

In the recent paper [?], Hsu et al. introduced a technique for obtaining tail bounds for the largest eigenvalue of matrix martingale sums that, in contrast to prior works, yields results that depend on an intrinsic measure of dimension rather than the ambient dimension of the matrices. The main contribution of this paper is a generalization of this bound to all the eigenvalues of such a sum, using the minimax technique introduced in [?].

There is a subtle shortcoming in the presentation of the results in [?]: the tail bounds provided are of the uncentered form

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_i \mathbf{X}_i \right) > t \right\} \leq C(e^t - t - 1)^{-1},$$

where the constant  $C$  depends on properties of the summand and  $t$  is constrained to be positive. Such bounds are potentially quite loose: if the largest eigenvalue of the sum concentrates around a large positive number,  $C$  will be large and this bound will be overly pessimistic. More crucially, if the largest eigenvalue concentrates around a negative number with large modulus, this tail bound tells us nothing about the behavior of the tails in the region  $t \leq 0$ . Furthermore, it seems that it is not possible to directly extend the results of [?] to apply to all eigenvalues.

For these reasons, it is more desirable to have centered tail bounds of the form

$$\mathbb{P} \left\{ \lambda_{\max} \left( \sum_i \mathbf{X}_i \right) > \mu + t \right\} \leq C(e^t - t - 1)^{-1},$$

where  $\mu$  is an ‘average’ value of the maximum eigenvalue. In this paper, we observe that Hsu et al.’s technique is in fact a matrix generalization of the classical Cramér-Chernoff method. With this viewpoint, it becomes clear that a natural choice of  $\mu$  is  $\lambda_{\max}(\sum_i \mathbb{E} \mathbf{X}_i)$ . Furthermore, using the Cramér-Chernoff formalisms, it becomes quite straightforward to use the minimax technique introduced in [?] to extend Hsu et al.’s results to bounds on the tails of all the eigenvalues of sums of independent random matrices.

## CONVEX CONJUGATION

**Theorem 1** (Conjugate inverse theorem). *Let  $\psi$  be a convex and continuously differentiable function on  $[0, b)$  with  $0 < b \leq +\infty$ . Assume that  $\psi(0) = \psi'(0) = 0$  and  $\text{range } \psi' = \mathbb{R}^+$ . Set, for each  $x \geq 0$ ,*

$$\psi^*(x) = \sup_{\eta \in (0, b)} \eta x - \psi(\eta).$$

*Then  $\psi^*$  is a nonnegative, convex, and nondecreasing function on  $\mathbb{R}_+$ . Furthermore, for any strictly positive  $t \in \text{range } \psi^*$ ,*

$$\psi^{*-1}(t) = \min_{\eta \in (0, b)} \frac{t + f(\eta)}{\eta}.$$

*Remark 1.* If  $f$  is convex and continuously differentiable, then in order for  $f^*(x)$  to be achieved it suffices that  $x \in \text{range } f'$ .

*Proof.* Let  $\eta \in (0, b)$ , then

$$f^*(x) \geq \eta x - f(\eta),$$

which implies

$$x \leq \frac{f^*(x) - f(\eta)}{\eta}.$$

Because  $f^*(x)$  is achieved, we have

$$\frac{f^*(x) + f(\eta^*)}{\eta^*} = x$$

for some  $\eta^* \in (0, b)$ . □

## THE CLASSICAL CRAMÉR-CHERNOFF METHOD

The classical Cramér-Chernoff technique allows one to bound the tails of a random variable in terms of the Cramér transform of the convex conjugate of the cumulant generating function of the random variable [?]. In this section, we develop the Cramér transform technique in a way which allows a natural generalization to the matrix case in view of Hsu et al.’s main result.

We now introduce the Cramér-Chernoff technique. Let  $X$  be a random variable and  $t, \eta > 0$ . By the monotonicity of the scalar exponential function and Markov's inequality, we have

$$\begin{aligned}\mathbb{P}\{\eta X - \psi_X(\eta) \geq t\} &= \mathbb{P}\{\exp(\eta X - \psi_X(\eta)) \geq e^t\} \\ &\leq e^{-t} \mathbb{E}(e^{\eta X}) e^{-\psi_X(\eta)} = e^{-t} (\mathbb{E} e^{\eta X}) (\mathbb{E} e^{\eta X})^{-1} = e^{-t}.\end{aligned}\quad (1)$$

If  $X = \sum_i X_i$  is a sum of independent random variables, then the cumulant generating function of  $X$  is the sum of cumulant generating functions of the  $X_i$ , so we have

$$\mathbb{P}\left\{\eta \sum_i X_i - \sum_i \log \mathbb{E} e^{\eta X_i} \geq t\right\} \leq e^{-t}.\quad (2)$$

**Proposition 1** (Cramér transform inversion). *Let  $\psi$  be some convex and continuously differentiable function on  $[0, b)$  with  $0 < b \leq +\infty$ . Assume that  $\psi(0) = \psi'(0) = 0$  and set, for every  $x \geq 0$ ,*

$$\psi^*(x) := \sup_{\eta \in (0, b)} (\eta x - \psi(\eta)).$$

*Then  $\psi^*$  is a nonnegative convex and nondecreasing function on  $\mathbb{R}_+$ . Moreover, for every  $t \geq 0$ , the set  $\{x \geq 0 \mid \psi^*(x) > t\}$  is nonvoid and the generalized inverse of  $\psi^*$  at a point  $t$ , defined by*

$$\psi^{*-1}(t) = \inf\{x \geq 0 \mid \psi^*(x) > t\},$$

*can also be written as*

$$\psi^{*-1}(t) = \inf_{\eta \in (0, b)} \left\lceil \frac{t + \psi(\eta)}{\eta} \right\rceil.$$

Let  $X$  be a centered random variable and assume there is an interval  $I = [0, b)$  over which  $\psi_X$  is well-defined. Then, by (1), for any  $\eta \in I$ ,

$$\mathbb{P}\{\eta X - \psi_X(\eta) \geq t\} = \mathbb{P}\left\{X \geq \frac{t + \psi_X(\eta)}{\eta}\right\} \leq e^{-t}.$$

Optimize over  $\eta \in I$  to observe that

$$\mathbb{P}\{X \geq \psi_X^{*-1}(t)\} = \mathbb{P}\left\{X \geq \inf_{\eta \in (0, b)} \frac{t + \psi_X(\eta)}{\eta}\right\} \leq e^{-t}.$$

By definition, for any  $t \geq 0$ , we have  $\psi_X^{*-1}(\psi_X^*(t)) \leq t$ , so we conclude that the following tail bound holds:

$$\mathbb{P}\{X \geq t\} \leq e^{-\psi_X^*(t)}.$$

## 1. MINIMAX CRAMÉR-CHERNOFF TECHNIQUE

The following proposition is the main result of [?].

**Proposition 2.** *Let  $\{\mathbf{X}_i\}$  be a finite sequence of independent, random, self-adjoint matrices. Then for any  $\eta \in \mathbb{R}$  and any  $t > 0$ ,*

$$\begin{aligned}\mathbb{P}\left\{\lambda_{\max}\left(\eta \sum_i \mathbf{X}_i - \sum_i \log \mathbb{E}[\exp(\eta \mathbf{X}_i)]\right) \geq t\right\} \\ \leq \text{tr} \mathbb{E}\left[-\eta \sum_i \mathbf{X}_i + \sum_i \log \mathbb{E}[\exp(\eta \mathbf{X}_i)]\right] \cdot (e^t - t - 1)^{-1}.\end{aligned}$$

Clearly this result is a generalization of (1), at the following costs: the tail decay is slightly slower,  $(e^{-t} - t - 1)^{-1}$  versus  $e^{-t}$  in the scalar case, and the noncommutativity of the summands introduces a constant (the trace term) not present in the scalar case. By analogy with the tail bounds obtained using the matrix Laplace transform method, we view this constant as a measure of the dimensionality of  $\sum_i \mathbf{X}_i$ .

Using 2 and 1, we extend the Cramér-Chernoff method to the case where the summands are matrices.

**Theorem 2** (Matrix Cramér-Chernoff bound). *Let  $\mathbf{X} = \sum_i \mathbf{X}_i$  be a sum of independent random matrices and take*

$$\Psi_{\mathbf{X}}(\eta) := \sum_i \log \mathbb{E} e^{\eta(\mathbf{X}_i - \mathbb{E}\mathbf{X}_i)}.$$

*Let  $\Phi \geq \lambda_{\max}(\Psi_{\mathbf{X}})$  be a function satisfying the hypotheses of Proposition 1. Select  $\eta^*$  to satisfy*

$$\Phi^{*-1}(\Phi^*(t)) = \frac{t + \Phi(\eta^*)}{\eta^*}.$$

*Then*

$$\mathbb{P}\{\lambda_{\max}(\mathbf{X}) \geq t + \lambda_{\max}(\mathbb{E}\mathbf{X})\} \leq (\text{tr } \Psi_{\mathbf{X}}(\eta^*)) \cdot (e^{\Phi^*(t)} - \Phi^*(t) - 1)^{-1}.$$

*Proof.* We begin by using the subadditivity of the maximum eigenvalue mapping to center  $\mathbf{X}$ :

$$\mathbb{P}\{\lambda_{\max}(\mathbf{X}) \geq t + \lambda_{\max}(\mathbb{E}\mathbf{X})\} \leq \mathbb{P}\{\lambda_{\max}(\mathbf{X} - \mathbb{E}\mathbf{X}) \geq t\}.$$

It follows that we can assume without loss that the summands are centered.

Let  $\eta > 0$  and notice that by Proposition 2,

$$\begin{aligned} \mathbb{P}\{\lambda_{\max}(\eta\mathbf{X}) \geq t + \lambda_{\max}(\Psi_{\mathbf{X}}(\eta))\} &\leq \mathbb{P}\{\lambda_{\max}(\eta\mathbf{X} - \Psi_{\mathbf{X}}(\eta)) \geq t\} \\ &\leq \text{tr } \mathbb{E}[-\eta\mathbf{X} + \Psi_{\mathbf{X}}(\eta)] \cdot (e^t - t - 1)^{-1} \\ &= \text{tr } \Psi_{\mathbf{X}}(\eta) \cdot (e^t - t - 1)^{-1}. \end{aligned}$$

Since  $\Phi \geq \lambda_{\max}(\Psi_{\mathbf{X}})$ , it follows that

$$\mathbb{P}\left\{\lambda_{\max}(\mathbf{X}) \geq \frac{t + \Phi(\eta)}{\eta}\right\} \leq \text{tr } \Psi_{\mathbf{X}}(\eta) \cdot (e^t - t - 1)^{-1}.$$

Optimizing over  $\eta$  and applying Proposition 1, we obtain

$$\mathbb{P}\{\lambda_{\max}(\mathbf{X}) \geq \Phi^{*-1}(t)\} \leq \text{tr } \Psi_{\mathbf{X}}(\eta^*) \cdot (e^t - t - 1)^{-1}.$$

Using the functional relation  $\Phi^{*-1}(\Phi^*(t)) = t$ , we conclude with the bound

$$\mathbb{P}\{\lambda_{\max}(\mathbf{X}) \geq t\} \leq \text{tr } \Psi_{\mathbf{X}}(\eta^*) \cdot (e^{\Phi^*(t)} - \Phi^*(t) - 1)^{-1}.$$

□