Ideas towards operator norm bounds

Cole Franks, Rafael Oliveira, Akshay Ramachandran, Michael Walter

February 2020

Contents

1 Numerics ideas

What's going on in the kglasso stuff? They are generating an erdos renyi and making sure that the least eigenvalue is 1, then trying to approximate each. I wonder how it'd do without scaling the eigenvalues into oblivion?

What's easiest? Convert to matlab or nah?

2 Tensor Scaling

We will maintain similar notation. We have n samples of $X \sim \mathcal{N}(0, \frac{1}{n} \otimes_a \frac{1}{d_a} I_a)$ with $D := \prod_a d_a$. We don't have a KLR style analysis at the moment, but strong convexity is enough by the FM analysis, and this can be proven by just controlling each bipartite piece. So the operator scaling analysis does give us very good bounds for $\|\mu\|_{op}$ and expansion with $nD \gg \max_a d_a^2 \log^c(D)$. These bounds are not enough though, so in this section we will follow the FM analysis to give the requirements, then show the required strong convexity, and show how to maintain this under perturbation.

2.1 FM Analysis

Recall that $\forall a : Q^a \approx \frac{1}{d_a} I_a$, so if we can show $\forall a \neq b : \langle Q^{ab}, Z \otimes Y \rangle \lesssim \frac{\|Z\|_F \|Y\|_F}{\sqrt{d_a d_b}}$ then we have strong convexity with $\langle Z, \nabla^2 Z \rangle \gtrsim \sum_a \frac{\|Z_a\|_F^2}{d_a}$, i.e. the Hessian is strongly diagonally dominant. We will derive our requirements on strong convexity, perturbation bounds, and initial error. Assume we have the following strong convexity

$$\forall Z : \sum_a \frac{\|Z_a\|_F^2}{d_a} \leq \kappa^2 : \langle Y, \nabla^2_{e^Z}, Y \rangle \geq \lambda \sum_a \frac{\|Y_a\|_F^2}{d_a}$$

Choose X to be the geodesic towards the optimum and $g(t) := f(e^{tZ})$ with the opt at $t = 1$:

$$g'(1) = \int_0^1 g''(t) + g'(0) \geq \lambda \sum_a \frac{\|Z\|_F^2}{d_a} - |\langle \nabla_a, Z_a \rangle|$$

$$\geq \sum_a \frac{\|Z_a\|_F}{\sqrt{d_a}} \left(\lambda \frac{\|Z_a\|_F}{\sqrt{d_a}} - \sqrt{d_a} \|\nabla_a\|_F \right)$$

$$\geq \sqrt{\sum_a \frac{\|Z_a\|_F^2}{d_a}} \left(\lambda \sqrt{\sum_a \frac{\|Z_a\|_F^2}{d_a}} - \sqrt{\sum_a d_a \|\nabla_a\|_F^2} \right)$$

This is > 0 if $\forall a : \lambda > d_a \frac{\|\nabla_a\|_F}{\|Z_a\|_F}$ or $\lambda^2 > \frac{\sum_a d_a \|\nabla_a\|_F^2}{\sum_a d_a^{-1} \|Z_a\|_F^2}$.

Since standard perturbation bounds ($e^Z \approx I + Z$) only work for small Z, we will require

$$\forall a : \|\nabla_a\|_F \ll \frac{1}{d_a}$$

$$\forall a : \|Z_a\|_F^2 \ll 1 \implies \langle Y, \nabla^2_{e^z}, Y\rangle \geq \Omega(1) \sum_a \frac{\|Y_a\|_F^2}{d_a}$$

2.2 Moment Map

For $g \sim \frac{1}{nD} \otimes_a I_a$, we want to bound $\|Q^a - sI_a\|_F$ using a net:

$$\mathbb{E}\langle \sum_t g_t g_t^*, X_a\rangle = \sum_i x_i \chi(\frac{1}{d_a}, \frac{TD}{d_a}) = \langle \frac{1}{d_a} I_a, X\rangle = 0$$

$$\begin{aligned}
\log \mathbb{E} \exp \theta \langle \sum_t g_t g_t^*, X_a\rangle &= \log \mathbb{E} \exp \theta \sum_i x_i \chi(\frac{1}{d_a}, \frac{TD}{d_a}) \\
&= \sum_i \frac{-TD}{2d_a} \log\left(1 - 2\theta \frac{x_i}{TD}\right) \\
&\lesssim \theta^2 \frac{\|X\|_F^2}{2d_a TD} \qquad \forall \theta < \left(\frac{\|X\|_{op}}{TD}\right)^{-1}
\end{aligned}$$

$$\implies \mathbb{P}[\langle \sum_t g_t g_t^*, X_a\rangle \geq \epsilon \|X\|_F] \leq \begin{cases} \exp(-\Omega(\epsilon^2 TD d_a)) & \epsilon < \frac{\|X\|_F}{d_a \|X\|_{op}} \\ \exp(-\Omega(\epsilon TD)\frac{\|X\|_F}{\|X\|_{op}}) & \epsilon \geq \frac{\|X\|_F}{d_a \|X\|_{op}} \end{cases}$$

We will need the following settings of ϵ in future:

$$\epsilon \approx \frac{1}{\sqrt{d_a}} \implies \mathbb{P}[d_a \|Q^a - sI_a\|_F^2 \gtrsim c] \leq \exp(d_a^2 \log d_a - c\frac{TD}{\sqrt{d_a}})$$

For which we need $TD \gtrsim \max_a d_a^{5/2} \log d_a$.

$$\epsilon \approx \frac{1}{d_a} \implies \mathbb{P}[d_a \|Q^a - sI_a\|_F^2 \gtrsim \frac{c}{d_a}] \leq \exp(d_a^2 \log d_a - c^2 \frac{TD}{d_a})$$

For which we need $TD \gtrsim \max_a d_a^3 \log d_a$.

Remark 1. *Note we lose out on the subgaussian part of the bound only when $\frac{d_a\|X\|_{op}^2}{\|X\|_F^2}$ is large. It is quite possible that for our setting, we can bound e.g. the condition number or stable rank of is small w.h.p. In particular if we can show the only relevant part of the net has $s\|X\|_F^2 \geq d\|X\|_{op}$ then we only incur a $\sqrt{s}$ loss in required samples.*

Actually there seems to be a simpler way to prove these statements using the $\|\cdot\|_{op}$ bounds derived earlier.

$$\|Q^a - sI_a\|_{op} \leq \frac{\sqrt{f(d)}}{d} \implies d_a \|Q^a - sI_a\|_F^2 \leq f(d)$$

So this means in the first case, we need $TD \gtrsim \max_a d_a^2 \log D$ and the second case we need $TD \gtrsim \max_a d_a^3 \log D$. But by this analysis we only get $1/poly$ failure probability.

2.3 Net proof of Expansion

Recall again we have $Q := \sum_{i\in[n]} X_i X_i^*$ with i.i.d $X \sim \mathcal{N}(0, \frac{1}{n} \otimes_a \frac{1}{d_a} I_a)$. For symmetric test matrices Z, Y with eigenvalues $\{z_i\}, \{y_j\}$ respectively:

$$\mathbb{E}\langle Q, Z_a \otimes Y_b\rangle = \sum_{ij} z_i y_j \chi(\frac{1}{d_a d_b}, \frac{nD}{d_a d_b}) = \langle \frac{1}{d_a} I_a, Z\rangle\langle \frac{1}{d_b} I_b, Y\rangle = 0$$

$$\begin{aligned}\log \mathbb{E} \exp \theta\langle Q, Z_a \otimes Y_b\rangle &= \log \mathbb{E} \exp \theta \sum_{ij} z_i y_j \chi(\frac{1}{d_a d_b}, \frac{nD}{d_a d_b}) \\ &= \sum_{ij} \frac{-nD}{2 d_a d_b} \log\left(1 - 2\theta \frac{z_i y_j}{nD}\right) \\ &\lesssim \theta^2 \frac{\|Z\|_F^2 \|Y\|_F^2}{2 d_a d_b nD} \qquad \forall \theta < \left(\frac{\|Z\|_{op}\|Y\|_{op}}{TD}\right)^{-1}\end{aligned}$$

$$\mathbb{P}[\langle Q, X \otimes Y\rangle \geq \lambda \frac{\|Z\|_F \|Y\|_F}{\sqrt{d_a d_b}}] \leq \begin{cases} \exp(-\lambda^2 nD) & \lambda < \frac{\|Z\|_F\|Y\|_F}{\sqrt{d_a d_b}\|Z\|_{op}\|Y\|_{op}} \\ \exp(-\lambda nD \frac{\|Z\|_F\|Y\|_F}{\sqrt{d_a d_b}\|Z\|_{op}\|Y\|_{op}}) \end{cases}$$

So for $\lambda \approx \frac{1}{k}$ we need $nD \gtrsim \max_a d_a^3 \log D > \max_{a,b} \sqrt{d_a d_b}(d_a^2 + d_b^2) \log D$

Remark 2. *Note again we lose out when $\frac{d\|X\|_{op}^2}{\|X\|_F^2}$ is large. So we would like to show that w.h.p. the singular vectors of our bipartite operator have e.g. small condition number or large stable rank. Again if we can show the relevant part of the net has $s\|X\|_F^2 \geq d\|X\|_{op}^2$ then we incur an s factor loss in samples. This is reminiscent of the fact that eigenvectors for random matrices have many delocalization properties, so will look into that.*

2.4 Delocalization

I have a few claims which, if true, would give another proof (along with Pisier's) of constant expansion at the start with $1/poly$ failure probability. Unfortunately to make this robust we would either need a robust form of delocalization, which I think is false in general, or exponential failure probability, which again may be false in general.

Recall M^{ab} is the operator defined by the off-diagonal block of the Hessian. For every choice of bases (torus) U, V we have a matrix M_{UV}^{ab} where each entry is $\chi(\frac{1}{d_a d_b}, \frac{TD}{d_a d_b})$. In particular each entry is of exponential type.

Claim 1. *Let U, V be the (random) basis for the optimizers of $\|M\|_{op}$. Then (U, V) and M_{UV} are distributionally independent.*

Theorem 2 (Informal)**.** *For $M \in \mathbb{R}^{n\times n}$ populated by iid variables X such that $\mathbb{E}X = 0, \mathbb{E}X^2 = 1$ and X is of exponential type, we have delocalization of eigenvectors with $1/poly$ failure probability:*

$$v \in S^{n-1}, \|v\|_\infty \gtrsim \frac{\log^c n}{\sqrt{n}} \implies \forall \lambda \in \mathbb{C} : \|Mv - \lambda v\|_2 \gtrsim \frac{1}{\sqrt{n}}$$

Claim 3. *The above informal claim can be extended to singular vectors.*

Corollary 4. *Conditioned on the above delocalization event, our net only has to cover* $\{X \mid d\|X\|_{op}^2 \lesssim \log^c(d)\|X\|_F^2$, *and so we only require* $TD \gtrsim d_a^2 \log^c D$.

2.5 Robust proof of Expansion

In this section we will show expansion under perturbations of the form $\otimes_a e^{\delta_a}$. Note if $\|\delta\|_{op} \ll 1$ then we can approximate $e^{\delta} - I \approx O(\delta)$.

$$\mu := \mathbb{E}\langle Q, e^{\delta_a} Z e^{\delta_a} \otimes e^{\delta_b} Y e^{\delta_b} \otimes_c e^{2\delta_c} \rangle = \langle \frac{e^{2\delta_a}}{d_a}, Z \rangle \langle \frac{e^{2\delta_b}}{d_b}, Y \rangle \prod_{c \neq a,b} Tr[\frac{e^{2\delta_c}}{d_c}]$$

$$\log \mathbb{E} \exp \theta \langle Q, e^{\delta_a} Z e^{\delta_a} \otimes e^{\delta_b} Y e^{\delta_b} \otimes_{c \neq a,b} e^{2\delta_c} \rangle = \log \mathbb{E} \exp \theta \sum_{\boldsymbol{i}} x^{\boldsymbol{i}} \chi(\frac{1}{D}, T)$$

$$\leq \theta\mu + \theta^2 \frac{\|e^{\delta_a} Z e^{\delta_a}\|_F^2 \|e^{\delta_b} Y e^{\delta_b}\|_F^2 \prod_c \|e^{2\delta_c}\|_F^2}{2TD^2}$$

$$\forall \theta < \left(\frac{\|e^{\delta_a} Z e^{\delta_a}\|_{op} \|e^{\delta_b} Y e^{\delta_b}\|_{op} \prod_c \|e^{2\delta_c}\|_{op}}{TD} \right)^{-1}$$

Lemma 5. *For* $\|\delta\|_{op} \ll 1$

$$\frac{1}{d} Tr[e^{2\delta}] \leq 1 + O(\|\delta\|_{op})$$

$$\langle \frac{1}{d} I, e^{\delta} Z e^{\delta} - Z \rangle \leq \frac{O(\|\delta\|_{op})}{\sqrt{d}} \|Z\|_F$$

$$\|e^{\delta} Z e^{\delta}\|_F^2 \leq (1 + O(\|\delta\|_{op})) \|Z\|_F^2$$

Proof. Let $\delta := e^{\delta'} - I$ and note $\|\delta\|_{op} \approx \|\delta'\|_{op}$ for small δ'. So we bound

$$Tr[e^{2\delta'} - I] = \langle I, e^{2\delta'} - I \rangle \leq d\|2\delta + \delta^2\|_{op} \leq dO(\|\delta\|_{op})$$

For the second line we use Cauchy Schwarz $\|Z\|_1 \leq \sqrt{d}\|Z\|_F$

$$\langle I, (I+\delta) Z (I + \delta') - Z \rangle = \langle 2\delta + \delta^2, Z \rangle \leq O(\|\delta\|_{op}) \|X\|_1 \leq O(\|\delta\|_{op}) \sqrt{d} \|Z\|_F$$

The second line is similar but simpler using $\|AB\|_F \leq \|A\|_{op} \|B\|_F$ ☐

So using the above lemma, we have bounds for $\langle Z, I_a \rangle = \langle Y, I_b \rangle = 0$:

$$\mu \leq \frac{\|Z\|_F \|Y\|_F}{\sqrt{d_a d_b}} O(\|\delta_a\|_{op} \|\delta_b\|_{op}) \prod_c (1 + O(\|\delta_c\|_{op})$$

$$\log \mathbb{E} \exp \theta (... - \mu) \leq \theta^2 \frac{\|Z\|_F^2 \|Y\|_F^2}{2TD} \prod_c (1 + O(\|\delta_c\|_{op}))$$

$$\forall \theta < \left(\frac{\|Z\|_{op}\|Y\|_{op}}{TD}\right)^{-1} \prod_c (1 - O(\|\delta_c\|_{op}))$$

So basically, as long as $\|\delta\|_{op} \ll 1$ everything is of the same order as in the unperturbed case, and therefore if we run a net on all parts simultaneously (of size $\exp(\sum_a d_a^2)$) we get roughly the same probabilistic bounds as the start.

This creates a bottleneck though as in general the inequality $\|X\|_{op} \leq \|X\|_F$ could be tight, so in order to guarantee a small enough perturbation we can only move in the ball $\|X\|_F \ll 1$. This is why we wrote out the conditions required for $\|\nabla_a\|_F \ll \frac{1}{d_a}$, as therefore we can assume $\|Z_a\|_F \ll 1$ and require $\lambda \approx 1$.

Remark 3. *Here is where we definitely would like a KLR style analysis to exploit the fact that we actually have robustness in* $\|\cdot\|_{op}$.

2.6 Deterministic Robust Proof of Expansion

We follow Michael's write-up in our setting. We can assume we have sufficiently strong expansion initially and we would like to show that any scaling ($e^Y \cdot g$ with $\forall a : \|Y_a\|_F \ll 1$) only changes the Hessian by a small amount.

Fact 6. *Recall the formula for Hessian at* v*:*

$$\langle X, \nabla_v^2, X\rangle = \langle \frac{vv^*}{\|v\|_2^2}, \left(\sum_a X_a \otimes I_{\overline{a}}\right)^2 \rangle - \left(\sum_a \langle \frac{vv^*}{\|v\|_2^2}, X_a \otimes I_{\overline{a}}\rangle\right)^2$$

Initially, we have expansion of the form

$$\langle X, \nabla^2, X\rangle \geq (1-\epsilon)\sum_a \frac{\|X_a\|_F^2}{d_a} - \lambda \sum_{a\neq b} \frac{\|X_a\|_F\|X_b\|_F}{\sqrt{d_a d_b}} \geq (1-\epsilon')\sum_a \frac{\|X_a\|_F^2}{d_a}$$

So for how large of a perturbation can we show that every quadratic form changes $<< \sum_a \frac{\|X_a\|_F^2}{d_a} =: \|X\|^2$?

Lemma 7.

$$\|\sum_a X_a \otimes I_{\overline{a}}\|_{op} \leq \sum_a \|X_a\|_{op} \leq \sum_a \|X_a\|_F \leq \sqrt{\left(\sum_a d_a\right)\left(\sum_a \frac{\|X_a\|_F^2}{d_a}\right)}$$

All inequalities can be tight, say at $X_a = \sqrt{d_a}E_{11}$*. In Michael's notation, for standard tensor rep*

$$\sup_X \frac{\|\Pi(X)\|_{op}}{\sum_a \|X_a\|_F} = k \ ; \qquad \sup_X \frac{\|\Pi(X)\|_{op}^2}{\|X\|^2} = \sum_a d_a$$

Theorem 8. *For* $\|v\|_2 = 1$ *and* $w := e^Y \cdot v/\|e^Y \cdot v\|_2$ *with perturbation* $\|\sum_a Y_a \otimes I_{\overline{a}}\|_{op} = \|\Pi(Y)\|_{op} \ll 1$*:*

$$|\langle X, \nabla_w^2, X\rangle - \langle X, \nabla_v^2, X\rangle| \leq O(\sum_a d_a)\|X\|^2\|\Pi(Y)\|_{op}$$

Proof. Assume wlog $\sum_a \frac{\|X_a\|_F^2}{d_a} = 1$:

$$|\langle X, \nabla_v^2, X\rangle - \langle X, \nabla_w^2, X\rangle|$$

$$\le |\langle ww^* - vv^*, \left(\sum_a X_a\right)^2\rangle| + |\langle ww^*, \sum_a X_a\rangle^2 - \langle vv^*, \sum_a X_a\rangle^2|$$

$$= |\langle ww^* - vv^*, \Pi(X)^2\rangle| + |\langle ww^* + vv^*, \Pi(X)\rangle\langle ww^* - vv^*, \Pi(X)\rangle|$$

$$\le \|ww^* - vv^*\|_1 \|\Pi(X)\|_{op}^2 + 2\|\Pi(X)\|_{op}\|ww^* - vv^*\|_1\|\Pi(X)\|_{op}$$

$$= 3(\sum_a d_a)\|ww - vv^*\|_1$$

Now we just have to bound perturbations of rank one matrices:

$$\|ww^* - vv^*\|_1 \le \sqrt{2}\|ww^* - vv^*\|_F \le 2\sqrt{2}\|w - v\|_2$$

Finally we can just bound the action of Y:

$$\|w - v\|_2 \le \|\frac{e^Y \cdot v}{\|e^Y \cdot v\|_2} - e^Y \cdot v\|_2 + \|e^Y \cdot v - v\|_2$$

$$\le |1 - \|e^Y \cdot v\|_2|\|w\|_2 + \|e^Y - I\|_{op}\|v\|_2 \le 2\|e^Y - I\|_{op} \le O(1)\|\Pi(Y)\|_{op}$$

□

Claim 9. *For unit vectors $v, w : \|vv^* - ww^*\|_F^2 \le 2\|v - w\|_2^2$*

Proof. Assume wlog $v = e_1, w = (x, y) \in \mathbb{R}^2$ and note

$$vv^* - ww^* = \begin{pmatrix} 1 - x^2 & -xy \\ -xy & -y^2 \end{pmatrix}$$

has $Tr = 1 - x^2 - y^2 = 0$ and $\det = (1 - x^2)(-y^2) - x^2y^2 = -y^2$; so the eigenvalues are $\pm y$.

$$\implies \|vv^* - ww^*\|_F^2 = 2y^2 = 2(1 - x^2)$$

$$\|v - w\|_2^2 = (1 - x)^2 + y^2 = 2(1 - x)$$

Finally note $1 - x^2$ is concave and $2(1 - x)$ is a tangent line at $x = 1$. □

Recall the sufficient condition for $\lambda = \Omega(1)$-expansion

$$\sum_a d_a\|\nabla_a\|_F^2 \lesssim \sum_a d_a^{-1}\|Y_a\|_F^2$$

So we require $\langle X, \nabla_{e^Y}, X\rangle \gtrsim \|X\|^2$. We can get this bound for all $\|\Pi(Y)\|_{op} \ll (\sum_a d_a)^{-1}$. Therefore we can get a proper perturbation bound for $\forall a : \|Y_a\|_F \ll (\sum_a d_a)^{-1}$ which implies $\|Y\|^2 \ll (\max_a d_a)^{-3}$, so we require $\forall a : \|\nabla_a\|_F \ll (\max_a d_a)^{-2}$. This is worse than the requirement from the larger net argument above.

[CF: cole agrees - needs $D > d_a^5$!]

2.7 Better Deterministic Robustness

Lemma 10. *For perturbation $v \to \otimes_a e^{\delta_a} \cdot v =: w$ where $\forall a : \|\delta_a\|_{op} \ll 1$, and let $\{\sigma_1^{ab}, \sigma_2^{ab}\}$ be the matrix $\|\cdot\|_F \to \|\cdot\|_F$ norm and matrix norm on subspace $\perp$ to (I, I) for each bipartite part respectively:*

$$\forall a, b : \sigma_2^{ab}(ww^*) - \sigma_2^{ab}(vv^*) \leq O\left(\sum_a \|\delta_a\|_{op}\right) \sigma_1^{ab}(vv^*)$$

The same is true for the diagonal blocks.

Proof. To lower bound the diagonal block, we just need a spectral lower bound on $\{Q_a\}$, since $\langle vec(X), M^a(vec(X))\rangle := \langle Q_a, X^2\rangle$.

$$\|e^{\delta_a} Q_a e^{\delta_a} - Q_a\|_{op} =\leq O(\|\delta_a\|_{op})\|Q_a\|_{op}$$

Now we address a perturbation on $b \neq a$. For a spectral lower bound, we choose test $Z \succeq 0$ and let $\delta := e^{2\delta_b} - I$:

$$\langle e^{\delta_b} vv^* e^{\delta_b} - vv^*, I_{\overline{a}} \otimes Z_a\rangle = \langle vv^*, \delta \otimes Z\rangle = \langle Z, V^* \delta V\rangle$$

Here $V \in \mathbb{R}^{d_b \times d_a}$ is the matricized version of v. But now since $Z \succeq 0$, the argument is clear

$$\leq \langle Z, V^* |\delta| V\rangle \leq \|\delta\|_{op} \langle Z, V^* I V\rangle = \|\delta\|_{op} \langle vv^*, I_{\overline{a}} \otimes Z\rangle$$

The argument for the off-diagonal blocks is similar. We first argue the change of σ^{ab} is small under perturbations where $\forall c \neq a, b : \delta_c = 0$. Let M^{ab} be the matrix versions of the bipartite operators:

$$\langle vec(Y), M_v^{ab}(vec(Z))\rangle := \langle vv^*, I_{\overline{ab}} \otimes Z \otimes Y\rangle$$

$$\langle vec(Y), M_w^{ab}(vec(Z))\rangle := \langle ww^*, I_{\overline{ab}} \otimes Z \otimes Y\rangle$$

$$\implies M_w = (e^{\delta_b} \otimes e^{\delta_b}) M_v (e^{\delta_a} \otimes e^{\delta_a})$$
$$\implies \|M_w - M_v\|_{op} \leq O(\|\delta_a\|_{op} + \|\delta_b\|_{op})\|M_v\|_{op}$$

where in the last step we used $\delta \ll 1$.

The more difficult part of the argument to see (at least for me) is how σ^{ab} changes if some other part $c \neq a, b$ is changed, i.e. $\forall d \neq c : \delta_d = 0$. First we define $\delta := e^{2\delta_c} - I$, and test vectors Z, Y:

$$\langle ww^* - vv^*, I_{\overline{ab}} \otimes Z \otimes Y\rangle = \langle vv^*, \delta \otimes Z \otimes Y\rangle = \langle Z \otimes Y, V^* \delta V\rangle$$

Here $V \in \mathbb{R}^{d_c \times d_a d_b}$ is the matricized version of v, i.e. the k-th element of ij-th column is $(V_{ij})_k := v_{ijk}$. Now in order to use our operator norm bounds, we need to deal with cancelations, so we split into positive and negative parts $Z := Z_+ - Z_-, Y := Y_+ - Y_-$:

$$|\langle Z \otimes Y, V^* \delta V\rangle| \leq |\langle Z_\pm \otimes Y_\pm, V^* \delta V\rangle|$$

Now we analyze one of these terms (by abuse of notation $Z, Y \succ 0$):

$$\leq \langle Z \otimes Y, V^* |\delta| V \rangle \leq \|\delta\|_{op} \langle Z \otimes Y, V^* V \rangle = \|\delta\|_{op} \langle vv^*, I_{\overline{ab}} \otimes Z \otimes Y \rangle$$

Each of these terms we can bound by $\sigma_1^{ab} \|Z\|_F \|Y\|_F$. (Note we can save a 2-factor on these four terms since they are decompositions of Z, Y). So by iterating this argument over all c, we get the desired bound. $\square$

Remark 4. *Instead of F, we could use any pair of dual norms and get the same result. In particular, we will use these results to bound the $1 \to 1$ and $op \to op$ norms of the channels. Explicitly, if we redefine $\sigma := \|\Phi\|_{p \to p}$ and (p, q) are Holder conjuguates:*

$$\langle vv^*, \delta \otimes Z \otimes Y \rangle \leq \|\delta\|_{op} \langle vv^*, I_{\overline{ab}} \otimes Z \otimes Y \rangle \leq \sigma \|\delta\|_{op} \|Z_{\pm}\|_p \|Y_{\pm}\|_q$$

$$\leq \sigma \|\delta\|_{op} (2\|Z_+\|_p^p + 2\|Z_-\|_p^p)^{1/p} (2\|Y_+\|_q^q + 2\|Y_-\|_q^q)^{1/q} = 2\sigma \|\delta\|_{op} \|Z\|_p \|Y\|_q$$

2.8 Conclusion

At the end of the day we require $nD \gtrsim \max_a d_a^3 \log D$ to get small enough ∇ and robust expansion. Note this is $d \log D$ away from the existence/ uniqueness threshold for the solution.

2.9 Ideas for KLR style analysis

1. The delocalization type arguments may be helpful since we know the movement is spread out.
2. Follow gradient flow and show we have strong convergence for $\Omega(\log \max_a d_a)$ time, after which the operator norm of the gradient can never be bigger than the initial one. Therefore if we have the first point, it is enough to cover the $O(1)$-size operator norm ball.
3. Even if we get that perturbations come from $O(1)$-size operator norm ball, without some improved delocalization (or something) we would still need an extra d factor.
4. One important lemma in KLR is that the operator norm of moment map is roughly monotone decreasing over all time (where the fudge is of the order of the change in objective function $\|v\|_2^2$ over time). This is not true for $k = 3$. But we have some extra structure e.g.: the error of a row and column are positively correlated.
5. Let $\epsilon(t) := \max_a d_a \|\nabla_a(t)\|_{op}$. Maybe we can show

   $$T_\epsilon := \inf\{t \mid \epsilon(t) > 100t\epsilon(0)\}$$

   is large enough. In particular by the argument above T_ϵ cannot be larger than $\log d$. Maybe there's some contradiction whereby we can always increase T_ϵ.

Recall the gradient flow is

$$\partial_t v(t) = \sum_a I_{\overline{a}} \otimes X_a \cdot v(t)$$

where $X_a := d_a \mu_a(t)$. Then we have

$$\partial_t \sum_a d_a \|\mu_a(t)\|_F^2 = \sum_a d_a^2 \langle Q^a, \mu_a^2 \rangle + \sum_{a \neq b} d_a d_b \langle Q^{ab}, \mu_a \otimes \mu_b \rangle$$

$$\geq (1-\epsilon) \sum_a d_a^2 \frac{\|\mu_a\|_F^2}{d_a} - \lambda \sum_{a \neq b} d_a d_b \frac{\|\mu_a\|_F \|\mu_b\|_F}{\sqrt{d_a d_b}}$$

$$\geq (1-\epsilon') \sum_a d_a \|\mu_a\|_F^2$$

2.10 Incoherence

Definition 1. *For two orthonormal bases $\{u_i\}, \{v_i\} \subseteq \mathbb{R}^d$, the coherence between them is*

$$\gamma(U, V) := \sqrt{d} \max_{i,j} |\langle u_i, v_j \rangle|$$

Let $(\epsilon/d_a, u)$ be the eigenpair associated with operator norm of μ_a and recall that under gradient flow we move in direction $X := -\sum_a e_a \otimes d_a \mu_a\}$

$$-\partial_{t=0} \|\mu_a\|_{op} = -\langle e_a \otimes uu^*, \nabla^2 X \rangle$$

$$= \langle \rho^{\{a\}}, d_a \mu_a uu^* \rangle + \sum_{b \neq a} \langle \rho^{\{a,b\}}, uu^* \otimes d_b \mu_b \rangle$$

$$= \frac{1}{d_a}(1+\varepsilon) d_a \frac{1}{d_a} \varepsilon + \sum_{b \neq a} d_b \langle \rho^{\{a,b\}}, uu^* \otimes \mu_b \rangle$$

Let's focus on a single (a, b) term, and denote $T : L(\mathbb{R}^{d_b}) \to L(\mathbb{R}^{d_a})$ as

$$\langle T(Y_b), X_a \rangle := \langle \rho^{\{a,b\}}, X_a \otimes Y_b \rangle$$

So the term we are analyzing is $\langle T(\mu_b), uu^* \rangle$, which can be viewed as a diagonal of the matrix $T(\mu_b)$. We have good bounds on this matrix:

$$\langle I_a, T(\mu_b) \rangle = \langle \rho^{\{a,b\}}, I_a \otimes \mu_b \rangle = \|\mu_b\|_F^2$$

Now define $Y := T(\mu_b) - \frac{\|\mu_b\|_F^2}{d_a} I_a$ to be projection orthogonal to the identity:

$$\|Y\|_F = \sup_X \frac{\langle Y, X \rangle}{\|X\|_F} = \sup_{X \perp I} \frac{\langle Y, X \rangle}{\|X\|_F} = \sup_{X \perp I, \|X\|_F = 1} \langle \rho^{\{a,b\}}, X \otimes \mu_b \rangle \leq \frac{\lambda}{\sqrt{d_a d_b}} \|\mu_b\|_F$$

Now if $Y = \sum_i y_i v_i v_i^*$ is the eigendecomposition, we have a bound on the variance of the eigenvalues. Let $\mu_a = \sum_i x_i u_i u_i^*$, and if we have a bound on the coherence of U, V, we can bound the competing term:

$$d_b \langle \rho^{\{a,b\}}, uu^* \otimes \mu_b \rangle = d_b \langle \rho^{\{a,b\}}, \frac{1}{d_a} I_a \otimes \mu_b + (uu^* - \frac{1}{d_a} I_a) \otimes \mu_b \rangle$$

$$= \frac{d_b \|\mu_b\|_F^2}{d_a} + d_b \langle Y, uu^* \rangle = \frac{d_b \|\mu_b\|_F^2}{d_a} + d_b \sum_i y_i \langle u, v_i \rangle^2$$

$$\leq \frac{d_b \|\mu_b\|_F^2}{d_a} + d_b \left(\sum_i y_i^2 \langle u, v_i \rangle^2 \right)^{1/2} \left(\sum_i \langle u, v_i \rangle^2 \right)^{1/2}$$

$$\leq \frac{d_b \|\mu_b\|_F^2}{d_a} + d_b \frac{\gamma(U,V)}{\sqrt{d_a}} \|Y\|_F \leq \frac{d_b \|\mu_b\|_F^2}{d_a} + \lambda\gamma \frac{\sqrt{d_b} \|\mu_b\|_F}{d_a}$$

So adding up all these terms, we get a competing force of

$$\sum_{b \neq a} \frac{d_b}{d_a} \|\mu_b\|_F^2 + \frac{\lambda\gamma}{d_a} \sqrt{d_b} \|\mu_b\|_F < \frac{1}{d_a} \left[\|\mu\|_*^2 + \lambda\gamma\sqrt{k} \|\mu\|_* \right]$$

Now the a-th part is pushing $\approx \varepsilon_a / d_a = \|\mu_a\|_{op}$. For comparison, if $\epsilon_b := d_b \|\mu_b\|_{op}$, then we have the bound

$$\|\mu\|_*^2 = \sum_b d_b \|\mu_b\|_F^2 \leq \sum_b d_b^2 \|\mu_b\|_{op}^2 = \sum_b \epsilon_b^2$$

So the above says if $\lambda\gamma \ll 1/k$, any eigenvalue that is on the same order as the average will be shrinking in magnitude.

Fact 11. *Random unitaries are very incoherent:*

$$\max_{ij} \langle u_i, e_j \rangle^2 \leq \frac{\log d}{d}$$

with $1/poly(d)$ failure probability.

Claim 12. *The eigenbases of $\{\mu_a\}, \{T_{ba}(\mu_b)\}$ are roughly as incoherent as random unitaries. This is true even after running gradient flow for some time.*

The problem remaining is that eigenvectors are not continuous objects, so a bound on incoherence at various discrete times does not imply a bound for all continuous times.

2.11 Experiments

1. Check $\|\mu(t)\|_{op}$ is (close to) monotone decreasing
2. Check $\int_0^T \|\mu(t)\|_{op} \gg \|\int_0^T \mu(t)\|_{op}$. I'm assumming this last term is roughly the operator norm of the scaling, but can check that too: what is the operator norm of log of the scaling at time T?
3. Plot $d\|\mu(t)\|_{op}^2 / \|\mu(t)\|_F^2$. Should be $\tilde{O}(1)$ for some amount of time.

2.12 Operator Norm Monotonicity

Our quantum setting is that we have $\{X_1, ..., X_n\}$ gaussian samples $X \sim \mathcal{N}(0, \frac{1}{n} \otimes_a \frac{1}{d_a} I_a)$ and we want to show the optimizers of the log-Likelihood function is "nearby". In the classical setting we have k-tensor with entries from the χ-square distribution with mean $\frac{1}{nD}$ and n degrees of freedom: $\chi(\frac{1}{D}, n)$. For any given torus $T \subseteq G$, the Kempf-Ness function of our quantum input on this torus looks the same as a classical input.

Claim 13. *If we condition on T being the "minimizing torus" so that the optimizer of the log-likelihood function lies in this torus, then the input looks like the classical setting.*

By the above reduction, if we can prove a bound on the optimizer for a fixed torus, then we have shown the same bound in the quantum setting.

Fact 14. *For $A \in \mathbb{R}^{n\times m}$, $\|A\|_{\infty\to\infty} \leq \max_{i\in n} \sum_{j\in[m]} |A_{ij}|$.*

We consider A to be a bipartite marginal, so with entries $\chi(\frac{1}{d_a d_b}, \frac{nD}{d_a d_b})$. With high probability the row sums are $\in \frac{1\pm\varepsilon}{d_a}$ and the column sums are $\in \frac{1\pm\varepsilon}{d_b}$, so $\|A\|_{\infty\to\infty} \leq \frac{1+\varepsilon}{d_a}$. We want to show that the gradient flow decreases the operator norm of the moment map:

$$-\partial_t \langle \mu, e_a \otimes e_i \rangle = (\rho_a)_i d_a (\mu_a)_i + \sum_{b\neq a} \langle e_i, A^{ab} d_b \mu_b \rangle$$

Here $(\rho_a)_i \in \frac{1\pm\varepsilon}{d_a}$, and the first term has the same sign as $(\mu_a)_i$ and so is pushing the correct way. We would like to show all other forces (from $b \neq a$) cannot push enough to go the wrong way.

$$-\partial_t |(\mu_a)_i| \geq |(\mu_a)_i| d_a (\rho_a)_i - \sum_{b\neq a} \|A^{ab}\|_{\infty\to\infty} \|d_b \mu_b\|_\infty$$

Then if e.g. $\forall b \neq a : \|A^{ab}\|_{\infty\to\infty} \ll 1/kd_a$ and we consider (a, i) to be the marginal with highest error, this error is decreasing exponentially.

Let $\langle y, \mathbf{1}_b \rangle = 0$:

$$\mathbb{E} \sum_j y_j \chi(\frac{1}{d_a d_b}, \frac{nD}{d_a d_b}) = \sum_j y_j \frac{1}{d_a d_b} = 0$$

$$\log \mathbb{E} \exp \theta \sum_j y_j \chi(\frac{1}{d_a d_b}, \frac{nD}{d_a d_b}) = \sum_j \frac{-nD}{2 d_a d_b} \log\left(1 - \frac{2\theta y_j}{nD}\right)$$

$$\lesssim \theta^2 \frac{\|y\|_2^2}{nD d_a d_b} \qquad \forall \theta \lesssim \left(\frac{\|y\|_\infty}{nD}\right)^{-1}$$

$$\implies \mathbb{P}[|\cdot| \geq t] \leq \begin{cases} \exp\left(-\frac{\Omega(nD d_a d_b t^2)}{\|y\|_2^2}\right) & \forall t \lesssim \frac{\|y\|_2^2}{d_a d_b \|y\|_\infty} \\ \exp\left(-\frac{\Omega(nDt)}{\|y\|_\infty}\right) & o.w. \end{cases}$$

To bound $\|A\|_{\infty\to\infty} \ll 1/kd_a$, we choose $t \ll \frac{\|y\|_\infty}{kd_a}$ which is in the second case. So we get probability $\le \exp(-\frac{nD}{kd_a})$. If we run a net over $\mathbb{R}^{d_b} \perp \mathbf{1}_b$ of size $\exp(\tilde{O}(d_b))$ and union bound over all rows, we get $\|A\|_{\infty\to\infty} \ll 1/kd_a$ whp whenever $nD \gg kd_a(d_b + \log d_a)$.

To make this robust to perturbations we could either give a deterministic bound on how a perturbation affects $\|A\|_{\infty\to\infty}$; or we could run the net over all perturbations near I of size $\exp(\sum_a d_a)$. Either way we should get that $\|\mu\|_{op}$ is decreasing for all time when $nD \sim \tilde{O}_k(\max_{ab} d_a d_b)$

2.13 Some Remarks on Extensions

Remark 5. *One fantastic thing about Michael's proof of* **??** *is that it generalizes quite well to a statement about $\|\cdot\|_{p\to p}$. Explicitly, if $\sigma := \|M\|_{p\to p}$, then we have the following bound:*

$$|\langle vv^*, I_{\overline{ab}} \otimes Z_\pm \otimes Y_\pm\rangle| \le \sigma \|Z_\pm\|_p \|Y_\pm\|_q$$

$$\|Z_+\|_p\|Y_+\|_q + \|Z_+\|_p\|Y_-\|_q + \|Z_-\|_p\|Y_+\|_q + \|Z_-\|_p\|Y_-\|_q$$
$$\le (2\|Z_+\|_p^p + 2\|Z_-\|_p^p)^{1/p}(2\|Y_+\|_q^q + 2\|Y_-\|_q^q)^{1/q} = 2\|Z\|_p\|Y\|_q$$

[AR: We may want to use this result for $p \in \{1,\infty\}$ to bound the confounding forces towards $\|\mu\|_{op}$.]

We can in fact move a more general statement quite simply. This line of reasoning also applies to the operator on $\perp I$, but shows why Pisier's theorem is spectacular.

Fact 15. *$\|\nabla^2_{ab}\|_{op\to op} = \|\rho^a\|_{op}$, and dually $\|\nabla^2_{ab}\|_{1\to 1} = \|\rho^b\|_{op}$.*

Proof.

$$\|\nabla^2_{ab}\|_{op\to op} = \sup_{\|X\|_{op}\le 1, \|Y\|_1 \le 1} \langle \rho^{ab}, X \otimes Y\rangle$$

Choose U, V to be the bases of the optimizers. Then restricting to this torus, we really have a matrix $M_{ij} := \langle \rho^{ab}, u_i u_i^* \otimes v_j v_j^*\rangle$, and $\|\nabla^2_{ab}\|_{op\to op} = \|M\|_{\infty\to\infty}$ for this basis.

$$\|M\|_{\infty\to\infty} = \max_i \sum_j |M_{ij}| = \max_i \sum_j \langle \rho^{ab}, u_i u_i^* \otimes v_j v_j^*\rangle = \max_i \langle u_i u_i^*, \rho^a\rangle \le \|\rho^a\|_{op}$$

The second statement follows since $(\nabla^2_{ab})^* = \nabla^2_{ba}$. □

Theorem 16 (Riesz-Thorin)**.** *Given $\|T\|_{p_0\to q_0}, \|T\|_{p_1\to q_1} < \infty$, we have the following log convexity: if $\frac{1}{p_\theta} = \frac{\theta}{p_0} + \frac{1-\theta}{p_1}$ and q_θ defined similarly:*

$$\|T\|_{p_\theta\to q_\theta} \le \|T\|_{p_0\to q_0}^\theta \|T\|_{p_1\to q_1}^{1-\theta}$$

Corollary 17. *For $p_0 = 1, p_1 = \infty$, we have $\frac{1}{p} = \frac{1/p}{1} + \frac{1-1/p}{\infty}$.*

$$\|\nabla^2_{ab}\|_{p\to p} \le \|\nabla^2_{ab}\|_{1\to 1}^{1/p} \|\nabla^2_{ab}\|_{op\to op}^{1-1/p} = \|\rho^b\|_{op}^{1/p} \|\rho^a\|_{op}^{1-1/p}$$

In our setting, we actually care about $\nabla^2(I-P)$ where we project away from I. As shown elsewhere, for matrices we can actually establish improved bounds by a net argument, which by the above corollary shows that a matrix of χ^2 variables actually shrinks all p-norms of inputs $\perp 1$. Unfortunately the net argument for the quantum setting loses an important dimension factor. Further, for the low sample regime, it can be shown that the above projection does not improve the bound $\|\nabla^2_{ab}(I-P)\|_{op\to op}$. This is another reason Pisier's result is astonishing.

2.14 A silly trick

The same idea that showed monotonicity of the operator norm for operator scaling can be used to very slightly improve the required expansion for k-tensors. In the proper norm, our gradient direction will be $\{d_a \mu^a\}$, and so to use robustness we need a bound on perturbation size $\sum_a d_a \|\mu^a\|_{op} =: \sum_a \varepsilon_a$ for all time. Let $\varepsilon := \max_a \varepsilon_a$. In the proper norm, we have weak two sided bounds at any time:

$$\varepsilon^2 = d_a^2 \|\mu_a\|_{op}^2 \leq d_a^2 \|\mu^a\|_F^2 \leq d_a \left(\sum_b d_b \|\mu^b\|_F^2 \right) \leq d_a \left(\sum_b \varepsilon_b^2 \right) \leq d_a k \varepsilon^2$$

Now recall the change in operator norm under negative gradient flow:

$$-\partial_t \varepsilon := -d_a \partial_t \langle uu^*, \mu^a \rangle = d_a \left(\langle \rho^a, (uu^*)(d_a \mu^a) \rangle + \sum_{b \neq a} \langle \rho^{ab}, uu^* \otimes d_b \mu^b \rangle \right)$$

$$\geq (1 \pm \varepsilon) \left(\varepsilon_a - \sum_{b \neq a} \varepsilon_b \right) \geq (1 \pm \varepsilon)(k-2)\varepsilon$$

Here we've used that ε_a is the biggest. To see the transition for the first to second line, recall $\|d_b \mu^b\|_{op} = \varepsilon_b \leq \varepsilon$, and for any basis $\{v_i\} \subseteq \mathbb{R}^{d_b}$ we have

$$\langle \rho^{ab}, uu^* \otimes d_b \mu^b \rangle \leq \langle \rho^{ab}, uu^* \otimes \varepsilon_b I_b \rangle = \varepsilon_b \langle \rho^a, uu^* \rangle = \varepsilon_b \frac{1 \pm \varepsilon}{d_a}$$

Here we've used $\pm$ depending on whether ε comes from a largest or a smallest eigenvalue. So what the above inequalities show is $\partial_t(\log \varepsilon) \leq (1+\varepsilon)(k-2) \approxeq (k-2)$. From here on I'll be ignoring low order terms. I'll mention later that they add a small error to the main theorem. Therefore we assume we initially have $1 - o(1) \approxeq 1$ strong convergence at time 0 and further that this is true up to another $o(1)$ term while our perturbation size is $o(1)$.

Claim 18. $\forall t : \varepsilon(t) \lesssim d^{\frac{(k-2)}{1+2(k-2)}} \varepsilon(0)$

Proof. Let T be the first time when $\varepsilon(t) \geq d^{(k-2)c} \varepsilon(0)$:

$$(\log \varepsilon)' \leq k-2 \implies T \geq \frac{\log \varepsilon(T) - \log \varepsilon(0)}{k-2} = c \log d$$

On the other hand let T^* be the last time when $\varepsilon(T^*) \leq o(1)$, so up to this time we have very strong convergence:

$$\forall t \leq T^* : \sum_a d_a \|\mu^a\|_F^2(t) \leq e^{-t} \sum_a d_a \|\mu^a\|_F^2(0) \leq e^{-t} k\varepsilon(0)^2$$

In particular if $T^* > \log k d^{1-2(k-2)c}$:

$$\varepsilon(T^*)^2 \leq e^{-T^*} dk\varepsilon(0)^2 < d^{2(k-2)c}\varepsilon(0)^2 = \varepsilon(T)^2$$

This would be a contradiction if $T \geq T^*$, so matching terms we get

$$c \log d = \log k + (1 - 2(k-2)c) \log d \iff c = \frac{1}{1 + 2(k-2)}$$

Again we ignore the lower order $\log k$ term. So if we have strong convergence up to time T^* then the claim is proven. □

The above claim shows that if $d^{\frac{(k-2)}{1+2(k-2)}}\varepsilon < o(1)$, and we have $1-o(1)$ at time 0, then in fact we have $1 - o(1)$ strong convergence at all time and we remain in an $d^{\frac{(k-2)}{1+2(k-2)}}\varepsilon < o(1)$ sized operator norm ball the whole time. Therefore it suffices to take $d^{3-\frac{1}{1+2(k-2)}}$ samples to get strong convergence. Forgive me the k's and this is our savings on samples.

2.15 The Final Concentration

Above we were trying to bound competing forces on the largest eigenvalue of μ for all time to make sure our scaling remains inside a small $\|\cdot\|_{op}$ ball. [AR: Here we will show it is enough to have small forces initially.]

[AR: Correction: the following line of argument isn't really good enough. This naive way of bounding the competing forces means that our $\|\mu\|_{op}$ jumps into the worst regime (exponentially increasing) within a constant amount of time. This is because our bound on the force is proportional to the size of the perturbation.]

If our input at time t is scaled by e^{δ}, then we have the following rough bound on our errors:

$$\|\mu_b(t) - \mu_b(0)\|_{op} = \sup_v \langle (e^{\delta} - I) \cdot \rho^{ab}, I_a \otimes vv^* \rangle \lesssim \delta \|T^*\|_{op_a \to op_b} \|I_a\|_{op} \|vv^*\|_1 \leq \delta \|\rho^b\|_{op}$$

$$\|T(\mu_b)(t)\|_{op} - \|T(\mu_b)(0)\|_{op} \leq \|T(\mu_b)(t) - T(\mu_b)(0)\|_{op} = \sup_u \langle uu^*, T(\mu_b)(t) - T(\mu_b)(0) \rangle$$

Let $\mu_b(t) := \mu_b(0) + Y =: \mu_b + Y$

$$\begin{aligned}\|T^t(\mu_b(t)) - T^0(\mu_b(0))\|_{op} &\leq \|T^t(\mu_b + Y) - T^t(\mu_b)\|_{op} + \|T^t(\mu_b) - T^0(\mu_b)\|_{op} \\ &\leq \|T^t\|_{op_b \to op_a} \|Y\|_{op} + \delta \|T^0\|_{op_b \to op_b} \|\mu_b\|_{op}\end{aligned}$$

$$\leq \delta \|\rho^a(t)\|_{op} \|\rho_b\|_{op} + \delta \|\rho_a\|_{op} \|\mu_b\|_{op}$$

Assuming $\|\mu_b\|_{op} \ll \delta/d_b$, this is on the order of $\delta/d_a d_b$. Recall the formula for the top eigenvalue change:

$$-\partial_t d_a \|\mu_a\|_{op} = \|d_a \rho^a\|_{op} \|d_a \mu^a\|_{op} + \sum_{b \neq a} d_a d_b \langle \rho^{\{a,b\}}, uu^* \otimes \mu_b \rangle$$

$$\geq (1 \pm d_a \|\mu_a\|_{op}) \|d_a \mu^a\|_{op} - \sum_{b \neq a} d_a d_b \|T_{b \to a}(\mu_b)\|_{op}$$

$$\geq (1 \pm \epsilon_a) \epsilon_a - (k-1)\delta$$

This is the competing force at time t, assuming it is very small at time 0. Therefore we have that any (normalized) eigenvalue of μ that is greater than $(k-1)\delta$ at time t is decreasing in absolute value. It should be enough to have $\delta \ll 1/k \log D$, and then in total the scaling should be of operator norm $\ll 1$, so we will have convergence.

3 Better upper bound for matrix model using HCIZ

3.1 Attempt at better upper bound for matrix normal model

Note if $d_a > d_b$, the error of the a marginal will be much larger than that of the b marginal. So we could try to analyze what happens after one iteration of flip-flop where we normalize the a to identity. We claim our input if our input is $X_1, ..., X_n \subseteq \mathbb{R}^{d_a \times d_b}$ iid from $N(0, I_{a \times b})$, and we normalize so that the rows of X form an orthogonal basis for $\mathbb{R}^{d_a}$, our new input Y will be uniform from the Stiefel manifold of bases for d_a dimensional subspaces of $\mathbb{R}^{nd_b}$. Given this we can try to run a net argument to bound the b marginal:

$$Y := \left(\sum_{i=1}^n X_i X_i^* \right)^{-1/2} X$$

$$\mu_b(Y) = \sup_{\xi \in \mathbb{R}^b} \langle \xi \xi^*, \sum_i X_i^* \left(\sum_j X_i X_i^* \right)^{-1} X_i \rangle$$

All this to say I would like to bound the following using HC formula:

Theorem 19. *For Hermitian $A, B \in \mathbb{R}^N$, wlog diagonal and increasing $a_1 < ... < a_N, b_1 < ... < b_N$ we have:*

$$\int_{U(N)} \exp\langle A, UBU^* \rangle dU = \prod_{k=1}^{N-1} k! \frac{\det\{e^{a_i b_j}\}_{ij}}{\prod_{j>i}(a_j - a_i)(b_j - b_i)}$$

I would like to choose A to be a rank d_a orthogonal projection and B a rank n orthogonal projection in $\mathbb{R}^{N=nd_b}$. Because these have non-distinct eigenvalues we use L'hopital's rule to take limits.

$$\lim_{a_1\to 0} e^{a_1 b_j} = 1$$

$$\lim_{a_1\to 0} a_j - a_1 = a_j$$

So the matrix has first row all ones and denominator $\prod_{j>i>1}(a_j - a_i)\prod_{j>1} a_j$. In this case if we set $a_2 = 0$, the second row will become the same as the first, so the matrix becomes singular. Luckily the denominator also has an a_2 term so we use L'hopitals:

$$\lim_{a_2\to 0} \partial_{a_2} e^{a_2 b_j} = b_j$$

$$\lim_{a_2\to 0} \partial_{a_2} \prod_{j>i>1}(a_j - a_i)\prod_{j>1} a_j = \prod_{j>i>2}(a_j - a_i)\prod_{j>2} a_j^2$$

The last step is by the product rule, since the only non-vanishing term is when the ∂ is applied to the a_2 term.

Now the matrix has first row $\mathbf{1}$ and second row $\boldsymbol{b}$. Therefore setting $a_3 = 0$ will make the matrix singular, and so will the derivative. Again luckily there is an order two zero in the denominator a_3^2 so we apply L'hopitals twice:

$$\lim_{a_3\to 0} \partial_{a_3}^2 e^{a_2 b_j} = b_j^2$$

$$\lim_{a_3\to 0} \partial_{a_3}^2 \prod_{j>i>2}(a_j - a_i)\prod_{j>2} a_j^2 = 2! \prod_{j>i>3}(a_j - a_i)\prod_{j>3} a_j^3$$

Continuing this way we get the first $N - d_a$ rows

$$\begin{pmatrix} 1 & 1 & ... & 1 \\ b_1 & b_2 & ... & b_N \\ b_1^2 & b_2^2 & ... & b_N^2 \\ & & ... & \\ b_1^{N-d_a-1} & b_2^{N-d_a-1} & ... & b_N^{N-d_a-1} \end{pmatrix}$$

And the denominator terms including A are

$$\prod_{k=1}^{N-d_a-1} k! \prod_{j>i>N-d_a}(a_j - a_i) \prod_{j>N-d_a} a_j^{N-d_a}$$

Now we will set $b_1 = ... = b_{N-n} = 0$ by the same process:

$$\lim_{b_1\to 0} b_1^0 = 1, \quad \lim_{b_1\to 0} b_1^{(i-1)\geq 1} = 0, \quad \lim_{b_1\to 0} e^{a_i b_1} = 1$$

$$\lim_{b_1\to 0} \prod_{j>i}(b_j - b_i) = \prod_{j>i>1}(b_j - b_i)\prod_{j>1} b_j$$

Therefore the first column is $(1, 0, ..., 0, 1, 1, ..., 1)$, i.e. the first entry and the last d_a entries are 1. In the second iteration setting $b_2 = 0$ would make the matrix singular and also make the denominator vanish:

$$\lim_{b_2\to 0} \partial_{b_2} b_2^0 = 0, \quad \lim_{b_2\to 0} \partial_{b_2} b_2^1 = 1, \quad \lim_{b_2\to 0} \partial_{b_2} b_2^{(i-1)>1} = 0, \lim_{b_2\to 0} \partial_{b_2} e^{a_i b_2} = a_i$$

$$\lim_{b_2\to 0} \partial_{b_2} \prod_{j>i>1} (b_j - b_i) \prod_{j>1} b_j = \prod_{j>i>2} (b_j - b_i) \prod_{j>2} b_j^2$$

The second column is therefore $(0, 1, 0, ..., 0, a_{N-d_a+1}, ..., a_N)$. One more iteration:

$$\lim_{b_3\to 0} \partial_{b_2}^2 b_2^0 = 0, \quad \lim_{b_2\to 0} \partial_{b_2}^2 b_2^1 = 0, \quad \lim_{b_2\to 0} \partial_{b_2}^2 b_2^2 = 2!, \quad \lim_{b_2\to 0} \partial_{b_2}^2 b_2^{(i-1)>2} = 0, \lim_{b_2\to 0} \partial_{b_2}^2 e^{a_i b_2} = a_i^2$$

$$\lim_{b_2\to 0} \partial_{b_2}^2 \prod_{j>i>2} (b_j - b_i) \prod_{j>2} b_j^2 = 2! \prod_{j>i>3} (b_j - b_i) \prod_{j>3} b_j^3$$

Wlog at this point we assume $d_a > n$ so $N - d_a < N - n$. So continuing this way we have the top left block $(N - d_a) \times (N - n)$, the first $N - d_a$ columns are diag$\{(i-1)!\}$ and the next $d_a - n$ columns are all 0. The bottom left block $d_a \times N - n$ is

$$\begin{pmatrix} 1 & a_{N-d_a+1} & a_{N-d_a+1}^2 & \cdots & a_{N-d_a+1}^{N-n-1} \\ 1 & a_{N-d_a+2} & a_{N-d_a+2}^2 & \cdots & a_{N-d_a+2}^{N-n-1} \\ & & \cdots & & \\ 1 & a_N & a_N^2 & \cdots & a_N^{N-n-1} \end{pmatrix}$$

The denominator terms involving b are

$$\prod_{k=1}^{N-n-1} k! \prod_{j>i>N-n} (b_j - b_i) \prod_{j>N-n} b_j^{N-n}$$

Now we would like to set $b_{N-n+1} = ... = b_N = 1$.

$$\lim_{b_{N-n+1}\to 1} b_{N-n+1}^{i-1} = 1, \quad \lim_{b_{N-n+1}\to 1} e^{a_i b_{N-n+1}} = e^{a_i}$$

$$\lim_{b_{N-n+1}\to 1} \prod_{j>i>N-n} (b_j - b_i) = \prod_{j>i>N-n+1} (b_j - b_i) \prod_{j>N-n+1} (b_j - 1)$$

We have left out the term $\prod_{j>N-n} b_j^{N-n}$ as we will set them all to one and so this term will end up being 1. So the first column in this block is $(1, ..., 1, e^{a_{N-d_a+1}}, ..., e^{a_N})$. If we set $b_{N-n+2} = 1$ then the column will be the same and the denominator has a root:

$$\lim_{b\to 1} \partial_b b^{i-1} = i - 1, \quad \lim_{b\to 1} \partial_b e^{a_i b} = a_i e^{a_i}$$

$$\lim_{b_{N-n+2}\to 1} \partial_{b_{N-n+2}} \prod_{j>i>N-n+1} (b_j - b_i) \prod_{j>N-n+1} (b_j - 1) = \prod_{j>i>N-n+2} (b_j - b_i) \prod_{j>N-n+2} (b_j - 1)^2$$

One more iteration:

$$\lim_{b\to 1}\partial_b^2 b^{i-1} = (i-1)(i-2), \qquad \lim_{b\to 1}\partial_b^2 e^{a_i b} = a_i^2 e^{a_i}$$

$$\lim_{b_{N-n+3}\to 1}\partial^2_{b_{N-n+2}}\prod_{j>i>N-n+2}(b_j-b_i)\prod_{j>N-n+2}(b_j-1)^2 = \prod_{j>i>N-n+3}(b_j-b_i)2!\prod_{j>N-n+3}(b_j-1)^3$$

Continuing on this way we can write the right block. For $i \in \{0,...,N-d_a-1\}$ and $j \in \{0,...,n-1\}$, the entry in position $(i+1, N-n+j+1)$ is $(i)_j := i!/(i-j)!$. In the bottom right block we have for $i \in \{N-d_a+1,...,N\}$ and $j \in \{0,...,n-1\}$, the entry in position $(i, N-n+j+1)$ is $a_i^j e^{a_i}$. We have set all b variables to scalars, so there are no more b variables in the denominator and we have added a term $\prod_{k=1}^{n-1} k!$.

Now we would like the set $a_{N-d_a+1} = ... = a_N = t$.

$$\lim_{a_{N-d_a+1}\to t} a^{j-1}_{N-d_a+1} = t^{j-1}$$

$$\lim_{a_{N-d_a+1}\to t}\prod_{j>i>N-d_a}(a_j-a_i)\prod_{j>N-d_a}a_j^{N-d_a} = \prod_{j>i>N-d_a+1}(a_j-a_i)\prod_{j>N-d_a+1}(a_j-t)\prod_{j>N-d_a+1}a_j^{N-d_a}t^{N-d_a}$$

Again setting $a_{N-d_a+2} = 1$ would make the matrix singular and the denominator would vanish so we have

$$\lim_{a\to t}\partial_a a^{j-1} = (j-1)t^{j-2}$$

$$\lim_{a_{N-d_a+2}\to t}\partial_{a_{N-d_a+2}}\prod_{j>i>N-d_a+1}(a_j-a_i)\prod_{j>N-d_a+1}(a_j-t)\prod_{j>N-d_a+1}a_j^{N-d_a}t^{N-d_a}$$

$$= \prod_{j>i>N-d_a+2}(a_j-a_i)\prod_{j>N-d_a+2}(a_j-t)^2\prod_{j>N-d_a+2}a_j^{N-d_a}t^{2(N-d_a)}$$

Continuing on this way we can find the bottom left block. For $i \in \{0,...,d_a-1\}$ and $j \in \{0,...,N-n-1\}$, the entry in position $(N-d_a+i+1, j+1)$ is $(j)_i t^{j-i}$. For $j \in \{0,...,n-1\}$ the entry in position $(N-d_a+i+1, N-n+j+1)$ is $\partial^i_{x=t}x^j e^x$. This is not an explicit formula but we will arrive at some recurrences involving these terms. The denominator is $\prod_{k=1}^{d_a-1} k! t^{d_a(N-d_a)}$. In total we will rewrite the whole matrix after all the L'hopitals

$$\begin{pmatrix} X^{00} & X^{01} \\ X^{t0} & X^{t1}\end{pmatrix}$$

The subscripts indicate the setting of a_i, b_j in that block. Then $X^{00} \in \mathbb{R}^{(N-d_a)\times(N-n)}$ and its top left square block of size $(N-d_a)\times(N-d_a)$ is a diagonal matrix with entries

$$(X^{00})_{i+1,i+1} = i!$$

The rest is 0. $X^{01} \in \mathbb{R}^{(N-d_a)\times n}$ has entries

$$(X^{01})_{i+1,j+1} = (i)_j$$

Note this is 0 for $i < j$. $X^{t0} \in \mathbb{R}^{d_a \times (N-n)}$ has entries

$$(X^{t0})_{i+1,j+1} = (j)_i t^{j-i}$$

Again this is 0 for $j < i$. $X^{t1} \in \mathbb{R}^{d_a \times n}$ is the most complicated and has entries

$$(X^{t1})_{i+1,j+1} = \partial^i_{x=t} x^j e^x$$

It is easy to calculate the beginning terms: $\partial^0_{x=t} x^j e^x = t^j e^t$ and $\partial^i_{x=t} x^0 e^t = e^t$. Further note the following recurrence:

$$\partial^i_x (x^j e^x) = \partial^{i-1}_x (j x^{j-1} e^x + x^j e^x)$$

We can rewrite this in block form where now i, j index the $(i+1, j+1)$ entry of the given block:

$$\begin{pmatrix} \operatorname{diag}\{i!\}_{i=0}^{N-d_a-1} & 0_{(N-d_a)\times(d_a-n)} & \{(i)_j\}_{i=0,j=0}^{i=N-d_a-1,j=n-1} \\ \{(j)_i t^{j-i}\}_{i=0,j=0}^{i=d_a-1,j=N-n-1} & & \{\partial^i_x x_j e^x\}_{i=0,j=0}^{i=d_a-1,j=n-1} \end{pmatrix}$$

We give new names for ease of use:

$$\begin{pmatrix} A & 0 & B \\ C & & D \end{pmatrix}$$

So the HC formula has reduced to the determinant of the above matrix divided by the following factorial terms:

$$\left(\prod_{k=1}^{N-d_a-1} k!\right)\left(\prod_{k=1}^{N-n-1} k!\right)\left(\prod_{k=1}^{n-1} k!\right)\left(\prod_{k=1}^{d_a-1} k!\right)\left(t^{d_a(N-d_a)}\right) \Big/ \left(\prod_{k=1}^{N-1} k!\right)$$

The final term is from the numerator of the original formula.

To calculate the determinant we will use Schur complement:

$$\det \begin{pmatrix} A & 0 & B \\ C & & D \end{pmatrix} = \det(A)\det(D - CA^{-1}B)$$

with slight abuse of notation as the first $d_a - n$ columns of the top right block are 0 and so don't affect the Schur complement. In particular if $\{B_{j+1}\}$ are the columns of B and $\{C_{i+1}\}$ are the rows of C, we want to calculate $\{C^*_{i+1} A^{-1} B_{j+1}\}$. We can find an explicit for when $i = 0$ or $j = 0$ and then we will show that these inner products satisfy the same recurrence as the D block.

$C_1 = (1, t, t^2, ..., t^{N-d_a-1})$, and $B_1 = (1, 1, ..., 1)$. Since A is diagonal $(A^{-1})_{i+1} = (i!)^{-1}$. Therefore:

$$C^*_1 A^{-1} B_1 = \sum_{k=0}^{N-d_a-1} \frac{C_{1,k+1} B_{1,k+1}}{A_{k+1}} = \sum_{k=0}^{N-d_a-1} \frac{t^k}{k!}$$

Note this is the first $N - d_a - 1$ terms of the Taylor approximation for e^t at 0. For B_{j+1} the first j entries are 0, and after this the terms are j falling terms.

$$C_1^* A^{-1} B_{j+1} = \sum_{k=0}^{N-d_a-1} \frac{C_{1,k+1} B_{j+1,k+1}}{A_{k+1}} = \sum_{k=j}^{N-d_a-1} \frac{t^k k!/(k-j)!}{k!} = t^j \sum_{k=0}^{N-d_a-j-1} \frac{t^k}{k!}$$

Similarly we can note the first i terms of C_{i+1} are 0:

$$C_{i+1}^* A^{-1} B_1 = \sum_{k=i}^{N-d_a-1} \frac{t^{k-i} k!/(k-i)! \cdot 1}{k!} = \sum_{k=0}^{N-d_a-i-1} \frac{t^k}{k!}$$

By our previous easy calculation of the first row and column of D, this gives easy approximations for the first row and column of the Schur complement:

$$(D - CA^{-1}B)_{i+1,1} = e^t - \sum_{k=0}^{N-d_a-i-1} \frac{t^k}{k!} \in \left(0, \frac{t^{N-d_a-i}}{(N-d_a-i)!}\right)$$

$$(D - CA^{-1}B)_{1,j+1} = t^j \left(e^t - \sum_{k=0}^{N-d_a-j-1} \frac{t^k}{k!}\right) \left(0, \frac{t^{N-d_a}}{(N-d_a-j)!}\right)$$

We will also show the same recurrence as D:

$$C_{i+1}^* A^{-1} B_{j+1} = C_i^* A^{-1} B_{j+1} + j C_i^* A^{-1} B_j$$

This is accomplished by the following recurrence relations:

$$C_{i+1,k+1} = (k)_i t^{k-i} = k(k-1)_{i-1} t^{k-i} = k C_{i,k}$$

$$A_{k+1} = k! = k(k-1)! = k A_k$$

$$B_{j+1,k+1} = (k)_j = k(k-1)_{j-1} = k B_{j,k}$$

$$B_{j+1,k} = (k-1)!/(k-1-j)! = (k-j)(k-1)!/(k-j)! = (k-j) B_{j,k}$$

Again noting the first j terms of C_{j+1} are 0:

$$C_{i+1}^* A^{-1} B_{j+1} = \sum_{k=j}^{N-d_a-1} \frac{C_{i+1,k+1} B_{j+1,k+1}}{A_{k+1}} = \sum_{k=j}^{N-d_a-1} \frac{k C_{i,k}}{k A_k} (k B_{j,k})$$

$$= j \sum_{k \geq j} \frac{C_{i,k} B_{j,k}}{A_k} + \sum_{k \geq j} \frac{C_{i,k}}{A_k} ((k-j) B_{j,k}) = j C_i^* A^{-1} B_j + C_i^* A^{-1} B_{j+1}$$

So we have in fact shown all entries of the matrix $D - CA^{-1}B$ are non-negative, and we have a simple recurrence as we go down the rows. One more simplification for the sake of bounds is to write everything in terms of the first row and column, for which we have clean bounds. In general our matrix M satisfies $M_{i+1,j+1} = M_{i,j+1} + j M_{i,j}$. Under this condition we show

Above the main diagonal, i.e. $j \geq i$, the recurrence stops at the first row:

$$M_{i+1,j+1} = \sum_{k=0}^{i} \binom{i}{k} (j)_k M_{1,j+1-k} = \sum_{k=0}^{i} \frac{i!j!}{k!(i-k)!(j-k)!} M_{1,j+1-k}$$

Otherwise, if $i > j$ we have contribution from the first column

$$M_{i+1,j+1} = \sum_{k=0}^{j} \frac{i!j!}{k!(i-k)!(j-k)!} M_{1,j+1-k} + \sum_{k=0}^{i-j} \binom{j+k}{k} j! M_{(i-j)+1-k,1} - \binom{i}{j} j! M_{11}$$

This last term is because we counted M_{11} in both sums. We can rewrite this

$$M_{i+1,j+1} = \sum_{k=0}^{j} \frac{i!j!}{k!(i-k)!(j-k)!} M_{1,j+1-k} + \sum_{k=j+1}^{i} (k)_j M_{i+2-k,1}$$

$$= \sum_{k=0}^{j} \frac{i!j!}{k!(i-k)!(j-k)!} M_{1,j+1-k} + \sum_{k=0}^{i-j} (i-k)_j M_{k+1,1}$$

These formulas can be proved simply by induction. From here we can use the following approximations:

$$\det(X) = \sum_{\sigma \in S_N} (-1)^{|\sigma|} \prod_i X_{i,\sigma(i)} \leq \sum_{\sigma \in S_N} \prod_i |X_{i,\sigma(i)}| \leq \prod_i \left(\sum_j |X_{ij}| \right)$$

$$\det(X) \leq \prod_i \|X_i\|_2$$

I'm having a bit of trouble actually doing bounds though.

4 Old stuff

4.1 Different Inner Product

Definition 2. *For desired marginals $\{R_a^2\}_{a \in [k]}$ (assume for now R are Hermitian though we can pick different square roots if required), define inner product*

$$\langle Z, Y \rangle_R := \sum_a \langle R_a Z R_a^*, Y \rangle$$

$$\|Z\|_R^2 := \langle Z, Z \rangle_R = \sum_a \|R_a Z_a\|_F^2$$

We restate the projective likelihood function and define gradient and Hessian in this metric:

Definition 3.

$$f_{\boldsymbol{X}}(\Theta_1,\ldots,\Theta_n)=\log\left\langle\sum_{i\in[n]}X_iX_i^*,\bigotimes_{a\in[k]}\Theta_a\right\rangle-\sum_{a\in[k]}\frac{1}{d_a}\log\det\Theta_a$$

Also $\rho:=\sum_i X_iX_i^*$ *and* $\{\rho^S\}_{S\subseteq[k]}$ *are marginals.*

Fact 20.

$$(\nabla f(I))_a=d_a\rho^{\{a\}}-I_a$$

Proof. We can define ∇f dually as $\forall Z:\langle\nabla f(I),Z\rangle_R:=\partial_{t=0}f(e^{tZ})$

$$\partial_{t=0}f(e^{tZ_a})=\partial_{t=0}\langle\rho,I_{\overline{a}}\otimes e^{tZ_a}\rangle-\partial_{t=0}\frac{1}{d_a}\log\det e^{tZ_a}$$

$$=\left\langle\rho^{\{a\}}-\frac{1}{d_a}I_a,Z_ae^{tZ_a}\right\rangle|_{t=0}=\left\langle R_a^{-1}\left(\rho^{\{a\}}-\frac{1}{d_a}I_a\right)R_a^{-1},Z_a\right\rangle_R$$

Similarly we define the Hessian as

$$\partial_{s=t=0}f(e^{tZ_a+sY_b})=\langle\rho,\{I_{\overline{a}}\otimes Z_a,I_{\overline{b}}\otimes Y_b\}\rangle$$

$$\implies(\nabla^2f(I))_{aa}=\langle R_a^{-1}\rho^{\{a\}}R_a^{-1},\{Z,Y\}\rangle_R$$

$$\implies(\nabla^2f(I))_{ab}=\langle\rho^{\{a,b\}},Z\otimes Y\rangle$$

□

[AR: Not sure how to define Hessian. I think I'd like the Hessian to be

$$\sum_a E_{aa}\otimes(1\pm\epsilon)I+\sum_{a\neq b}E_{ab}\otimes\pm\lambda$$

for some small ϵ,λ. Then the Hessian will be $1-\epsilon-(k-1)\lambda$-strongly convex.]

Lemma 21 (Restatement of **??**)**.** *Let f be geodesically convex everywhere. All the below quantities are wrt metric $\langle\cdot,\cdot\rangle_R$. Assume f and λ-strongly geodesically convex ball of radius κ about I; further assume the geodesic gradient satisfies $\|\nabla f(I)\|_R=\varepsilon<\lambda\kappa$. Then there is an optimizer within an ε/λ-ball.*

Proof of **??**. The proof is exactly the same except the following:

$$g'(0)=\langle\nabla f(I),Z\rangle_R\geq-\|\nabla f(I)\|_R\|Z\|_R\geq-\varepsilon$$

□

Remark 6. *Note the perturbation lemma then gives the following strategy. By Cole's lemma, we have that $c\|\nabla f(I)\|_R\geq\|\nabla f(I)\|_{op}$. If we can say the same thing for the optimizer Z, then it is enough for $\lambda\kappa\geq\Omega(1/c)>\varepsilon$ and we can improve sample complexity to $nD>c\max_a d_a^2$.*

A similar thing is true if we can show the above inequality for the gradient flow for $\log\max_a d_a$ time.

Lemma 22. *λ-strong convexity is a sufficient condition for fast convergence of the gradient flow:*

$$-\partial_{t=0}\|\nabla f(e^{-t\nabla f(I)})\|_R^2 = -\partial_{t=0}^2 f(e^{-t\nabla f(I)}) = \langle \nabla^2 f, \nabla f \otimes \nabla f\rangle \geq \lambda \|\nabla f\|_R^2$$

[AR: Not sure how to write the third term above, the inner product with Hessian]

4.2 Old proof of ??

Proof of **??**. Take any quadratic form of the Hessian for $\{Z_a \perp I_a\}$:

$$\partial_{t=0}^2 f(e^{tZ}) = \sum_a \langle Q^a, Z_a^2\rangle + \sum_{a\neq b}\langle Q^{ab}, Z_a \otimes Z_b\rangle$$

$$\geq \sum_a \lambda_{\min}(Q^a)\|Z_a\|_F^2 - \sum_{a\neq b}\|Q^{ab}(I-P_{ab})\|_{op}\|Z_a\|_F\|Z_b\|_F$$

Now we can use our high probability bounds derived above:

$$\forall a : Q^a \succeq \frac{1-\epsilon}{d_a} I_a; \qquad \forall a \neq b : \|Q^{ab}(I-P_{ab})\|_{op} < \frac{\lambda}{\sqrt{d_a d_b}}$$

$$\implies \partial_{t=0}^2 f(e^{tZ}) \geq \sum_a \frac{1-\varepsilon}{d_a}\|Z_a\|_F^2 - \sum_{a\neq b}\frac{\lambda}{\sqrt{d_a d_b}}\|Z_a\|_F\|Z_b\|_F$$

$$\geq \sum_a \frac{1-\varepsilon+\lambda}{d_a}\|Z_a\|_F^2 - \lambda\left(\sum_a \frac{1}{\sqrt{d_a}}\|Z_a\|_F\right)^2$$

$$\geq \sum_a \frac{1-\varepsilon+\lambda}{d_a}\|Z_a\|_F^2 - k\lambda\sum_a \frac{1}{d_a}\|Z_a\|_F^2$$

$$= (1-\varepsilon-(k-1)\lambda)\|Z\|^2$$

Choosing ε, λ small enough gives the theorem. □

Proof: [CF: Akshay's conceptual proof of **??***].* We can in fact show that ∇^2 is well-conditioned using the following:

$$-\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} \preceq \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} \preceq \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

[AR: Ya ∇_{ab} notation is fine, just needed something that was a matrix of the right dimensions, so shorthand M was to avoid weird things with ρ]
[AR: It's fine if they're of different sizes, we enumerate the basis of the whole space as $\cup_a e_a \otimes \{e_{i\in[d_a]}\}$]
[CF: $E_{aa} \otimes \nabla_{aa}^2$ is $kd_a^2 \times kd_a^2$ dimensional. So how does this make sense? Maybe needs to be updated along the lines of the next proof.]

$$\nabla^2 f = \sum_a E_{aa} \otimes \nabla_{aa}^2 + \sum_{a\neq b} E_{ab} \otimes \nabla_{ab}^2$$

Now we can again use the high-probability bounds derived above: [TODO: actually cref them]

$$\nabla^2_{aa} \in \frac{1 \pm \varepsilon}{d_a}; \qquad \forall a \neq b : \|\nabla^2_{ab}\|_{op} \leq \frac{\lambda}{\sqrt{d_a d_b}} \tag{1}$$

$$\nabla^2 \preceq \sum_a E_{aa} \otimes \left(\frac{1+\varepsilon}{d_a} I_a\right) + \sum_{a<b} E_{aa} \otimes \left(\frac{\lambda}{d_a} I_a\right) + E_{bb} \otimes \left(\frac{\lambda}{d_b} I_b\right)$$

$$\preceq \sum_a E_{aa} \otimes \frac{1+\varepsilon+(k-1)\lambda}{d_a} I_a$$

The same sequence of inequalities can be reversed to show a lower bound. So in fact we can show the above bounds on blocks shows $1 + O(\varepsilon + k\lambda)$-condition number bound on the Hessian in norm $\|\cdot\|_d$. □

4.3 Steifel concentration bound

4.3.1 Gaussian Version

We will eventually compare all our bounds to gaussian inputs, so we will analyze marginal error for gaussians first.

Lemma 23. *Let $X_1, ..., X_n$ be iid from $N(0, I_d)$, then we have the following with failure probability $\leq \exp(-d/2)$*

$$\left\|\sum_i X_i X_i^* - nI_d\right\|_{op} \leq 3\sqrt{\frac{d}{n}}$$

Proof. The proof goes by a simple net argument. For an arbitrary $\xi \in S^{d-1}$ consider the MGF:

$$\log \mathbb{E} \exp t\langle \xi\xi^*, \sum_i X_i X_i^* \rangle = \sum_i \log \mathbb{E} \exp t\langle \xi, X_i\rangle^2 = \frac{-n}{2}\log(1-2t)$$

The last line is because $\langle \xi, X\rangle^2$ is distributed as a chi-squared with one degree of freedom. For small $|t| < \frac{1}{4}$, we can approximate this by the second order term:

$$\frac{-n}{2}\log(1-2t) = \frac{n}{2}\sum_{k\geq 1}\frac{(2t)^k}{k} \leq nt + 2nt^2$$

If we define $Z_\xi := \langle \xi\xi^*, \sum_i X_i X_i^* - nI_d\rangle$, then we have shown this random variable is $(4n, 4)$-subexponential. Using standard results on sub-exponential variables (i.e. Bernstein bound), we get

$$Pr\left[|\langle \xi\xi^*, \sum_i X_i X_i^* - nI_d\rangle| \geq c\sqrt{n}\right] \leq \exp(-c^2 n/8)$$

So we can simultaneously ask for this bound for $\exp(O(n))$ many unit vectors. To lift this to a result on operator norm, for a fixed instantiation of X we consider

$$\nu := \left\| \sum_i X_i X_i^* - nI_d \right\|_{op} = \sup_{\xi \in S^{d-1}} |\langle \xi\xi^*, \sum_i X_i X_i^* - nI_d \rangle|$$

Let ξ' be the optimizer, and let N be our η-net over S^d, i.e. every element of S^d is η-close to some element in N. In particular $\xi' = \xi + \delta$ where $\xi \in N, \|\delta\|_2 \leq \eta$.

$$\nu = |\langle (\xi')(\xi')^*, \sum_i X_i X_i^* - nI_d \rangle| = |\langle (\xi+\delta)(\xi+\delta)^*, \sum_i X_i X_i^* - nI_d \rangle|$$

$$\leq |\langle \xi\xi^*, \sum_i X_i X_i^* - nI_d \rangle| + (2\|\delta\|_2 + \|\delta\|_2^2)\nu$$

Rearranging terms, we see that for $\eta < \frac{1}{5}$, the sup over N gives a 2-approximation for the sup over S^{d-1}. Therefore we can choose $|N| \leq \exp(d \log 10)$:

$$Pr\left[\left\| \sum_i X_i X_i^* - nI_d \right\|_{op} \geq c\sqrt{n} \right] \leq Pr\left[\sup_{\xi \in N} Z_\xi \geq 2c\sqrt{n} \right] \leq |N| \exp(-c^2 n/2)$$

So choosing $c = 3\sqrt{d/n}$ gives the result. $\square$

4.3.2 Stiefel Version

Note if $d_a > d_b$, the error of the a marginal will be much larger than that of the b marginal. So we could try to analyze what happens after one iteration of flip-flop where we normalize the a to identity. We claim our input if our input is $X_1, ..., X_n \subseteq \mathbb{R}^{d_a \times d_b}$ iid from $N(0, I_{a \times b})$, and we normalize so that the rows of X form an orthogonal basis for $\mathbb{R}^{d_a}$, our new input Y will be uniform from the Stiefel manifold of bases for d_a dimensional subspaces of $\mathbb{R}^{nd_b}$. Given this we can try to run a net argument to bound the b marginal and compare it to the gaussian case above:

$$\mu_b(Y) = \sum_i X_i^* \left(\sum_j X_i X_i^* \right)^{-1} X_i - \frac{d_a}{d_b} I_b$$

The normalization for the Identity term comes from comparing traces. Now if we consider

$$Y := \left[\left(\sum_{i=1}^n X_i X_i^* \right)^{-1/2} X_1, ..., \left(\sum_{i=1}^n X_i X_i^* \right)^{-1/2} X_n \right]$$

as a matrix in $\mathbb{R}^{d_a \times nd_b}$, then by our claim above its rows come from orthonormal bases of a uniformly random d_a dimensional subspace of $\mathbb{R}^{nd_b}$. Therefore we can rewrite the variable we will consider a net over:

$$\langle \xi\xi^*, \sum_i X_i^* \left(\sum_i X_i X_i^*\right)^{-1} X_i \rangle = \langle Y^*Y, \xi\xi^* \otimes I_n \rangle$$

By the discussion above, Y^*Y is distributed as a uniformly random orthogonal projection on $\mathbb{R}^{nd_b}$ of rank d_a. Since $\xi\xi^* \otimes I_n$ is unitarily equivalent to any rank n orthogonal projection on $\mathbb{R}^{nd_b}$, we are left with calculating the MGF for the "overlaps" of two projections. This has already been done for us in the links commented below: specifically Lemma III.5 in `https://arxiv.org/pdf/quant-ph/0407049.pdf`, which cites Lemma II.3 in `https://arxiv.org/pdf/quant-ph/0307104.pdf` (thanks Michael!).

Theorem 24 (Lemma III.5 in `https://arxiv.org/pdf/quant-ph/0407049.pdf`). *Our ambient space is $\mathbb{R}^N$, and we consider a fixed n dimensional orthogonal projector Q. If P is a uniformly random orthogonal projector of rank a, then we have*

$$Pr[\langle P, Q\rangle \notin (1 \pm \varepsilon)\frac{an}{N}] \leq \exp(-\Omega(an\varepsilon^2))$$

So in our case we have ambient space $\mathbb{R}^{nd_b}$ and set $Q = \xi\xi^* \otimes I_n$ for any $\xi \in \mathbb{R}^{d_b}$.

4.3.3 Spreading

Because I like the vocabulary used in `https://arxiv.org/abs/2006.14009`, I will rewrite the proof for completeness using ideas of "spreading". I claim no originality, this is merely a translation of terms to fit in this theory.

[CF: Perhaps we can give some references for where spreading comes up to justify using the terminology instead of just putting it out there.]

Definition 4 (Def 2.1). *We say that random variable $Y \in \mathbb{R}^d$ is a spread of random variable $X \in \mathbb{R}^d$ if there exists a coupling such that $\mathbb{E}[Y|X] = X$.*

Another way to say this is if we can sample Y by first sampling X, then sampling a martingale.

Lemma 25 (Lemma 2.2). *If Y is a spread of X, then for any convex $\phi : \mathbb{R}^d \to \mathbb{R}$ we have (by Jensen's inequality)*

$$\mathbb{E}_Y \phi(Y) = \mathbb{E}_X \mathbb{E}[\phi(Y) \mid X] \geq \mathbb{E}_X \phi(\mathbb{E}[Y \mid X]) = \mathbb{E}_X \phi(X)$$

With this we can already compare the MGF's for gaussians vs the sphere:

Corollary 26. *If $v \sim S^{N-1}$ and $g \sim N(0, I_N)$ then for any $X \succeq 0$ we have*

$$\mathbb{E}_{S^{N-1}} \exp\langle X, vv^*\rangle \leq \mathbb{E}_{N(0,I_N)} \exp\langle X, gg^*\rangle$$

Proof. We just have to notice $gg^* = \|g\|_2^2 \frac{gg^*}{\|g\|_2^2}$, so there is a coupling where we first sample a random direction v, and then sample the magitude of the gaussian. The MGF comparison follows since $\phi(Z) := \exp\langle Z, X\rangle$ is convex for $X \succeq 0$. $\square$

The above corollary is the content of Lemma 2.3 of `https://arxiv.org/pdf/quant-ph/0307104.pdf`, and the proof strategy is exactly the same, I just added the spreading vocabulary. Similarly, since we want to compare MGFs of higher rank projections, there is a trick in Lemma 3.5 of `https://arxiv.org/pdf/quant-ph/0407049.pdf` that we can copy here with slightly different notation.

Definition 5. *As a weakening, for $K \subseteq \mathbb{R}^d$ we say $Y \in \mathbb{R}^d$ is a spread of $X \in \mathbb{R}^d$ wrt K if there exists a coupling of X, Y such that*

$$\forall \xi \in K : \langle \xi, \mathbb{E}[Y|X]\rangle = \langle \xi, X\rangle$$

Note for $K = \mathbb{R}^d$, the definitions are the same.

[CF: I would state the next lemma like "$Tr_n[uu^*]$ is a spread of VV^* wrt $L(\mathbb{R}^d)$, therefore (insert inequality)". This better motivates the additional terminology of spreading.]

Corollary 27. *Let our vector space be $\mathbb{R}^d \otimes \mathbb{R}^n$, and consider $V \in \mathbb{R}^{d\times a}$ distributed an a uniformly random rank $a < n$ isometry on $\mathbb{R}^d$, and consider u a uniformly random unit vector on $\mathbb{R}^d \otimes \mathbb{R}^n$. Then for any $L(\mathbb{R}^d) \ni X \succeq 0$:*

$$\mathbb{E}\exp\langle X, \frac{1}{a}VV^*\rangle \le \exp\langle X, Tr_n[uu^*]\rangle$$

Proof. We first exhibit the coupling between V, u. Let $v_1, ..., v_a$ be the columns of V, then lift to

$$V' := \{v_1 \otimes e_1, ..., v_a \otimes e_a\} \subseteq \mathbb{R}^d \otimes \mathbb{R}^n$$

$$u := V'1_a = \sum_i v_i \otimes e_i \in \mathbb{R}^d \otimes \mathbb{R}^n$$

Choosing $\{e_i\}$ as any orthonormal vectors does not change the partial trace [CF: use Tr not Tr] $VV^* = Tr_n[V'V'^*] = Tr_n[uu^*]$. As V runs over all possible bases, and $\{e_i\}$ runs over all possible bases, we recover u uniformly random over the sphere in $\mathbb{R}^d \otimes \mathbb{R}^n$. Finally since the partial traces are equal, we get that uu^* is a spread of $V'V'^*$ wrt $\{X \otimes I_n \mid X \in L(\mathbb{R}^d)\}$; equivalently $Tr_n[uu^*]$ is a spread of VV^* wrt $L(\mathbb{R}^d)$. The comparison of MGFs then follows by Jensen's inequality. $\square$

So we have the following sequence of inequalities for MGFs:

$$\mathbb{E}\exp\langle Y^*Y, \xi\xi^* \otimes I_n\rangle$$

$$\leq \mathbb{E}_{u\sim S^{d_a\times d_a\times nd_b}} \exp d_a\langle uu^*, \xi\xi^* \otimes I_n \otimes I_{d_a}\rangle$$

$$\leq \mathbb{E}_{g\sim N(0, I_{d_a\times d_a\times nd_b})} d_a\langle gg^*, \xi\xi^* \otimes I_n \otimes I_{d_a}\rangle$$

The desired bound then follows from the gaussian part. We use the correct normalization to be consistent with the rest of the paper below.

Corollary 28. *If $X_1, ..., X_n$ are iid from $N(0, \frac{1}{nd_a d_b} I_{d_a\times d_b})$, then let Y be the output after one iteration of flip-flop, so Y is uniformly distributed as $\frac{1}{d_a}$ times the d_a-dimensional Stiefel manifold on $\mathbb{R}^{nd_b}$. By definition $\mu_a(Y) = 0$. We have the following bound on $\mu_b(Y)$ with failure probability $\leq \exp(-\Omega(d_b))$:*

$$\left\| \sum_i Y_i^* Y_i - \frac{1}{d_b} I_b \right\|_{op} \leq O(1) \frac{1}{d_b} \sqrt{\frac{d_b}{nd_a}}$$

4.3.4 Net and union bound

[TODO: do the net bound like in the vershynin survey]