School of Mathematical Sciences

AI for Mathematics

小组成员: 方尤乐 尚恩龙 郭子荀
宋苜之 刘浩甜

一、 Exercise 2.15:

Problem: Suppose that $A, B \in S_+^n$, **Proof:** $\langle A, B\rangle \geq 0$

首先，我们先声明一些对原问题具有帮助的引理。

引理 1 $\langle A, B\rangle \doteq \sum_{i=1}^n \sum_{j=1}^n a_{ij}b_{ij} = Tr(A'B)$

引理 2 $\forall A \in S_+^n,\ \exists D \in R^{n\times n}\ s.t. A = D'D$

引理 3 $\forall A, B \in C^{n\times n}, Tr(AB) = Tr(BA)$

引理 4 $\forall D \in R^{n\times n},\ (D'D)_{ii} \geq 0$

于是，原问题可被分解为如下步骤（反推顺序）：

1. $\langle A, B\rangle \geq 0 \leftarrow Tr(A'B) \geq 0$ **(by Lemma 1)**
2. $Tr(A'B) \geq 0 \leftarrow Tr(CC'B) \geq 0$ **(by Lemma 2)**
3. $Tr(CC'B) \geq 0 \leftarrow Tr(C'BC) \geq 0$ **(by Lemma 3)**
4. $Tr(C'BC) \geq 0 \leftarrow Tr(C'D'DC) \geq 0$ **(by Lemma 2)**
5. $Tr(C'D'DC) \geq 0 \leftarrow Tr((DC)'(DC))$
5. $Tr((DC)'(DC)) \geq 0 \leftarrow \sum_{i=1}^n (DC)'(DC)_{ii} \geq 0$ **(by Lemma 4)**

二、 Exercise 2.13:

Problem: Compute the derivatives of the following functions of matrix variables:

(a) $f(X) = a^T Xb, X \in R^{m\times n}, a \in R^m, b \in R^n$

(b) $f(X) = Tr(X^T AX), X \in R^{m\times n}, A \in R^{m\times m}$

(c) $f(X) = ln\, det(X), X \in R^{n\times n}\ \cap\ det(X) > 0$

Definition: ($G\breve{a}teaux$ 可微)

设 f(X) 为矩阵变量函数，如果存在矩阵 $G \in R^{m\times n}$，对任意方向 $V \in R^{m\times n}$ 满足

$$\lim_{t\to 0} \frac{f(X+tV) - f(X) - t\langle G, V\rangle}{t} = 0 \tag{1}$$

则称 f 关于 X 是 $G\breve{a}teaux$ 可微的。满足上式的 G 称为 f 在 X 处在 $G\breve{a}teaux$ 可微意义下的导数（简称为 G 导数）。

(a)

证明 1

$$
\begin{aligned}
&\frac{f(X+tV)-f(X)}{t}\\
&=\frac{a^T(X+tV)b-a^TXb}{t}\\
&=a^TVb=Tr(a^TVb)=Tr(ba^TV)\ \textbf{\textit{(by Lemma 3)}}\\
&=\left\langle ab^T,V\right\rangle\ \textbf{\textit{(by Lemma 1)}}
\end{aligned}
$$

(b)

引理 5 $\forall A\in R^{n\times n},Tr(A)=Tr(A^T)$

证明 2

$$
\begin{aligned}
&\frac{f(X+tV)-f(X)}{t}=\frac{Tr((X+tV)^TA(X+tV))-Tr(X^TAX)}{t}\\
&=\frac{Tr(X^TAX+tX^TAV+tV^TAX+t^2V^TAV)-Tr(X^TAX)}{t}\\
&=\frac{tTr(X^T(A+A^T)V)+t^2Tr(V^TAV)}{t}\ \textbf{\textit{(by Lemma 5)}}\\
&=Tr(X^T(A+A^T)V)+tTr(V^TAV)\\
&=\left\langle (A+A^T)X,V\right\rangle+O(t)\ \textbf{\textit{(by Lemma 1)}}\\
&=\left\langle (A+A^T)X,V\right\rangle+o(1)\to\left\langle (A+A^T)X,V\right\rangle
\end{aligned}
$$

(c)

Try 1:

引理 6 *Schur 分解:* $\forall A\in C^{n\times n},\ \exists\, U,R,\ A=U^HRU$ ，其中 U 为酉矩阵，R 为上三角矩阵。

证明 3

$$
\begin{aligned}
&\frac{f(X+tV)-f(X)}{t}=\frac{ln\ det(X+tV)-ln\ det(X)}{t}=\frac{ln\ det(I+tX^{-1}V)}{t}\\
&=^{Schur}\frac{ln\ det(I+tU^HRU)}{t}\ \textbf{\textit{(by Lemma 6)}}\\
&=\frac{ln\ det(U^H(I+tR)U)}{t}=\frac{ln\ det(I+tR)}{t}\\
&=\frac{\sum_{i=1}^n ln(1+tr_{ii})}{t}=^{Taylor}\frac{\sum_{i=1}^n tr_{ii}+o(t)}{t}\\
&=\sum_{i=1}^n r_{ii}+o(1)=Tr(R)+o(1)=Tr(RUU^H)+o(1)=Tr(U^HRU)+o(1)=Tr(X^{-1}V)+o(1)\\
&=\left\langle (X^{-1})^T,V\right\rangle+o(1)\to\left\langle (X^{-1})^T,V\right\rangle\ \textbf{\textit{(by Lemma 1)}}
\end{aligned}
$$

上面的方法虽然在自然语言证明上不存在问题，但涉及到复数域的极限语言在 lean 以及形式化证明中实在过于繁琐（也可能是因为我们还并不熟悉 lean）。因此，我们默认在 Schur 分解中的 U 为正交阵，R 为实上三角矩阵来进行形式化证明。即如下引理：

引理 7 $\forall A \in R^{n\times n}$, $\exists U, R,\ A = U^T RU$ 其中 U 为正交矩阵，R 为实上三角矩阵。

然而，事实上，上述结论并不正确。不过，以此为理论依据的形式化证明过程也是非常具有参考意义的，得出的最终结果也是合理且正确的。

于是，为了更加严谨且完备的说明此问题，我们又尝试了一种新的方法，该办法可以严谨的求出此矩阵函数的导数。

Try 2: 方法二的具体步骤如下所示：

1. 首先，我们考虑行列式函数，将其按行展开，利用归纳法证明行列式函数可导，从而说明矩阵函数 f 的 G 导数是存在的。（接下来，我们只需要求出 G 即可）

2. 取 $V = \delta_{ij}$(即第 i 行第 j 列处的元素为 1，其余元素全为 0 的矩阵)，将其带入 G 可微的极限表达式中。

$$\lim_{t\to 0} \frac{ln\ det(X + t\delta_{ij}) - ln\ det(X)}{t} = \langle G, \delta_{ij}\rangle = G_{ij} \tag{2}$$

3. 考虑到矩阵 X 与 $X + t\delta_{ij}$，只有一个元素不一样。因此，我们可以将两个矩阵都通过从第 i 行展开的方式来计算行列式，两个矩阵展开后只有第 j 项有所不同，因此 $det(X + t\delta_{ij})$ 可被展开合并为 $det(X) + t * [X.adjugateji]$（第二项为代数余子式）。于是，可进行如下化简：

$$\begin{aligned}
&\frac{ln\ det(X + t\delta_{ij}) - ln\ det(X)}{t} = \frac{ln\ (det(X) + t * [X.adjugateji]) - ln\ det(X)}{t} \\
&= \frac{ln\ (1 + t * (X.adjugateji/det(X)))}{t} \to (X.adjugateji/det(X)) \\
&= (X^{-1})^T_{ij}
\end{aligned}$$

4. 由上式得到 $\forall i\ j,\ G_{ij} = (X^{-1})^T_{ij}$，因此 $G = (X^{-1})^T$。

5. 最后，根据上式以及极限的唯一性，也可得出导数 G 是唯一的，即所求得的 G 即为矩阵函数的唯一导数。

事实上，方法二步骤中每一个看似显然的结论都需要比较繁琐的形式化证明过程**（包括 G 导数的定义、性质及其存在性，行列式展开，代数余子式的性质，log 函数的极限逼近，极限的唯一性等等）**。不过，上述步骤中所蕴含的所有结论都已经被我们形式化证明完毕。

更为具体的形式化证明细节于 lean 代码中展示。