frac{t}{L})) \]
For symmetric matrices $X_i$ of size $n \times n$, $v = \| \sum_i \mathbb{E}[X_i^2] \|$, $d_1 = d_2 = n$.

For nonzero mean matrices, the above bound, along with Weyl’s inequality, implies that, w.p. $\geq 1 - 2 \exp(\log \max(d_1, d_2) - c \min(\frac{t^2}{v}, \frac{t}{L}))$,

$$s_{\text{min}}\left(\sum_{i=1}^{m} \mathbb{E}[X_i]\right) - t \leq s_{\text{min}}\left(\sum_{i=1}^{m} X_i\right) \leq s_{\text{max}}\left(\sum_{i=1}^{m} X_i\right) \leq s_{\text{max}}\left(\sum_{i=1}^{m} \mathbb{E}[X_i]\right) + t$$

i.e. the min and max singular values of the sum are close to those of the expected values w.h.p.

Proof: See Vershynin book Sec 5.4 or the original reference “User-friendly tail bounds for sums of random matrices” by Joel Tropp. Main ideas:
- first prove the result for sums of symmetric matrices, then extend to any general matrices using the dilation trick;
- for symmetric matrices: bound the MGF of $\lambda_{\text{max}}(\sum_{i=1}^{m} X_i)$ conditioned on $X_1, X_2, X_{m-1}$, using Leib’s inequality or Gold-Thompson inequality; followed by averaging over $X_{m-1}$ and repeating the steps. Do this one at a time, eventually get a bound on the MGF. Not straightforward.
Matrix Bernstein vs subGaussian rows' result:
Matrix Bernstein bounds the norm of the sum of bounded, independent, zero-mean random matrices.
SubG rows' result: $A'A$ can also be interpreted as the sum of rank-one matrices $A'A = \sum_i (A^i)(A^i)'$ with $A^i$ being sub-Gaussian. Matrix Bernstein applies to this setting if the $A^i$ are bounded.
Matrix Bern is a better result for bounded r. matrices because the probability contains $\exp(\log n - (\text{terms}))$ while use of the eps-net argument for the subG rows results in the probability containing $\exp(n - (\text{terms}))$. 
Upper bound on max singular value of matrix $A$ with each
entry subG I

**Theorem (Thm 4.4.5 of book)**

$A$ is $m \times n$, each entry $A_{ij}$ is zero mean, independent, subG with subG norm at most $K$. Then

$$\|A\| \leq CK\left(\sqrt{m} + \sqrt{n} + t\right) \text{ w.p. at least } 1 - 2 \exp(-t^2)$$

Proof: use $\|A\| = \max_{x,y} x^T Ay = \sum_{ij} A_{ij} x_i y_j$, eps-net, for a fixed $x, y$, $A_{ij} x_i y_j$ is subG-$(K|x_i||y_j|)$ and so we can use subG Hoeffding.

Notice:

1. This result does not require the rows of $A$ to be isotropic. But then it only gives upper bound.

2. In particular this allows for upper bounding of the norm of a matrix in which some entries of a row are even zero.

3. Application: symmetric matrix with above/on diagonal entries subG.
Upper bound on max singular value of matrix $A$ with each entry subG II

Corollary (Cor 4.4.8 of book)

$A$ is $n \times n$, symmetric, each entry above and on diagonal is zero mean, independent, subG with subG norm at most $K$.

$$||A|| \leq CK(\sqrt{n} + t) \text{ w.p. at least } 1 - 4 \exp(-t^2)$$

Proof: $A = A^{top} + A^{bottom}$, $||A|| \leq 2||A^{top}||$, $A^{top}$ has zeros below the diagonal. So rows have a few zero entries. Can still apply Thm 4.4.5 though.

Application of this result: adjacency matrix of a graph.
Davis Kahan sin theta thereom 1

Reference: book and Spectral Methods for Data Science (by Yuxin Chen and others)

1 for subspace estimation:
Symmetric matrices $S, \hat{S}$. $U, \hat{U}$ are top $r$ eigenvectors

$$\text{SubsDist}(U, \hat{U}) \leq \frac{||S - \hat{S}||}{\lambda_r(S) - \lambda_{r+1}(\hat{S})} \leq \frac{||S - \hat{S}||}{\lambda_r(S) - \lambda_{r+1}(S) - ||S - \hat{S}||}$$

second inequality follows by Weyl.

subspace distance equals sine of largest principal angle between the subspaces

$$\text{SubsDist}(U, \hat{U}) := ||(I - \hat{U}\hat{U}^T)U||$$

2 for individual eigenvectors:

$$\sin \theta(u_i, \hat{u}_i) \leq \frac{||S - \hat{S}||}{\min_{j \neq i} |\lambda_j(S) - \lambda_i(S)|}$$

Here $\sin \theta(u_i, \hat{u}_i) = \sqrt{1 - (u_i^T\hat{u}_i)^2}$
network, node, connection – graph, vertex, edge

Graph with \( n \) vertices can have at most \( n(n - 1)/2 \) edges

Assuming everywhere node i not connected to itself.

Degree of a node: number of edges from that node.

Max degree of a graph: maximum degree of all nodes

Adjacency matrix of a graph: \( n \times n \) matrix \( A \) s.t. \( A_{ij} = 1 \) if \( i, j \) connected and zero otherwise

Random graph: nodes i,j connected with a certain probability

Erdos Renyi graph, \( ER(p) \): any pair of nodes connected w.p. \( p \) independent of all others
Communities in a network: simple model: two communities, each of size $n/2$, all connections independent, nodes within same community connecting w.p $p$, those from different communities w.p. $q < p$.

Goal: develop an algorithm to find the communities. We do not know which nodes are connected with what probability. We only know the connectivity.

Solution:

1. Define the adjacency matrix $A: n \times n$ and symmetric with 1-0 entries
2. Compute second eigenvector $A$. Call it $u_2$
3. The signs of $u_2$ provide an estimate of the community labels: if $(u_2)_i > 0$ $i$ is in commun 1, else it is in commun 2.