矩阵迹的凸性不等式

虚空若叶睦

2025 年 8 月 26 日

> 给定实数 $p \geqslant 1$. 对任何半正定的 Hermite 矩阵 $A, B \in M_n(\mathbb{C})$, 证明:
>
> $$\operatorname{tr}\big[(A+B)^p\big] \geqslant \operatorname{tr}\big[A^p\big] + \operatorname{tr}\big[B^p\big].$$

第一幕: p 是正整数

某日, 若叶睦在茶话会上看到 MyGO!!!!! 和 Ave Mujica 的少女们在讨论这样一个矩阵不等式: 对任意半正定的 Hermite 矩阵 $A, B \in M_n(\mathbb{C})$, 证明:

$$\operatorname{tr}\big[(A+B)^{2025}\big] \geqslant \operatorname{tr}\big[A^{2025}\big] + \operatorname{tr}\big[B^{2025}\big].$$

大家讨论的氛围热闹非凡, Karamata 不等式, Jensen 不等式, Hölder 不等式等各种工具都用上了, 却都似乎没法解决这个问题. 若叶睦看着这个不等式, 缓缓开口道:

“这个不等式有初等且简单的证法. 可以把指数上的 2025 换成一般的正整数 p.”

大家瞬间安静下来, 向小睦投来讶异的目光, 连忙询问如何求解这个问题. 小睦说道:

“第一步, 不妨设 $B = E_{11}$. 第二步, 用二项式定理.”

这里 E_{11} 是只有第一行第一列的元素为 1, 其它元素全为 0 的 $n \times n$ 矩阵. 大家似懂非懂地点了点头, 脑海里却还是一头雾水. 八幡海铃率先开口了:

“这第一步我就不明白. 为什么可以假设 $B = E_{11}$?”

小睦眨巴了一下眼睛, 从手边的袋子里拿出了一根黄瓜和一把水果刀, 然后仔细地把黄瓜切成了十份. 小睦亲手种下的黄瓜, 是茶话会的大家平时最喜爱的食物之一. 小睦随后说道:

“首先, B 是可以正交对角化的, 所以不妨设 $B=\operatorname{diag}\{\lambda_1,\cdots,\lambda_n\}$. 这样, $\lambda_1,\cdots,\lambda_n$ 都是非负的, 并且原不等式等价于

$$\operatorname{tr}\left(A+\begin{pmatrix}\lambda_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \lambda_n\end{pmatrix}\right)^p \geqslant \operatorname{tr}\big[A^p\big]+\sum_{i=1}^n \lambda_i^p. \tag{1.1}$$

接下来只需要证明 (1.1).”

八幡海铃忽然眼睛一亮, 轻轻张开了口, 露出稍微吃惊的表情, 浑然不知身边的椎名立希一直在盯着她看. 小睦走向了海铃, 冷不丁地把黄瓜块塞进了海铃的嘴里. 这一动作着实把立希吓了一大跳, 立希抓住小睦的手臂, 并对小睦说:

“等一下小睦, 海铃现在是我的女... 唔唔?!”

立希的嘴里也被小睦塞了一块黄瓜. 小睦微笑着向大家说:

“别着急哦, 大家都有份. 我会一个个喂给你们吃哦.”

小睦不紧不慢地继续说:

“接下来, 如果原不等式在 $B=E_{11}$ 的情况下成立, 那么也应该在 $B=\lambda_i E_{ii}$ 的情形下成立, 其中 $i=1,\cdots,n$. 由于 A 可以取任何半正定矩阵, 所以可以将 $\lambda_n,\cdots,\lambda_1$ 依次加到 A 上, 并得到下面的一串不等式:

$$\operatorname{tr}\left(A+\begin{pmatrix}\lambda_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \lambda_n\end{pmatrix}\right)^p \geqslant \operatorname{tr}\left(A+\begin{pmatrix}0 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \lambda_n\end{pmatrix}\right)^p+\lambda_1^p \geqslant \operatorname{tr}\big[A^p\big]+\sum_{i=1}^n \lambda_i^p.$$

这样就得到了 (1.1). ”

大家对小睦的论证很是满意, 这样的确可以把原问题化简到一个相当简单的情形.

“接下来要证明的是 $B=E_{11}$ 的情形, 也就是不等式

$$\operatorname{tr}\big[(A+E_{11})^p\big] \geqslant \operatorname{tr}\big[A^p\big]+1. \tag{1.2}$$

刚刚有提到过, 这个可以利用二项式定理展开来证明.”

长崎素世满意地咽下了小睦送来的黄瓜, 并说道:

“二项式定理在 a, b 都是实数的情况下我是知道的, 也就是

$$(a+b)^p = \sum_{k=0}^{p} \binom{p}{k} a^k b^{p-k}.$$

但是, 如果对于矩阵的情形, A, B 不一定交换, 如何保证定理的正确性呢? ”

小睦缓缓走到素世的身边, 千早爱音便警觉地挽起了素世的胳膊, 似乎是在向小睦宣誓自己的主权. 小睦只是向爱音眯起眼睛笑了笑, 便向素世说:

“很好的观察. 不过, 这里并不会直接用到二项式系数. 由于 $E_{11}^2 = E_{11}$, 所以 $(A+E_{11})^p$ 的展开式中的每一项实际上形如

$$A^{p_1} E_{11} \cdots A^{p_k} E_{11}, \tag{1.3}$$

其中 $p_1, \cdots, p_k$ 为非负整数. 我们只要证明每个形如 (1.3) 的矩阵的迹是非负的就可以了. 注意到 $E_{11} = e_1 e_1^\dagger$, 所以 (1.3) 的迹就等于

$$\operatorname{tr}\left[A_1^{p_1} e_1 e_1^\dagger \cdots A_k^{p_k} e_1 e_1^\dagger\right] = \operatorname{tr}\left[(e_1^\dagger A^{p_1} e_1) \cdots (e_1^\dagger A^{p_k} e_1)\right] \geqslant 0.$$

这样就得到了原问题的证明.”

素世对于爱音的亲密举动毫不介意, 似乎已经习惯了她们之间的关系. 素世继续说:

“我明白了. 因为 A^{p_1} 是半正定的, 所以它的第一行第一列 $e_1^\dagger A^{p_1} e_1$ 也是非负的.”

爱音认真听完两人的讨论, 高声欢呼起来:

“SO! DA! YO! 素世小睦你们好厉害呀! 哇啊!! 痛痛痛!”

爱音似乎是被身边的人重重地掐了一下, 呜呜了好一阵才安静下来.

这时, 茶话会的角落里传来了一声清脆的咖啡杯放在玻璃桌子上的声音. 一位梳着天蓝色长发, 戴着舞会假面的大小姐皱着眉头, 用她缠绕着黑色蕾丝的右手轻轻揉捻着刚刚小睦送来的黄瓜. 这位大小姐, 明明还没有开始吃却似乎已沉浸在无穷的回味中. 即便她还没有摘下面具, 大家也早已知道, 她就是丰川家族的大小姐, Ave Mujica 的核心成员, 丰川祥子. 她用略显高傲的声音问道:

“小睦, 你的解答我很满意. 但我还有一个问题: 如果 p 不是整数, 只有 $p \geqslant 1$ 是实数的条件, 应该如何证明这个不等式?”

第二幕: $p \geqslant 2$ 是实数

面对来自 ob 一串字母小姐突然提出的问题, 若叶睦却一下子慌了神, 在大脑里开始飞速运转着 p 是实数的情形的证明方法, 而一旁的丰川祥子只是继续淡淡地看着她.

“Karamata 不等式? Jensen 不等式? Mortis 不等式? Hölder 不等式? 都不对. 等等, Mortis, Mortis?? 啊哈哈... 我又被唤醒了啊...”

若叶睦在一瞬间瘫坐在身后的椅子上, 并突然把头转向正上方, 向天空露出了一抹诡异的微笑. 要乐奈吸完了抹茶巴菲里的最后一点奶油, 舔了舔嘴角, 再次说出了那句名言:

“有趣的女人.”

丰川祥子对眼前的场景并不感到奇怪, 因为她知道若叶睦体内住着一个奇怪的人格. 祥子还是用略显高傲的语气对小睦说道:

“你就是 Mortis, 对吧. 说说看你有什么好的想法吧.”

祥子端起还没喝完的咖啡, 正准备继续品尝时, 却没注意到小睦正带着诡异的微笑快步向她走来. 小睦低下头, 用左手掀起了祥子的刘海, 并主动靠近祥子的脸, 两人的鼻子都几乎要贴在一起了. 祥子满脸通红地望着小睦, 即使是她也没想到, 平时文静乖巧又可爱的小睦竟然会对她做出如此大胆的操作. 坐在祥子身旁的三角初华早已恨得牙痒痒, 但碍于祥子本人的在场, 她不得不暂时先忍下这口气. 小睦笑盈盈地对祥子说:

“我知道你很想知道答案. 我会变成这个样子, 也都是因为你哦. 这道题, 现在需要一点矩阵微分的知识.”

说着说着, 小睦的手指从祥子光滑的颈部慢慢滑向胸口, 然后便开始玩弄祥子胸前的那对巨大的黑色蝴蝶结. 似乎是察觉到身边的三角初华的怒气值已经达到满格, 小睦迅速起身并回到了茶话会的中央区域. 她随即向大家开启了自己的表演:

“我现在有一个解法, 虽然它只能解决 $p \geqslant 2$ 的情形. 现在, 我们只需要证明: 对任何半正定矩阵 A, 都有

$$\operatorname{tr}\big[(A+E_{11})^p\big] \geqslant \operatorname{tr}\big[A^p\big] + 1. \tag{2.1}$$

然后利用和 p 整数情形一样的归纳法, 就可以证明一般情形的不等式成立."

祐天寺若麦对转换人格之后的小睦非常感兴趣, 举起手说:

"好的好的! 那么要怎么用矩阵微分证明 (2.1) 呢?"

小睦微笑着举起了她的手, 回应着若麦的问题. 小睦继续说:

"定义一个实值函数 $g(t) = \mathrm{tr}\big[(A + tE_{11})^p\big]$. 那么, $g(t)$ 的导数将是

$$g'(t) = p \cdot \mathrm{tr}\big[(A + tE_{11})^{p-1}E_{11}\big] \geqslant 0. \tag{2.2}$$

因此 $g(t)$ 是单调递增的."

还没等大家开始提问, 小睦便继续说:

"我知道你们会好奇, 为什么 A 和 E_{11} 不一定交换而仍可正常使用链式法则. 这实际上是因为迹函数的循环性. 具体地说, 假设 $f : [0, +\infty) \to \mathbb{R}$ 是一个连续可微函数, 而 $X(t)$ 是可微的正定矩阵函数, 那么就有

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{tr}\big[f(X(t))\big] = \mathrm{tr}\big[f'(X(t))X'(t)\big]. \tag{2.3}$$

这里, 当 $X = P\,\mathrm{diag}\{\lambda_1, \cdots, \lambda_n\}P^\dagger$ 是半正定矩阵的时候, $f(X)$ 的定义是

$$f(X) = P\,\mathrm{diag}\big\{f(\lambda_1), \cdots, f(\lambda_n)\big\}P^\dagger.$$

这是把连续函数作用在矩阵上时的运算规则. "

由于大家对矩阵微分都不太熟悉, 对于公式 (2.3) 也是半信半疑地听着.

"首先, (2.3) 有一种直观的理解方式. 假设 $X(t)$ 的特征值为 $\lambda_1(t), \cdots, \lambda_n(t)$, 并且 $X(t)$ 有连续的谱分解

$$X(t) = \sum_{i=1}^{n} \lambda_i(t)u_i(t)u_i(t)^\dagger, \tag{2.4}$$

其中 $u_i(t)$ 为 $\lambda_i(t)$ 所对的单位特征向量. 于是

$$\mathrm{tr}\big[f(X(t))\big] = \sum_{i=1}^{n} f(\lambda_i(t)) \Longrightarrow \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{tr}\big[f(X(t))\big] = \sum_{i=1}^{n} f'(\lambda_i(t))\lambda_i'(t).$$

在 (2.4) 两端对 t 求导, 并利用 $u_i(t)$ 和 $u_i'(t)$ 及其它的 $u_j(t)$ 的正交性有

$$\lambda_i'(t) = u_i(t)^\dagger X'(t) u_i(t).$$

综合以上两个等式即可得到

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{tr}\big[f(X(t))\big] &= \sum_{i=1}^{n} f'(\lambda_i(t)) u_i(t)^\dagger X'(t) u_i(t)\\ &= \mathrm{tr}\Bigg[\sum_{i=1}^{n} f'(\lambda_i(t)) u_i(t) u_i(t)^\dagger X'(t)\Bigg] = \mathrm{tr}\big[f'(X(t))X'(t)\big].\end{aligned}$$

从而 (2.3) 得证. 不过, (2.3) 严格的证明还是需要直接利用导数的定义. 由于 f 连续可微, 可以 f 可以在任意有界区间内被多项式 C^1 逼近. 因此可以不妨设 $f(x) = x^k$, 其中 k 为正整数. 此时 (2.3) 即等价于

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{tr}\big[X^k(t)\big] = k \cdot \mathrm{tr}\big[X^{k-1}(t)X'(t)\big]. \tag{2.5}$$

当 h 充分小时, 由迹函数的循环性有

$$\begin{aligned}&\mathrm{tr}\big[X^k(t+h)\big] - \mathrm{tr}\big[X^k(t)\big]\\ &= \mathrm{tr}\big[(X(t)+hX'(t))^k - X^k(t)\big] + o(h)\\ &= h\cdot \mathrm{tr}\Bigg[\sum_{i=0}^{k-1} X^i(t)X'(t)X^{k-1-i}(t)\Bigg] + o(h)\\ &= kh\cdot \mathrm{tr}\big[X^{k-1}(t)X'(t)\big] + o(h).\end{aligned}$$

因此有 (2.5) 得证. 因此在迹函数内部是可以自由使用链式法则的. ”

若麦继续为小睦喝彩着, 而丰川祥子也久违地露出了微笑. 她心想:

> “不愧是我选中的骑士. 无论是若叶睦还是 Mortis, 她们都具有极其强大的潜质, 并将帮我实现心愿. 我要把 Ave Mujica 打造成为最强大的数学组合, 而最终的舞台将是: 国际少女数学家大会.”

若叶睦还在继续她的讲解. 之前定义了函数 $g(t) = \mathrm{tr}\big[(A+tE_{11})^p\big]$, 现在来看看 $g'(t)$ 的表达式 (也就是等式 (2.2)) 能带来什么. 小睦继续说道:

“现在, 来介绍一个 Jensen 不等式的矩阵版本. 设 $f:[0,+\infty)\to\mathbb{R}$ 是凸函数, A 为半正定矩阵, $v\in\mathbb{C}^n$ 为任意单位向量, 则有

$$v^\dagger f(A)v \geqslant f(v^\dagger Av). \tag{2.6}$$

这个不等式的证明用谱分解就可以了. 若设 $A=\sum_{i=1}^n \lambda_i u_i u_i^\dagger$, 则 (2.6) 等价于

$$\sum_{i=1}^n |u_i^\dagger v|^2 f(\lambda_i) \geqslant f\biggl(\sum_{i=1}^n |u_i^\dagger v|^2\lambda_i\biggr).$$

由于 $\displaystyle\sum_{i=1}^n |u_i^\dagger v|^2=|v|^2=1$, 这就是普通的 Jensen 不等式. 于是 (2.6) 得证.

最后, 使用微积分基本定理就能得到最终的结果了. 利用 (2.2)(2.6) 可以得到

$$\begin{aligned}
&\operatorname{tr}\bigl[(A+E_{11})^p\bigr]-\operatorname{tr}\bigl[A^p\bigr]\\
={}&g(1)-g(0)=\int_0^1 g'(t)\mathrm{d}t && \text{(微积分基本定理)}\\
={}&p\int_0^1 \operatorname{tr}\bigl[(A+tE_{11})^{p-1}E_{11}\bigr]\mathrm{d}t && \text{($g'(t)$ 在 (2.2) 中的表达式)}\\
={}&p\int_0^1 e_1^\dagger(A+t)^{p-1}e_1\mathrm{d}t && \text{(迹的循环性)}\\
\geqslant{}&p\int_0^1 (A_{11}+t)^{p-1}\mathrm{d}t && \text{(矩阵 Jensen 不等式)}\\
={}&(A_{11}+t)^p\big|_0^1=(A_{11}+1)^p-A_{11}^p\geqslant 1.
\end{aligned}$$

于是就得到了 (2.1). 这里, Jensen 不等式要求 x^{p-1} 是凸函数, 因此要求 $p\geqslant 2$.”

大家听到这里, 纷纷对小睦的证明惊叹不已, 就连一直没怎么说话的高松灯也鼓掌起来. 这时, 丰川祥子站起身来, 走到了小睦的身边. 祥子轻轻拉起黑色的裙摆, 向小睦鞠了一躬. 这是淑女必备的礼节之一, 即便祥子已经许久不住在丰川家的豪宅中. 祥子微笑着对小睦说:

“睦子米, 谢谢你的解答. 在我, Oblivionis 的心中, 你已经是一名合格的骑士了. 让我们携手, 让茶话会, 让 Ave Mujica 成为世界上最棒的...”

还没等祥子说完, 小睦便俏皮地耷拉下了脑袋, 对祥子说:

> “嗯, 骑士? 想让我帮你的忙, 难道不应该称呼我为主人吗?”

小睦用她小巧可爱的嘴巴靠近丰川祥子的耳朵, 用只有她们两人能听见的声音说:

> “祥, 如果你能从现在起叫我 Mortis 大人, 并且一辈子都效忠于我, 给我做饭, 帮我打扫, 陪我唱歌, 哄我睡觉, 一起 [redacted] 的话, 我也许可以考虑一下加入你的 Ave Mujica 哦! 对了, 如果你见到小睦的话, 请让她增加一点我的出场时间吧!”

祥子听罢顿时愣在原地, 脸上充满不解和害怕, 竟又似乎有些兴奋和害羞, 如此尴尬的场景也使她一句话也说不出来. 三角初华虽然听不到两人的对话, 但她的愤怒和心痛早已在祥子和小睦的亲密举动中到达顶点. 她还是尽力保持着偶像的情形, 强忍住内心的怒火, 装作温柔又好奇的语气问小睦:

> “那 $1 < p < 2$ 的情形呢? 这种情况下不能用 Jensen 不等式吧? 还真是困扰呢... 但如果是 Mortis 的话, 一定能够解决的吧? 你一定能的吧?”

小睦的心中似乎被什么东西深深刺痛了一下, 那种强烈的求知欲和好胜心立刻将她拉回了原问题的解答上. 她再次朝向天空, 脑子里开始飞速运转, 嘴里又开始念叨起各种矩阵不等式的名字, 但她却没能再得到使自己满意的解答. 在沉默了许久之后, 小睦突然开始向天空尖叫:

> “小祥! 小祥! 小祥! 不要啊啊!!!!”

随后便晕倒在了地上, 眼前的不等式顿时消失在一片漆黑中.

第三幕：梦境中的相遇

在梦境. 在现阶段, 梦境是若叶睦和她的第二人格 (大家称她为 Mortis) 为数不多的可以直接对话的情景. Mortis 已经不记得她是如何跌入梦境的, 她睁开眼看到的, 不过是一片苍茫的林海, 点缀着朵朵玉兰花. 在林海的中央, 摆着一张似乎有些年头的深蓝色咖啡桌, 旁边坐着一位与自己样貌极为相似的淡绿长发的少女——那正是若叶睦. 小睦的手边放着一叠又一叠的论文, 似乎正在努力在搜寻自己要找的答案.

在察觉到 Mortis 的到来后, 小睦微微抬起了头, 望着 Mortis 的眼睛说道:

> "刚才我都看见了哦, Mortis. 你想要 $p > 1$ 时的答案吧?"

Mortis 这才回忆起刚才发生的种种, 确认自己是被三角初华的刁钻问题难住才会变成这样. 也许是对刚才的谢幕不太满意, Mortis 丝毫不耐烦地对小睦说:

> "所以呢? 为什么你不去努力思考, 反而在这里悠哉游哉地看起论文来了? 你不知道我刚刚为了把不等式推广到 $p \geqslant 2$ 花了多大的劲吗?"

小睦随即回答道:

> "我从来没有觉得执着于思考自己无法解决的问题是一件开心的事情. 所以, 我在尝试从别人的结果里汲取灵感. 我的想法快要完成了."

Mortis 听完更激动了:

> “为什么要相信别人的想法? 一味地逃避独立思考, 就能让祥子对你刮目相看了吗? 我诞生的原因, 不正是因为你无法承受祥子对你的过高期待吗? 用自己的大脑思考吧! 用自己的思想成为最棒的数学少女吧!”

听到祥子的名字, 小睦不由震动了她的身体. 在低下头沉默许久后, 小睦轻声说:

> “祥子... 我不想再逃避了...”

小睦随后抬起头看向 Mortis, 说道:

> “Mortis, 我们的生命, 记忆, 知识和灵魂共享. 你所展现的一切, 都来自我的真实但不为人知的一面. 我愿意接受你的思想, 但你不要独自承受一切...”

小睦紧接抱了上去, 用她的下巴尖轻轻靠着 Mortis 细腻的脖颈. Mortis 还没准备好如何接受这一切, 一束刺眼的阳光就从林间的缝隙中倾泻下来, 照亮了梦境的每一处角落, 而自己的身体仿佛也在这灼热的光芒中化为了尘埃... 是梦醒了.

......

> “小睦? 小睦? 你醒了吗?”

小睦醒来后发现她正躺在自己家的床上, 而呼唤着她的名字的正是丰川祥子.

> “太好了... 你醒了啊...”

仔细观察一番, 小睦才发现祥子的双眼都已经哭红了, 看来她在茶话会上晕倒以后就一直待在自己的身边. 小睦的心中久违地泛起一丝暖意, 并温柔地对祥子说:

> “我已经没事了哦. 而且, 我知道 $p > 1$ 的情况要怎么解决了.”

在接下来的时间里, 小睦和祥子开始一起搜集文献并撰写报告. 最后的结果相当成功, 她们找到了针对 $p > 1$ 的情形的严格证法. 虽然证明长度比预想的要长很多, 但所幸在两人的多次确认下, 她们的证明应当是正确的.

第四幕：直接向凸函数进攻!

本文档的结果基于以下两篇文章.

- Aujla, J. S., & Silva, F. C. (2003). Weak majorization inequalities and convex functions. *Linear Algebra and Its Applications*, 369, 217-233.
- Aujla, J. S., & Bourin, J. C. (2007). Eigenvalue inequalities for convex and log-convex functions. *Linear Algebra and Its Applications*, 424(1), 25-35.

先引入一些基本的记号. 记 $M_n(\mathbb{C})$ 是 n 阶复矩阵的集合, $H_n(\mathbb{C})$ 是 n 阶 Hermite 矩阵的集合, $H_n^+(\mathbb{C})$ 是 n 阶半正定 Hermite 矩阵的集合, $U_n(\mathbb{C})$ 是 n 阶酉矩阵的集合. 当 $A \in H_n(\mathbb{C})$ 时, 记 A 的特征值向量为

$$\lambda(A) = \big(\lambda_1(A), \cdots, \lambda_n(A)\big) \in \mathbb{R}^n.$$

如无特殊说明, 假设 $\lambda_1(A) \geqslant \cdots \geqslant \lambda_n(A)$. 接下来引入一些偏序关系:

- 对 $A, B \in H_n(\mathbb{C})$, 记
$$A \leqslant B \Longleftrightarrow B - A \in H_n^+(\mathbb{C}).$$
- 对 $A, B \in H_n(\mathbb{C})$, 记
$$\lambda(A) \leqslant \lambda(B) \Longleftrightarrow \lambda_i(A) \leqslant \lambda_i(B), \quad i = 1, \cdots, N.$$

我们最终的目标是证明一个强大的结果:

主要定理

设 f 是 $[0, +\infty)$ 上严格递增的凸函数, 且 $f(0) = 0$. 则对任意 $A, B \in H_n^+(\mathbb{C})$, 存在 $U, V \in U_n(\mathbb{C})$, 使得
$$f(A+B) \geqslant Uf(A)U^\dagger + Vf(B)V^\dagger.$$

作为上述结果的一个推论, 可以得到:

$$\operatorname{tr}\big[f(A+B)\big] \geqslant \operatorname{tr}\big[Uf(A)U^\dagger\big] + \operatorname{tr}\big[Vf(B)V^\dagger\big] = \operatorname{tr}\big[f(A)\big] + \operatorname{tr}\big[f(B)\big].$$

特别的, 当 $p>1$ 时, $f(x)=x^p$ 是满足定理要求的凸函数, 故有

$$\operatorname{tr}\big[(A+B)^p\big] \geqslant \operatorname{tr}\big[A^p\big]+\operatorname{tr}\big[B^p\big],$$

从而目标问题得证. 为证明主要定理, 先考虑下面的引理:

引理 1 对 $A\in H_n(\mathbb{C})$, 有

$$\lambda_k(A)=\max_{\substack{\mathcal{M}\subset\mathbb{C}^n\\ \dim\mathcal{M}=k}}\min_{\substack{x\in\mathcal{M}\\ |x|=1}} x^\dagger Ax, \qquad k=1,\cdots,n.$$

这个结果被称为极小-极大定理, 它的证明可以在北京大学出版社的《数值分析》中找到, 也可以用下面的过程直接说明.

证明 设 A 的特征值和特征向量对为 $(\lambda_i,u_i)_{i=1}^n$, 其中 $\lambda_1\geqslant\cdots\geqslant\lambda_n$. 则 A 有谱分解

$$A=\sum_{i=1}^n\lambda_iu_iu_i^\dagger \Longrightarrow x^\dagger Ax=\sum_{i=1}^n\lambda_i|u_i^\dagger x|^2.$$

接下来只需证明: 对 $\mathbb{C}^n$ 的一切 k 维线性子空间 $\mathcal{M}$, 都有

$$\lambda_k\geqslant\min_{\substack{x\in\mathcal{M}\\ |x|=1}}\sum_{i=1}^n\lambda_i|u_i^\dagger x|^2. \tag{4.1}$$

令 $\mathcal{S}=\operatorname{span}\{u_k,\cdots,u_n\}$ 是后 $n-k+1$ 个特征向量张成的子空间, 我们证明: $\mathcal{M}$ 中存在一个向量 x, 使得 $|x|=1$ 且 $x\in\mathcal{S}$. 设 $\mathcal{M}$ 的一个基为 $\{v_1,\cdots,v_k\}$, 则 $\mathcal{M}$ 中的元素可表示为

$$x=\sum_{j=1}^k t_jv_j, \quad t_1,\cdots,t_k\in\mathbb{C}.$$

要使得 $x\in\mathcal{S}$, 只需 $u_i^\dagger x=0$, 其中 $i=1,\cdots,k-1$. 因此我们得到方程组

$$\sum_{j=1}^k t_j(u_i^\dagger v_j)=0, \quad i=1,\cdots,k-1. \tag{4.2}$$

方程组 (4.2) 含有 k 个未知量和 $k-1$ 个等式, 因此一定存在非零解, 也即存在非零的 $x\in$

$\mathcal{M} \cap \mathcal{S}$. 不妨设 $|x|=1$, 且这个 x 形如

$$x=\sum_{i=k}^{n} s_i u_i, \quad s_k, \cdots, s_n \in \mathbb{C}.$$

于是有 $\sum_{i=k}^{n} |s_i|^2 = 1$, 从而

$$\sum_{i=1}^{n} \lambda_i |u_i^\dagger x|^2 = \sum_{i=k}^{n} \lambda_i |s_i|^2 \leqslant \sum_{i=k}^{n} \lambda_k |s_i|^2 = \lambda_k.$$

故 (4.1) 成立, 引理 1 得证. ■

作为引理 1 的一个简单推论, 当 $A, B \in H_n(\mathbb{C})$ 且 $A \leqslant B$ 时, 有 $\lambda(A) \leqslant \lambda(B)$. 并且有

引理 2 对 $A, B \in H_n(\mathbb{C})$, $\lambda(A) \leqslant \lambda(B)$ 当且仅当存在 $U \in U_n(\mathbb{C})$ 使得 $A \leqslant U^\dagger B U$.

证明 不妨设 A, B 都是对角阵, 否则只需考虑 A, B 的对角标准型. 设

$$A = \operatorname{diag}\{\lambda_1, \cdots, \lambda_n\}, \quad B = \operatorname{diag}\{\mu_1, \cdots, \mu_n\},$$

且 $\lambda_1 \geqslant \cdots \geqslant \lambda_n$, $\mu_1 \geqslant \cdots \geqslant \mu_n$. 于是 $\lambda(A) \leqslant \lambda(B)$ 等价于 $\lambda_i \leqslant \mu_i$ 对 $i=1,\cdots,n$ 成立.

- 若 $\lambda_i \leqslant \mu_i$ 对 $i=1,\cdots,n$ 成立, 显然有 $A \leqslant B$, 故结论成立.
- 若 $A \leqslant U^\dagger B U$, 则由引理 1 可知 $\lambda(A) \leqslant \lambda(U^\dagger B U) = \lambda(B)$.

故引理 2 得证. ■

引理 3 设 f 是 $[0,+\infty)$ 上的凸函数, $A \in H_n^+(\mathbb{C})$, $v \in \mathbb{C}^n$ 为单位向量, 则

$$v^\dagger f(A) v \geqslant f(v^\dagger A v).$$

证明 这是之前已经证明过的 Jensen 不等式的矩阵形式. 设 A 的谱分解为

$$A = \sum_{i=1}^{n} \lambda_i u_i u_i^\dagger,$$

则由 Parseval 等式 $\sum_{i=1}^n |u_i^\dagger v|^2 = |v|^2 = 1$, 且 $f(A) = \sum_{i=1}^n f(\lambda_i) u_i u_i^\dagger$, 从而

$$v^\dagger f(A) v = \sum_{i=1}^n |u_i^\dagger v|^2 f(\lambda_i) \geqslant f\left(\sum_{i=1}^n |u_i^\dagger v|^2 \lambda_i \right) = f(v^\dagger A v).$$

这里用到了原版 Jensen 不等式. 故引理 3 得证. ■

引理 4 对任何 $T \in M_n(\mathbb{C})$, 存在 $U \in U_n(\mathbb{C})$ 使得 $T^\dagger T = U^\dagger T T^\dagger U$.

证明 设 T 的奇异值分解为 $T = P\Sigma Q^\dagger$, 其中 $P, Q \in U_n(\mathbb{C})$, $\Sigma \in H_n(\mathbb{C})$ 为对角矩阵. 则

$$TT^\dagger = P\Sigma^2 P^\dagger \sim \Sigma^2, \qquad T^\dagger T = Q\Sigma^2 Q^\dagger \sim \Sigma^2,$$

从而 $TT^\dagger \sim T^\dagger T$, 引理 4 得证. ■

将引理 1–引理 4 的内容汇总至下面的表格:

引理 1	$A \in H_n(\mathbb{C}) \Longrightarrow \lambda_k(A) = \max\limits_{\substack{\mathcal{M} \subset \mathbb{C}^n \\ \dim \mathcal{M} = k}} \min\limits_{\substack{x \in \mathcal{M} \\ \lvert x \rvert = 1}} x^\dagger A x$
引理 2	$A, B \in H_n(\mathbb{C})$, $\lambda(A) \leqslant \lambda(B) \Longleftrightarrow \exists U \in U_n(\mathbb{C})$ s.t. $A \leqslant U^\dagger B U$
引理 3	f 凸, $A \in H_n(\mathbb{C})$, $\lvert v \rvert = 1 \Longrightarrow v^\dagger f(A) v \geqslant f(v^\dagger A v)$
引理 4	$T \in M_n(\mathbb{C})$, $T^\dagger T \sim T T^\dagger$

接下来我们证明一个关键的定理, 它可以被视为另一种矩阵形式的 Jensen 不等式.

定理 1 设 f 是 $[0, +\infty)$ 上严格递增的凸函数, 且 $f(0) = 0$. 则对任意 $A \in H_n^+(\mathbb{C})$ 和满足 $\|X\| \leqslant 1$ 的 $X \in M_n(\mathbb{C})$, 有

$$\lambda\big(f(X^\dagger A X)\big) \leqslant \lambda\big(X^\dagger f(A) X\big).$$

证明 由于 f 是严格单调递增的, 故当 $k=1,\cdots,n$ 时,

$$\begin{aligned}
\lambda_k\big(f(X^\dagger AX)\big) &= f\big(\lambda_k(X^\dagger AX)\big) && \text{(单调性)}\\
&= f\left(\max_{\substack{\mathcal{M}\subset\mathbb{C}^n\\ \dim\mathcal{M}=k}}\min_{\substack{u\in\mathcal{M}\\ |u|=1}} u^\dagger X^\dagger AXu\right) && \text{(极小-极大, 引理 1)}\\
&= \max_{\substack{\mathcal{M}\subset\mathbb{C}^n\\ \dim\mathcal{M}=k}}\min_{\substack{u\in\mathcal{M}\\ |u|=1}} \Big[f\big(u^\dagger X^\dagger AXu\big)\Big] && \text{(单调性)}\\
&= \max_{\substack{\mathcal{M}\subset\mathbb{C}^n\\ \dim\mathcal{M}=k}}\min_{\substack{u\in\mathcal{M}\\ |u|=1}} \left[f\left(|Xu|^2\cdot\frac{(Xu)^\dagger A(Xu)}{|Xu|^2}+(1-|Xu|^2)\cdot 0\right)\right]\\
&\leqslant \max_{\substack{\mathcal{M}\subset\mathbb{C}^n\\ \dim\mathcal{M}=k}}\min_{\substack{u\in\mathcal{M}\\ |u|=1}} \left[|Xu|^2\cdot f\left(\frac{(Xu)^\dagger A(Xu)}{|Xu|^2}\right)\right] && \text{(原版 Jensen)}\\
&\leqslant \max_{\substack{\mathcal{M}\subset\mathbb{C}^n\\ \dim\mathcal{M}=k}}\min_{\substack{u\in\mathcal{M}\\ |u|=1}} \left[|Xu|^2\cdot\left(\frac{Xu}{|Xu|}\right)^\dagger f(A)\left(\frac{Xu}{|Xu|}\right)\right] && \text{(矩阵 Jensen, 引理 3)}\\
&= \max_{\substack{\mathcal{M}\subset\mathbb{C}^n\\ \dim\mathcal{M}=k}}\min_{\substack{u\in\mathcal{M}\\ |u|=1}} \Big[u^\dagger X^\dagger f(A)Xu\Big] = \lambda_k\big(X^\dagger AX\big). && \text{(极小-极大, 引理 1)}
\end{aligned}$$

因此我们得到 $\lambda_k\big(f(X^\dagger AX)\big)\leqslant\lambda_k\big(X^\dagger f(A)X\big)$, 定理 1 得证. 注意, 在上述证明中可能发生 $Xu=0$. 但 $Xu=0$ 时自然有

$$f(u^\dagger X^\dagger AXu)=f(0)=u^\dagger X^\dagger f(A)Xu,$$

因此不影响定理 1 的证明. ∎

利用定理 1, 我们最终可以得到主要定理的证明.

定理 2 设 f 是 $[0,+\infty)$ 上严格递增的凸函数, 且 $f(0)=0$. 则对任意 $A,B\in H_n^+(\mathbb{C})$, 存在 $U,V\in U_n(\mathbb{C})$, 使得

$$f(A+B)\geqslant Uf(A)U^\dagger+Vf(B)V^\dagger.$$

证明 不妨设 $A+B$ 是正定的. 定义矩阵 $X=(A+B)^{-\frac{1}{2}}A^{\frac{1}{2}}$, $Y=(A+B)^{-\frac{1}{2}}B^{\frac{1}{2}}$. 根据引理

2 和矩阵范数的定义, 有

$$\begin{aligned}\|X\| &= \sqrt{\lambda_1(XX^\dagger)} = \sqrt{\lambda_1\big((A+B)^{-\frac{1}{2}}A(A+B)^{-\frac{1}{2}}\big)} \\ &\leqslant \sqrt{\lambda_1\big((A+B)^{-\frac{1}{2}}(A+B)(A+B)^{-\frac{1}{2}}\big)} = 1,\end{aligned}$$

因此 $\|X\| \leqslant 1$, $\|Y\| \leqslant 1$. 根据引理 2 和定理 1, 存在 $U_0 \in U_n(\mathbb{C})$ 使得

$$f(A) = f(X^\dagger(A+B)X) \leqslant U_0^\dagger X^\dagger f(A+B)XU_0. \tag{4.3}$$

根据引理 4, $X^\dagger f(A+B)X \sim f(A+B)^{\frac{1}{2}}XX^\dagger f(A+B)^{\frac{1}{2}}$, 故由 (4.3) 存在 $U \in U_n(\mathbb{C})$ 使得

$$f(A) \leqslant U^\dagger f(A+B)^{\frac{1}{2}}XX^\dagger f(A+B)^{\frac{1}{2}}U,$$

从而有

$$Uf(A)U^\dagger \leqslant f(A+B)^{\frac{1}{2}}XX^\dagger f(A+B)^{\frac{1}{2}}. \tag{4.4}$$

类似于 (4.4), 还可以得到: 存在 $V \in U_n(\mathbb{C})$ 使得

$$Vf(B)V^\dagger \leqslant f(A+B)^{\frac{1}{2}}YY^\dagger f(A+B)^{\frac{1}{2}}. \tag{4.5}$$

注意到

$$XX^\dagger + YY^\dagger = (A+B)^{-\frac{1}{2}}(A+B)(A+B)^{-\frac{1}{2}} = I,$$

因此把 (4.4) 和 (4.5) 相加即可得到

$$f(A+B) \geqslant Uf(A)U^\dagger + Vf(B)V^\dagger,$$

从而定理 2 得证. ■

由定理 2 即可证明: 当 $p > 1$ 且 $A, B \in H_n^+(\mathbb{C})$ 时,

$$\mathrm{tr}\big[(A+B)^p\big] \geqslant \mathrm{tr}\big[A^p\big] + \mathrm{tr}\big[B^p\big].$$

作为练习, 读者可在阅读文献后证明 Rotfel'd 迹不等式: 当 $0 < p < 1$ 且 $A, B \in H_n^+(\mathbb{C})$ 时,

$$\operatorname{tr}\big[(A+B)^p\big] \leqslant \operatorname{tr}\big[A^p\big] + \operatorname{tr}\big[B^p\big].$$

尾声：你想要什么奖励?

在完成报告以后, 小睦和祥子都感觉累了, 于是便躺在床上休息. 祥子翻过身, 有些害羞地问小睦: “谢谢你, 小睦. 你想要什么奖励呢? 只要是你想要的, 我都可以满足哦...” 小睦坏笑着向祥子做了个鬼脸, 并说道:

“奬勵我一塊華為手錶!” (完)