Best Separable State: Notes for SoS Seminar 2024

Samuel B. Hopkins

December 8, 2024

These notes are a work in progress and have not been subjected to the scrutiny of a peer-reviewed publication.

In these notes we will give an account of the algorithm of Barak, Kothari, and Steurer for the *best separable state* problem. Although this problem has close connections to quantum information and computation, we will discuss and solve an entirely classical version of it.

Acknowledgements This presentation is based on notes by Alan (Xinyang) Chen and was also influenced by notes of Cedric Lin, Bill Fefferman, and Anand Natarajan. Both were received through personal communication.

The proof of Lemma 6.2, the crucial quantitative calculation, was accomplished by the students and community surrounding the Fall 2024 MIT SoS seminar course. Particularly: Prashanti Anderson, Ainesh Bakshi, Kiril Bangachev, Chenghao Guo, Alina Harbuzova, Henry Hu, Saba Lepsveridze, Mahbod Majid, and Amit Rajaraman.

Degree-4 Polynomial Optimization Let P be a degree-4 polynomial in n variables $x_1, \dots, x_n$. Abusing notation, we also let $P \in \mathbb{R}^{n^2 \times n^2}$ be any matrix such that $P(x) = (x \otimes x)^\top P(x \otimes x)$. Our goal is to solve the problem $\max_{x \in \mathbb{S}^{n-1}} P(x)$. Of course this problem is NP-hard; we would like to solve it up to some kind of error.

1 *How to measure error?*

The first kind of guarantee we might hope for is multiplicative error – that is, can we find x such that $P(x) \geq \alpha \cdot OPT$ for some $\alpha > 0$? This kind of multiplicative guarantee would violate ETH for any constant.

On the other end of the spectrum, we could measure error via the Frobenius norm of P. That is, we could look for x such that $P(x) \geq OPT - \varepsilon \cdot \|P\|_F$, where $\varepsilon > 0$ is small. It turns out that this problem is very closely related to dense 4-CSPs, and can be solved in time $n^{(1/\varepsilon)^{O(1)}}$.

We get a more difficult problem if we measure error in the *operator norm* of the matrix P – that is, we look for x such that $P(x) \geq OPT - \varepsilon \cdot \|P\|_{op}$. This is more difficult because $\|P\|_{op} \leq \|P\|_F$.

Problem 1.1. *Is there a better than brute-force algorithm (ie, with running* *time* $2^{o(n)}$*) when* ε *is an arbitrarily small constant which can find* x *such that* $P(x) \geq OPT - \varepsilon \cdot \|P\|_{op}$*?*

It is known that a running time of $n^{o(\log n)}$ for this problem is unlikely – it would violate standard complexity hypotheses. But an algorithm with running time $n^{O(\log n)}$ (for a slight quantum generalization of this problem) would have major consequences for quantum computing – in particular would imply a containment of complexity classes $QMA \subseteq EXP$.

2 *The perfect completeness case*

Barak, Kothari, and Steurer in 2017 gave the first nontrivial algorithm for this degree-4 polynomial optimization problem in any regime. Their algorithm works when $OPT = \|P\|_{op}$. Or, put a different way, their algorithm can distinguish between the possibilities: $OPT = \|P\|_{op}$ and $OPT \leq (1-\varepsilon)\|P\|_{op}$, for any constant $\varepsilon > 0$. For every constant $\varepsilon > 0$, the algorithm runs in time $2^{\sqrt{n}\cdot(\log n)^{O(1)}}$, beating brute force by replacing n with $\sqrt{n}$ in the exponent. The algorithm is based on rounding $\sqrt{n} \cdot (\log n)^{O(1)}$ degree SoS.

Goal: Our goal for the rest of these notes is as follows. Given P, we want to distinguish between two cases – $OPT = \|P\|_{op}$ ("YES") and $OPT \leq (1-\varepsilon)\|P\|_{op}$ ("NO"). We think of ε as a tiny constant and hide dependences on ε in $O(\cdot)$ notation.

If $OPT = \|P\|_{op}$, then the top eigenspace of P contains a vector of the form $x \otimes x$. So it will be enough to solve the following problem.

Goal, reformulated: Given the projector Π to a subspace of $\mathbb{R}^{n^2}$, distinguish the following two cases: YES, where there exists y such that $\Pi(y \otimes y) = y \otimes y$, and NO, where every y has $\|\Pi(y \otimes y)\| \leq (1-\varepsilon)\|y \otimes y\|$.

Our plan to solve this problem is search for a pseudoexpectation $\widetilde{\mathbb{E}}$ of degree $d = \sqrt{n}(\log n)^{O(1)}$ which satisfies the constraints $\|x\|^2 = 1$ and $\Pi(x \otimes x) = x \otimes x$. If such a pseudoexpectation exists, we will output "YES", and otherwise we will output "NO".

To prove that this algorithm works, we need to show that if such $\widetilde{\mathbb{E}}$ exists, then there exists y such that $\|\Pi(y\otimes)\| \geq (1-\varepsilon)\|y \otimes y\|$. Then Π cannot be in the NO case, so it must be in the YES case.

For this task it suffices to take such $\widetilde{\mathbb{E}}$ and round it to find such y. Note that because are only aiming to solve the YES vs NO *decision* problem rather than the *search* problem of actually finding y, our rounding only needs to be "existential" – that is, we don't need to give an efficient algorithm which takes $\widetilde{\mathbb{E}}$ and extracts such y. This will simply things for us, although one can also perform the search task also in time $2^{\sqrt{n}\cdot(\log n)^{O(1)}}$.

3 *Sufficient to make $\widetilde{\mathbb{E}} xx^\top$ approximately rank-one*

We first formulate a sufficient condition on $\widetilde{\mathbb{E}}$ which would make it easy to round. This condition is called "approximately rank one". Our goal afterwards will be to find a procedure which can make $\widetilde{\mathbb{E}}$ approximately rank one.

Lemma 3.1 (Approximately rank one suffices)**.** *Suppose that $\widetilde{\mathbb{E}} xx^\top$ has the property that $\lambda_{\max}(\widetilde{\mathbb{E}} xx^\top) \geq (1-\varepsilon)\|\widetilde{\mathbb{E}} xx^\top\|_F$. (And suppose that $\widetilde{\mathbb{E}}$ satisfies $\|x\|^2 = 1$ and $\Pi(x \otimes x) = x \otimes x$.) Then the maximum eigenvector y of $\widetilde{\mathbb{E}} xx^\top$ satisfies $\|\Pi(y \otimes y)\| \geq 1 - O(\sqrt{\varepsilon})$.*

Proof. Under the hypotheses of the lemma, if y is the (unit norm) top eigenvector of $\widetilde{\mathbb{E}} xx^\top$, with eigenvalue λ, then $\|\lambda yy^\top - \widetilde{\mathbb{E}} xx^\top\|_F \leq O(\sqrt{\varepsilon} \cdot \lambda)$. To see this, we expand

$$
\begin{aligned}
\|\lambda yy^\top - \widetilde{\mathbb{E}} xx^\top\|_F^2 &= \lambda^2 - 2\lambda \langle yy^\top, \widetilde{\mathbb{E}} xx^\top \rangle + \|\widetilde{\mathbb{E}} xx^\top\|_F^2 \\
&= \lambda^2 - 2\lambda^2 + \|\widetilde{\mathbb{E}} xx^\top\|_F^2 \\
&= \|\widetilde{\mathbb{E}} xx^\top\|_F^2 - \lambda^2 \\
&\leq (1 + O(\varepsilon))\lambda^2 - \lambda^2 \\
&\leq O(\varepsilon)\lambda^2 .
\end{aligned}
$$

So consider

$$
\|\Pi(y \otimes y)\| = \|\Pi(\tfrac{1}{\lambda} \widetilde{\mathbb{E}} x \otimes x - E)\|
$$

where $\|E\|$ is a vector with Euclidean norm $O(\sqrt{\varepsilon})$. This is at least $\|\Pi(\frac{1}{\lambda}\widetilde{\mathbb{E}} x \otimes x)\| - O(\sqrt{\varepsilon})$. Now, since $\widetilde{\mathbb{E}}$ satisfies $\Pi(x \otimes x) = x \otimes x$, we have $\|\Pi \widetilde{\mathbb{E}}(x \otimes x)\| = \|\widetilde{\mathbb{E}} x \otimes x\|$. And $\lambda \leq \|\widetilde{\mathbb{E}} x \otimes x\|_F$, which completes the proof. □

4 *Example: the Uniform Distribution*

Before we describe a procedure which can take a pseudoexpectation and produce a new one which satisfies the hypothesis of Lemma 3.1, we describe some intuition. As a thought experiment, imagine that $\widetilde{\mathbb{E}}$ is the uniform distribution on the unit sphere in n dimensions. The second moments are $\mathbb{E}_{x \sim \mathbb{S}^{n-1}} xx^\top = \frac{1}{n}\mathrm{Id}$.

If we wanted to condition on an event such that the conditional distribution satisfies Lemma 3.1, we could try and condition on x being in a spherical cap around some vector g. This should create a large eigenvalue in the g direction in the second moment matrix. Without loss of generality, let us take $g = e_1$, the first coordinate direction.

Then we are interested in the second moments $\mathbb{E}[xx^\top \mid |x_1| \geq t]$. How large do we have to make t in order to satisfy Lemma 3.1? By rotational symmetry, the conditional second moments are

$$\mathbb{E}[xx^\top \mid |x_1| \geq t] = \operatorname{diag}(t^2, \tfrac{1-t^2}{n-1}, \ldots, \tfrac{1-t^2}{n-1})\,.$$

The top eigenvalue is t^2, and the Frobenius norm is roughly $\sqrt{t^4 + 1/n}$. So, the second moments satisfy Lemma 3.1 if we take $t \gg n^{-1/4}$.

Now let us estimate, roughly, the probability of the event we are conditioning on here. Since x_1 is roughly Gaussian $\mathcal{N}(0, 1/n)$, we have $\Pr(|x_1| > n^{-1/4}) \approx \exp(-\sqrt{n})$.

The next crucial piece of intuition is that *it should be possible to "approximately" condition a degree* $\approx \sqrt{n}$ ***pseudoexpectation*** *on an event of probability* $\exp(-\sqrt{n})$. There is nothing special about $\sqrt{n}$ here – a good intuition is that degree $\approx k$ is sufficient to "mimic" conditioning on an event of probability $\approx 2^{-k}$.

5 *Reweighing*

How do we condition a pseudoexpectation on an event like "$x_1 \geq t$"? Unlike in the Boolean setting, where we were able to develop a direct analogue for pseudoexpectations of conditioning on events like $x_1 = 1$ or $x_1 = -1$, here we will not get an exact analogue of conditioning. Instead, we will approximate conditioning by "reweighing".

5.1 *Reweighing a probability distribution*

Suppose that μ is a probability distribution on a domain Ω, and w is any nonnegative function. Then there is a "reweighed" distribution μ' given by $\mu'(x) \propto \mu(x)w(x)$. This distribution will put more weight on x where w is relatively large.

If we have a function f, we can write its expectation under the reweighed distribution as $\mathbb{E}_{\mu'} f(x) = \mathbb{E}_\mu w(x)f(x) / \mathbb{E}_\mu w(x)$. If w is the indicator for some event, then this is equivalent to conditioning on that event.

5.2 *Reweighing a pseudoexpectation*

We can reweigh a pseudoexpectation using the same formula. The only modification is that the reweighing function must be a sum of squares. If w is a degree-r SoS and $\tilde{\mathbb{E}}$ is a degree-d pseudoexpectation in variables x, then we can define a new linear operator $\tilde{\mathbb{E}}'$ by

$$\tilde{\mathbb{E}}' p(x) = \frac{\tilde{\mathbb{E}} w(x) p(x)}{\tilde{\mathbb{E}} w(x)}\,.$$

It is easy to check that $\tilde{\mathbb{E}}'$ is a degree $d - r$ pseudoexpectation. Furthermore, if $\tilde{\mathbb{E}} \vDash q(x) \geq 0$, then also $\tilde{\mathbb{E}}' \vDash q(x) \geq 0$.

5.3 Reweighing the uniform distribution on $\mathbb{S}^{n-1}$

Continuing our example from before, let us show that there is a degree $O(\sqrt{n})$-degree reweighing of the uniform distribution on $\mathbb{S}^{n-1}$ which has the approximate rank-one property.

Let μ be uniform on $\mathbb{S}^{n-1}$. Let $g \sim \mathcal{N}(0, I)$. Consider the reweigh given by $\langle x, g\rangle^k$. We claim that with positive probability over g,

$$\frac{\mathbb{E}_x \langle x, g\rangle^{k+2}}{\mathbb{E}_x \langle x, g\rangle^k \|g\|^2} \geq \frac{k+2}{n}$$

and hence taking $k \approx \sqrt{n}$ is good enough for us. It will be enough to show that

$$\mathbb{E}_g[\mathbb{E}_x \langle x, g\rangle^{k+2}] \geq \frac{k+2}{n} \mathbb{E}_{x,g} \langle x, g\rangle^k \|g\|^2 .$$

Both left and right hand sides are simple calculations. The left hand side is simply $(k+2)!!$, the $k+2$-nd moment of a standard Gaussian. The right hand side is given by

$$\mathbb{E}_g \langle x, g\rangle^k (\langle x, g\rangle^2 + \|g'\|^2)$$

where g' is an $n-1$-dimensional standard Gaussian independent of $\langle x, g\rangle$. This is in turn equal to $(k+2)!! + (n-1) \cdot k!!$. So the ratio is approximately $(k+2)/(n-1)$, as we wanted.

This gives a lower bound on the top eigenvalue. Of course, the Frobenius norm might also have changed. When we do the analysis for real (on pseudoexpectations) we will need to address this issue.

6 Making a pseudoexpectation approximately rank-one via reweighing

Now to the heart of things: we show that there is a degree $O(\sqrt{n} \log n)$ reweighing which makes a pseudoexpectation approximately rank one.

6.1 Stabilizing the Frobenius Norm

We can assume that no degree-k reweighing can increase the Frobenius norm $\|\tilde{\mathbb{E}} xx^\top\|_F$ by more than a multiplicative factor of $(1+\varepsilon)$. Because, if such a reweighing did exist, we could apply it to increase the Frobenius norm by this factor $(1+\varepsilon)$. Since $1/\sqrt{n} \leq \|\tilde{\mathbb{E}} xx^\top\|_F \leq 1$, this can happen only $O(\log n)$ times.

6.2 Reweighing by random vector

Our plan is to reweigh using the function $\langle x, g\rangle^a$ where $a \leq k$ is an (even) integer and $g \sim \mathcal{N}(0, \tilde{\mathbb{E}} xx^\top)$.

Lemma 6.1. *Let $\varepsilon > 0$. Suppose $\widetilde{\mathbb{E}}$ is a pseudoexpectation of degree $k+2 = O(\sqrt{n})$ satisfying $\|x\|^2 = 1$. And suppose that every degree $\leq k$ reweighing $\widetilde{\mathbb{E}}'$ of $\widetilde{\mathbb{E}}$ has $\|\widetilde{\mathbb{E}}' xx^\top\|_F \leq (1+\varepsilon)\|\widetilde{\mathbb{E}} xx^\top\|_F$. Then then there exists a degree $\leq k$ reweighing $\widetilde{\mathbb{E}}'$ of $\widetilde{\mathbb{E}}$ such that*

$$\lambda_{\max}(\widetilde{\mathbb{E}}') \geq (1-\varepsilon)\|\widetilde{\mathbb{E}}' xx^\top\|_F \,.$$

Proof. It will be enough to show that there exists $a \leq k$ such that with positive probability over $g \sim \mathcal{N}(0, \widetilde{\mathbb{E}} xx^\top)$, if we reweigh $\widetilde{\mathbb{E}}$ using $\langle x, g\rangle^a$ to obtain $\widetilde{\mathbb{E}}'$, we get

$$\lambda_{\max}(\widetilde{\mathbb{E}}' xx^\top) \geq (1-\varepsilon)\|\widetilde{\mathbb{E}} xx^\top\|_F \,.$$

A lower bound on $\lambda_{\max}(\widetilde{\mathbb{E}}' xx^\top)$ is given by

$$\lambda_{\max}(\widetilde{\mathbb{E}}' xx^\top) \geq \frac{\widetilde{\mathbb{E}}' \langle x, g\rangle^2}{\|g\|^2} = \frac{\widetilde{\mathbb{E}} \langle x, g\rangle^{a+2}}{\widetilde{\mathbb{E}} \langle x, g\rangle^a \cdot \|g\|^2} \,.$$

By taking the product of these lower bounds from $a = 0$ to $a = k-2$ and telescoping the product, it's enough to show that with positive probability,

$$\frac{\widetilde{\mathbb{E}} \langle x, g\rangle^k}{\|g\|^k} \geq (1-\varepsilon)^{k/2} \cdot \|\widetilde{\mathbb{E}} xx^\top\|_F^{k/2}$$

since then by taking $k/2$-th roots on both sides we would find that some a satisfies the desired lower bound. Since we only need positive probability, it's enough to show

$$\mathbb{E}_g \widetilde{\mathbb{E}} \langle x, g\rangle^k \geq (1-\varepsilon)^{k/2} \cdot \|\widetilde{\mathbb{E}} xx^\top\|_F^{k/2} \cdot \mathbb{E}_g \|g\|^k \,.$$

The left-hand side satisfies

$$\begin{aligned} \widetilde{\mathbb{E}}\left[\mathbb{E}_g \langle x, g\rangle^k\right] &= M_k \cdot \widetilde{\mathbb{E}}\left[(\mathbb{E}_g \langle x, g\rangle^2)^{k/2}\right] \\ &\geq M_k \cdot (\widetilde{\mathbb{E}}\, \mathbb{E}_g \langle x, g\rangle^2)^{k/2} \\ &= M_k \cdot \|\widetilde{\mathbb{E}} xx^\top\|_F^k \end{aligned}$$

where $M_k = (k-1)!!$ is the k-th Gaussian moment.

What about the right-hand side?[1] Lemma 6.2, which follows this proof, shows that

$$\mathbb{E}\, \|g\|^k \leq \sum_{p \leq k/2} \binom{k/2}{p} (2p-1)!! \cdot \|\widetilde{\mathbb{E}} xx^\top\|_F^p \,.$$

We claim that there exists $C(\varepsilon)$ so that if $k \geq C(\varepsilon)/\|\widetilde{\mathbb{E}} xx^\top\|_F$ then

$$\sum_{p \leq k/2} \binom{k/2}{p} (2p-1)!! \cdot \|\widetilde{\mathbb{E}} xx^\top\|_F^p \leq (1+\varepsilon)^{k/2} (k-1)!! \cdot \|\widetilde{\mathbb{E}} xx^\top\|_F^{k/2} \,,$$

[1] Note that $\mathbb{E}\, \|g\|^2 = \widetilde{\mathbb{E}} \|x\|^2 = 1$. If $\|g\|^2$ "acts like" its expectation, we would have $\mathbb{E}\, \|g\|^k \approx 1$ as well (of course when k gets large this won't be true). In this case, we would just need $M_k \cdot \|\widetilde{\mathbb{E}} xx^\top\|_F^{k/2} \geq (1-\varepsilon)^{k/2}$, for which it suffices to choose $k = C\sqrt{n}$ for a large enough constant C.

which will complete the proof. There are only k terms in the sum, so it will be enough to show that for every p,

$$\binom{k/2}{p}(2p-1)!! \le (1+\varepsilon)^{k/2-o(k)}(k-1)!! \cdot \|\widetilde{\mathbb{E}}xx^\top\|_F^{k/2-p}\,. \qquad (6.1)$$

For some $\delta = \delta(\varepsilon)$ to be chosen later, we split into two cases.

Large p case: First, consider a term where $p \ge (1-\delta)\frac{k}{2}$. In this case, $(k-1)!!/(2p-1)!! \ge (k/2)^{k/2-p}$. So,

$$\frac{\binom{k/2}{p}(2p-1)!!}{(k-1)!!\|\widetilde{\mathbb{E}}xx^\top\|_F^{k/2-p}} \le \binom{k/2}{p} \cdot \frac{1}{(k/2)^{k/2-p}\|\widetilde{\mathbb{E}}xx^\top\|_F^{k/2-p}}$$

If $k/2 \ge \|\widetilde{\mathbb{E}}xx^\top\|_F^{-1}$, then the last term is at most 1, so the whole thing is at most

$$\binom{k/2}{p} \le 2^{kH(\delta)}\,.$$

Here $H(\delta)$ is the binary entropy function. This is at most $(1+\varepsilon)^{k-o(k)}$ as long as we choose δ such that $H(\delta) \le \delta \log \frac{1}{\delta} \le \varepsilon/100$, proving (6.1) for the large-p case.

Small p case: Now we need to handle the case $p \le (1-\delta)k/2$, where δ is chosen as above. Suppose $p = (1-\gamma)k/2$ for $\gamma \ge \delta$. Then we can use the bounds $\binom{k/2}{p} \le 2^{(k/2)H(\gamma)}$ and $(2p-1)!! \le 2^{o(k)} \cdot 2^{(1-\gamma)k/2}(k/(2e))^{(1-\gamma)k/2}$ to obtain

$$\binom{k/2}{p}(2p-1)!! \le 2^{(k/2)(H(\gamma)+o(1))} \cdot (k/e)^{(1-\gamma)k/2}\,.$$

At the same time $(k-1)!! \ge 2^{-o(k)}2^{k/2}(k/2e)^{k/2} = 2^{-o(k)}(k/e)^{k/2}$. So

$$\frac{\binom{k/2}{p}(2p-1)!!}{(k-1)!!} \le 2^{o(k)} \cdot 2^{(k/2)H(\gamma)} \cdot (k/e)^{-\gamma k/2}\,.$$

We just need the RHS of the above to be less than $(1+\varepsilon)^{k/2-o(k)}\|\widetilde{\mathbb{E}}xx^\top\|_F^{\gamma k/2}$. For this it is enough to choose $k \ge C\gamma^{-1}\|\widetilde{\mathbb{E}}xx^\top\|_F^{-1}$ for a big-enough constant C. Since $\gamma \le \delta = \delta(\varepsilon)$, this completes the analysis. □

Lemma 6.2. *Let $g \sim \mathcal{N}(0, \Sigma)$. For every even $n \in \mathbb{N}$,*

$$\mathbb{E}\,\|g\|^{2n} \le \sum_{p \le k} \binom{n}{p}(2p-1)!! \cdot (\operatorname{Tr}\Sigma)^{n-p} \cdot \|\Sigma\|_F^p\,.$$

Proof. Generally, one can write down moments of a random variable in terms of its cumulant. Cumulants κ_n are the coefficients of cumulant-generating function $K(t)$ given by

$$K(t) \triangleq \log \mathbb{E}\, e^{tX} = \sum_{n=1}^{\infty} \kappa_n \frac{t^n}{n!}.$$

It is well known (see wikipedia) that the formula for n-th moment given cumulants is given by

$$\mathbb{E}\, X^n = \sum_{\pi \in \Pi} \prod_{B \in \pi} \kappa_{|B|}$$

Here Π is the set of all set partitions of $[n]$ and $B \in \pi$ are parts.

Suppose $g \sim \mathcal{N}(0, \Sigma_d)$ where Σ has eigenvalues $\lambda_1, \dots, \lambda_d$. We can compute the cumulants as follows:

$$\kappa_n(\|g\|^2) = \kappa_n(\sum_{j=1}^n \lambda_j z_j^2) = \sum_{j=1}^n \lambda_j^n \kappa_n(z_j^2) = \sum_{j=1}^n \lambda_j^n 2^{n-1}(n-1)! = 2^{n-1}(n-1)!\,\mathrm{Tr}(\Sigma^n),$$

here $z_1, \dots, z_d \sim \mathcal{N}(0,1)$ are iid normals. We can now use this to see that

$$\mathbb{E}(\|g\|^{2n}) = \sum_{\pi \in \Pi} \prod_{B \in \pi} 2^{|B|-1}(|B|-1)!\,\mathrm{Tr}(\Sigma^{|B|})$$

Let us make a little digression. Note that if we apply this formula for $g \sim \mathcal{N}(0,1)$ we get that

$$(2n-1)!! = \mathbb{E}(g^{2n}) = \sum_{\pi \in \Pi} \prod_{B \in \pi} 2^{|B|-1}(|B|-1)!$$

Using this fact, we can upper bound[2]

$$\sum_{\substack{\pi \in \Pi \\ p \text{ singletons}}} \prod_{B \in \pi} 2^{|B|-1}(|B|-1)! \le \binom{n}{p}(2(n-p)-1)!!$$

Right hand side counts number of ways to choose p singletons and then partition the rest.

To finish, whenever $k \ge 2$ we bound $\mathrm{Tr}(\Sigma^k) \le \mathrm{Tr}(\Sigma^2)^{k/2} = \|\Sigma\|_F^k$.

$$\begin{aligned}
\mathbb{E}(\|g\|^{2n}) &= \sum_{\pi \in \Pi} \prod_{B \in \pi} 2^{|B|-1}(|B|-1)!\,\mathrm{Tr}(\Sigma^{|B|}) \\
&= \sum_{p=0}^n \sum_{\substack{\pi \in \Pi \\ p \text{ singletons}}} \prod_{B \in \pi} 2^{|B|-1}(|B|-1)!\,\mathrm{Tr}(\Sigma^{|B|}) \\
&\le \sum_{p=0}^n \mathrm{Tr}(\Sigma)^p \|\Sigma\|_F^{n-p} \sum_{\substack{\pi \in \Pi \\ p \text{ singletons}}} \prod_{B \in \pi} 2^{|B|-1}(|B|-1)! \\
&\le \sum_{p=0}^n \binom{n}{p}(2(n-p)-1)!!\,\mathrm{Tr}(\Sigma)^p \|\Sigma\|_F^{n-p} \\
&= \sum_{p=0}^n \binom{n}{p}(2p-1)!! \|\Sigma\|_F^p\,\mathrm{Tr}(\Sigma)^{n-p}.
\end{aligned}$$

□

[2] The upper bound is not very lossy as the sum is lower bounded by $\binom{n}{p}2^{n-p-1}(n-p-1)!$.