title: Matrix Chernoff Bounds for sums of independent random variables date: 2023-01-26 category: notes speaker: Tushant Mittal

abstract: "We do a matrix-version of our dear Chernoff bound."

Summary

In this talk, we proved a matrix-version of a simple Chernoff bound. The scalar version which we will generalize is the following.

Scalar Chernoff

Let $X = sumi *epsiai $ where $a*i in R$ are fixed, and, $eps*i$ are $pm 1$ Rademacher random variables. Then, \[prob{sum*i epsi*a*i > t} ~leq~ e^{frac{-t^2}{2sigma^2}},\;\; sigma^2 = ex{X^2} =\; sum*i a*i^2 \] Let's look at a standard proof of this to see where it breaks down. For a single Rademacher variable, one can show that $ex{e^{theta eps a}} = cosh(theta a) leq e^{frac{theta^2 a^2}{2}}$. The goal is to break the sum on n such variables as a product of these. Fix $theta > 0in R$. \[begin{align} prob{X > t} ~&=~ prob{e^{theta X} > e^{theta t}}\;\;small{text{ (Monotonicity of exponential)} }\ &leq e^{-theta t}\,ex{ e^{theta X}} \;\;small{text{ (Markov)} }\ ~&=~ e^{-theta t}\,ex{ prod*i e^{theta epsi ai}} \;\; small{(e^{a+b} = e^acdot e^b)} \ ~&=~ e^{-theta t}\,prod*iex{ e^{theta eps*i a*i}} \;\; small{text{ (Independence)} } \ ~&leq~ e^{-theta t}\,prod*i e^{frac{theta*^2 *a*i^2}{2}} \;\; small{text{ (Plugging in the 1-variable bound)} }\ ~&leq~ e^{frac{-t^2}{2(sum*i a*i^2)}} \;\; small{text{ (Optimizing for }theta) } end{align} \]

Matrix Chernoff

We will generalize the above setup as follows. Let $X = sum*i epsi*A*i $ where $A*i$ are fixed Hermitian matrices, and, $eps*i$ are $pm 1$ Rademacher random variables. Then, \[prob{lambda*{max}(X) > t} ~leq~ e^{frac{-t^2}{2sigma^2}} \]

We will give two proofs each of which give a slightly different $sigma$. The first question is to make sense of exponentials of matrix valued random variables. This, fortunately, is easy.

Lifting functions to matrices

Let $f: I to R$ be a function defined on an interval I. Let $A = QLambda Q^* $ be an Hermitian matrix such that all eigenvalues of A lies in I, i.e., $mathrm{Spec}(A)subseteq I$. Then, we can define $f(A) := Qf(Lambda) Q^*$, where $f(Lambda)$ is obtained by applying f entry-wise to the diagonal matrix $Lambda$, i.e., $f(Lambda) = mathrm{diag}(f(lambda*1), cdots, f(lambda*d))$.

Mimicking the scalar proof

One can repeat the scalar argument for the first two steps but it is not clear how to handle the term, $ex{lambda{max}(e^{theta X})}$. Ideally, we would like to have $prodi ex{lambdamax (e^{theta epsi A_i})}$. The key difficulty is that $e^{A+B}neq e^Ae^B$ for matrices. This can be resolved (by at least) two approaches.

\[begin{align} small{[text{Naive Hope}]}\;\;\; &ex{lambda{max}(e^{theta X})} ~=~ ex{lambda{max}left(prodi e^{theta epsi Ai}right)} ~leq~ prodi ex{lambdamax (e^{theta epsi Ai})} \ small{[text{Ahlswede-Winter}]}\;\;\; &ex{lambda{max}(e^{theta X})} ~leq~ ex{tr(e^{theta X})} ~leq~ dprodi ex{lambdamax (e^{theta epsi Ai})}\ small{[text{Tropp}]}\;\;\; &ex{lambda{max}(e^{theta X})} ~leq~ ex{tr(e^{theta X})} ~leq~ tr\, mathrm{exp sumi log ex{(e^{theta epsi Ai})}right) \ end{align} \]

Using the trace inequalities

AW By definition of the matrix exponential, $lambdamax(e^A) = e^{lambdamax(A)}$. Thus, treating $lambdamax(A)$ as a scalar random variable, we can plug it into the scalar inequality we used earlier, $ex{e^{theta eps a}} leq e^{frac{theta^2 a^2}{2}}$. Thus, we get, $ex{lambdamax (e^{theta epsi Ai})} leq e^{frac{theta^2 lambdamax(Ai)^2}{2}}$. It is now exactly like the scalar Chernoff and we get a variance term of $sumi lambdamax(A_i)^2$.

Tropp We need two more facts.

- Firstly, a matrix version of the 1-variable inequality. This is given by $log ex{e^theta epsi Ai} preceq frac{theta^2 A_i^2}{2}$. Here the order being used is the Loewner order ($Apreceq B$ if $B-A$ is PSD). The proof is analogous to the scalar proof and is given in Tropp's book [].
- The fact that trace-exponential is monotone in the sense that if $Apreceq B$, $tr\, e^A leq tr\, e^B$. This is easy to establish by using the fact that $Apreceq B$ implies $lambdai(A)leq lambdai(B)$ which itself follows by Courant-Fischer.

Now, we are ready. \[begin{align} log ex{e^theta epsi Ai} &~preceq~ frac{theta^2 Ai^2}{2}\;\;\; small{text{[Fact 1]}}\ sumi log ex{e^theta epsi Ai} &~preceq~ frac{theta^2 sumi Ai^2}{2}\ tr \; mathrm{exp}lef log ex{e^theta epsi Ai}right) &~leq~ tr \; mathrm{exp}left(frac{theta^2 sumi Ai^2}{2}right) \;\; \; small{text{[Fac 2]}}\ &~leq~ d\, lambdamax left(mathrm{exp}left(frac{theta^2 sumi Ai^2}{2}right)right)\ &~=~ d\, mathrm{exp}left(frac{theta^2 lambdamaxleft(sumi Ai^2right)}{2}right). end{align} \] This, gives us a variance term of $lambdamaxleft(sumi Ai^2right) $ which is better than the earlier one of $sumi lambdamaxleft(A_iright)^2$ by up to a factor of d. This matters as we have the factor of d in the exponent.

Deriving the trace inequalities

-Thompson Lie-Trotter formula says that $e^{A+B} = lim{nto infty} \left(e^{A/n}e^{B/n} \right)^{n}$. Such a formula can be used to derive the GT inequality, $tr(e^{A+B})leq tr(e^Ae^B)$. However, this inequality is false for three or more matrices. Ahlswede–Winter cleverly apply this in an iterative way by pairing this with the inequality $tr(e^A e^B) leq lambdamax(e^A)tr(e^B)$.

s Concavity Tropp's insight is that one must instead work with the cumulant generating function, $log e^{X}$. The advantage of this POV is that this approach generalizes to a much more general settings. Moreover, it gives tighter bounds.

Resources

- Tropp, Joel A. ”An Introduction to Matrix Concentration Inequalities.” arXiv.
- Lecture Notes on Ahlswede-Winter Inequality by Nicholas Harvey.
- Garg, Ankit, Lee, Yin T., Song, Zhao, and Nikhil Srivastava. ”A Matrix Expander Chernoff Bound.” arXiv.
- Talk by Joel Tropp at [Youtube Link]