伴随表示与 exp 的导数

Jean H. Gallier

Spring 2023

1 伴随表示：Ad 与 ad

给定任意两个向量空间 E 和 F，回想一下，从 E 到 F 的所有线性映射的向量空间都用 $\mathrm{Hom}(E,F)$ 标志。

从 E 到它本身的所有可逆线性映射的向量空间都是一个标志为 $\mathbf{GL}(E)$ 的群。

当 $E=\mathbb{R}^n$ 时，我们通常用 $\mathbf{GL}(n,\mathbb{R})$ 标志 $\mathbf{GL}(\mathbb{R}^n)$ (并且如果 $E=\mathbb{C}^n$，我们通常用 $\mathbf{GL}(n,\mathbb{C})$ 标志 $\mathbf{GL}(\mathbb{C}^n)$)。

所有 $n\times n$ 矩阵的向量空间 $\mathrm{M}_n(\mathbb{R})$ 也用 $\mathfrak{gl}(n,\mathbb{R})$ 标志 (并且 $M_n(\mathbb{C})$ 用 $\mathfrak{gl}(n,\mathbb{C})$ 标志)。

则 $\mathbf{GL}(\mathfrak{gl}(n,\mathbb{R}))$ 是从 $\mathfrak{gl}(n,\mathbb{R})=\mathrm{M}_n(\mathbb{R})$ 到它自身的所有可逆线性映射的向量空间。

对于任意矩阵 $A\in\mathrm{M}_A(\mathbb{R})$ (或 $A\in\mathrm{M}_A(\mathbb{C})$)，由下式定义映射 $L_A:\mathrm{M}_n(\mathbb{R})\to\mathrm{M}_n(\mathbb{R})$ 和 $R_A:\mathrm{M}_n(\mathbb{R})\to\mathrm{M}_n(\mathbb{R})$

$$L_A(B)=AB,\quad R_A(B)=BA,\quad \text{for all } B\in\mathrm{M}_n(\mathbb{R}).$$

观察到 $L_A\circ R_B=R_B\circ L_A$，对于所有的 $A,B\in\mathrm{M}_n(\mathbb{R})$ 成立。

对于任意矩阵 $A\in\mathbf{GL}(n,\mathbb{R})$，设

$$\mathbf{Ad}_A:\mathrm{M}_n(\mathbb{R})\to\mathrm{M}_n(\mathbb{R})\quad(\text{ conjugation by } A)$$

由下式给出

$$\mathbf{Ad}_A(B)=ABA^{-1}\quad \text{for all } B\in\mathrm{M}_n(\mathbb{R}).$$

观察到 $\mathbf{Ad}_A=L_A\circ R_{A^{-1}}$，并且 $\mathbf{Ad}_A$ 是一个具有逆为 $\mathbf{Ad}_{A^{-1}}$ 的可逆线性映射。

$\mathbf{Ad}_A$ 对可逆矩阵 $B\in\mathbf{GL}(n,\mathbb{R})$ 的约束产生的映射

$$\mathbf{Ad}_A:\mathbf{GL}(n,\mathbb{R})\to\mathbf{GL}(n,\mathbb{R})$$

也由下式给出

$$\mathbf{Ad}_A(B)=ABA^{-1}\quad \text{for all } B\in\mathbf{GL}(n,\mathbb{R}).$$

这一次，观察到 $\mathbf{Ad}_A$ 是一个群同态 (相对于乘法)，因为

$$\begin{aligned}\mathbf{Ad}_A(BC)&=ABCA^{-1}\\&=ABA^{-1}ACA^{-1}\\&=\mathbf{Ad}_A(B)\mathbf{Ad}_A(C).\end{aligned}$$

事实上，$\mathbf{Ad}_A$ 是一个群同构 (因为它的逆是 $\mathbf{Ad}_{A^{-1}}$)。

注意 $\mathbf{Ad}_A$ 不是在 $\mathbf{GL}(n,\mathbb{R})$ 上的线性映射，因为 $\mathbf{GL}(n,\mathbb{R})$ 不是向量空间！

然而，$\mathbf{GL}(n,\mathbb{R})$ 是 $\mathrm{M}_n(\mathbb{R})$ 的开子集，因为它是奇异矩阵集的补集

$$\{A \in \mathrm{M}_n(\mathbb{R}) \mid \det(A) = 0\},$$

的一个闭集，因为它是闭集 $\{0\}$ 由行列式函数的逆象，它是连续的。

由于 $\mathbf{GL}(n,\mathbb{R})$ 是 $\mathrm{M}_n(\mathbb{R})$ 的开子集，对于每一个 $B \in \mathbf{GL}(n,\mathbb{R})$，存在一个开球 $B(B,\eta) \subseteq \mathbf{GL}(n,\mathbb{R})$，使得 $B+X \in B(B,\eta)$ 对于所有的 $X \in \mathrm{M}_n(\mathbb{R})$ 具有 $\|X\| < \eta$，因此 $\mathbf{Ad}_A(B+X)$ 定义良好，并且

$$\begin{aligned}\mathbf{Ad}_A(B+X) - \mathbf{Ad}_A(B) &= A(B+X)A^{-1} - ABA^{-1}\\ &= AXA^{-1},\end{aligned}$$

这表明 $d\left(\mathbf{Ad}_A\right)_B$ 存在，并给出为

$$d\left(\mathbf{Ad}_A\right)_B(X) = AXA^{-1}, \quad \text{for all } X \in \mathrm{M}_n(\mathbb{R}).$$

特别地，对于 $B = I$，我们看到 $\mathbf{Ad}_A$ 在 I 处的导数 $d\left(\mathbf{Ad}_A\right)_I$ 是 $\mathfrak{gl}(n,\mathbb{R}) = \mathrm{M}_n(\mathbb{R})$ 的线性映射，用 $\mathrm{Ad}(A)$ 或 Ad_A (或 $\mathrm{Ad}A$) 标志，并给出为

$$\mathrm{Ad}_A(X) = AXA^{-1} \text{ for all } X \in \mathfrak{gl}(n,\mathbb{R}).$$

Ad_A 的逆是 $\mathrm{Ad}_{A^{-1}}$，因此 $\mathrm{Ad}_A \in \mathbf{GL}(\mathfrak{gl}(n,\mathbb{R}))$。

注意

$$\mathrm{Ad}_{AB} = \mathrm{Ad}_A \circ \mathrm{Ad}_B,$$

所以映射 $A \mapsto \mathrm{Ad}_A$ 是一个群同态，标志为

$$\mathrm{Ad} : \mathbf{GL}(n,\mathbb{R}) \to \mathbf{GL}(\mathfrak{gl}(n,\mathbb{R})).$$

同态 Ad 被称为 $\mathbf{GL}(n,\mathbb{R})$ 的伴随表示 (*adjoint representation*)。

我们还想计算 Ad 在 I 处的导数 $d(\mathrm{Ad})_I$。

对于所有的 $X, Y \in \mathrm{M}_n(\mathbb{R})$，当 $\|X\|$ 足够小时我们有 $I + X \in \mathbf{GL}(n,\mathbb{R})$，并且

$$\mathrm{Ad}_{I+X}(Y) - \mathrm{Ad}_I(Y) - (XY - YX) = \big(YX^2 - XYX\big)(I+X)^{-1}.$$

则，如果我们设

$$\epsilon(X,Y) = \frac{(YX^2 - XYX)(I+X)^{-1}}{\|X\|},$$

对于 $\|X\|$ 足够小时我们证明了下式

$$\mathrm{Ad}_{I+X}(Y) - \mathrm{Ad}_I(Y) = (XY - YX) + \epsilon(X,Y)\|X\|,$$

其中 $\|\epsilon(X,Y)\| \leq 2\|X\|\|Y\|\,\|(I+X)^{-1}\|$，并且其中 $\epsilon(X,Y)$ 在 Y 中是线性的。

设 $\mathrm{ad}_X : \mathfrak{gl}(n,\mathbb{R}) \to \mathfrak{gl}(n,\mathbb{R})$ 是线性映射，由下式给出

$$\mathrm{ad}_X(Y) = XY - YX = [X,Y],$$

并且 ad 是线性映射

$$\mathrm{ad} : \mathfrak{gl}(n,\mathbb{R}) \to \mathrm{Hom}(\mathfrak{gl}(n,\mathbb{R}), \mathfrak{gl}(n,\mathbb{R}))$$

由下式给出为

$$\mathrm{ad}(X) = \mathrm{ad}_X.$$

我们还定义 $\epsilon_X : \mathfrak{gl}(n,\mathbb{R}) \to \mathfrak{gl}(n,\mathbb{R})$ 作为线性映射，给出为

$$\epsilon_X(Y) = \epsilon(X,Y).$$

如果 $\|\epsilon_X\|$ 是 ϵ_X 的算子范数，我们有

$$\|\epsilon_X\| = \max_{\|Y\|=1} \|\epsilon(X,Y)\| \le 2\|X\| \left\|(I+X)^{-1}\right\|.$$

则，等式

$$\mathrm{Ad}_{I+X}(Y) - \mathrm{Ad}_I(Y) = (XY - YX) + \epsilon(X,Y)\|X\|,$$

其对所有 Y 都成立，产生

$$\mathrm{Ad}_{I+X} - \mathrm{Ad}_I = \mathrm{ad}_X + \epsilon_X\|X\|,$$

并因为 $\|\epsilon_X\| \le 2\|X\| \, \|(I+X)^{-1}\|$，我们有 $\lim_{X\mapsto 0}\epsilon_X = 0$，这表明 $d(\mathrm{Ad})_I(X) = \mathrm{ad}_X$；也就是，

$$d(\mathrm{Ad})_I = \mathrm{ad}.$$

符号 $\mathrm{ad}(X)$ (或 $\mathrm{ad}X$) 也被用于替代 ad_X。

映射 ad 作为一个线性映射

$$\mathrm{ad} : \mathfrak{gl}(n,\mathbb{R}) \to \mathrm{Hom}(\mathfrak{gl}(n,\mathbb{R}), \mathfrak{gl}(n,\mathbb{R}))$$

被称为 $\mathfrak{gl}(n,\mathbb{R})$ 的伴随表示 (*adjoint representation*)。有人会检查这个

$$\begin{aligned}\mathrm{ad}([X,Y]) &= \mathrm{ad}(X)\mathrm{ad}(Y) - \mathrm{ad}(Y)\mathrm{ad}(X)\\ &= [\mathrm{ad}(X), \mathrm{ad}(Y)],\end{aligned}$$

李括号在 $\mathfrak{gl}(n,\mathbb{R})$ 上的线性映射。

这意味着 ad 是一个李代数同态。可以检查此属性是否等效于以下称为 Jacobi 恒等式 (*Jacobi identity*) 的恒等式：

$$[X,[Y,Z]] + [Z,[X,Y]] + [Y,[Z,X]] = 0,$$

对于所有的 $X, Y, Z \in \mathfrak{gl}(n,\mathbb{R})$ 成立。

请注意

$$\mathrm{ad}_X = L_X - R_X.$$

最后，我们通过指数证明了 Ad 与 ad 的关系式。

命题 1. 对于任意 $X \in \mathrm{M}_n(\mathbb{R}) = \mathfrak{gl}(n,\mathbb{R})$，我们有

$$\mathrm{Ad}_{e^X} = e^{\mathrm{ad}_X} = \sum_{k=0}^{\infty} \frac{(\mathrm{ad}_X)^k}{k!};$$

也就是，

$$\begin{aligned}e^X Y e^{-X} =& e^{\mathrm{ad}_X} Y\\ =& Y + [X,Y] + \frac{1}{2!}[X,[X,Y]] + \frac{1}{3!}[X,[X,[X,Y]]]\\ &+ \cdots\end{aligned}$$

对于所有的$X, Y \in \mathrm{M}_n(\mathbb{R})$ 成立。

2 exp 的导数

对于在 A 处的指数映射的导数 $d\exp_A$，也可以找到一个公式，但这有点棘手。

它可被证明为

$$d(\exp)_A = e^A \sum_{k=0}^{\infty} \frac{(-1)^k}{(k+1)!} (\mathrm{ad}_A)^k,$$

所以

$$d(\exp)_A(B) = e^A \Big(B - \frac{1}{2!}[A,B] + \frac{1}{3!}[A,[A,B]] - \frac{1}{4!}[A,[A,[A,B]]] + \cdots \Big).$$

对于幂级数

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{(k+1)!} (\mathrm{ad}_A)^k,$$

它习惯地被写为[1]

$$\frac{\mathrm{id} - e^{-\mathrm{ad}_A}}{\mathrm{ad}_A},$$

并且对于 exp 的导数公式通常声明为

$$d(\exp)_A = e^A \left(\frac{\mathrm{id} - e^{-\mathrm{ad}_A}}{\mathrm{ad}_A} \right).$$

该指数公式告诉我们导数 $d(\exp)_A$ 是可逆的。

事实上，很容易看出，如果矩阵 X 的特征值是 $\lambda_1, \ldots, \lambda_n$，则矩阵的特征值

$$\frac{\mathrm{id} - e^{-X}}{X} = \sum_{k=0}^{\infty} \frac{(-1)^k}{(k+1)!} X^k$$

是

$$\frac{1 - e^{-\lambda_j}}{\lambda_j} \text{ if } \lambda_j \neq 0, \text{ and } 1 \text{ if } \lambda_j = 0.$$

因此，矩阵 $\frac{\mathrm{id}-e^{-X}}{X}$ 是可逆的，当且仅当对于某些 $k \in \mathbb{Z} - \{0\}$，没有 λ_j 是 $k2\pi i$ 的形式，所以 $d(\exp)_A$ 是可逆的，当且仅当对于某些 $k \in \mathbb{Z} - \{0\}$，ad_A 的特征值不是 $k2\pi i$ 形式。

[1] 译注：这样的写法更加简洁, 也反映了这个级数与指数函数关系的本质。

- 首先，我们可以将幂级数写成如下形式：

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{(k+1)!} (\mathrm{ad}_A)^k = \frac{1}{\mathrm{ad}_A} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} (\mathrm{ad}_A)^{k+1}.$$

- 其次，我们注意到 $\sum_{k=0}^{\infty} \frac{(-1)^k}{k!} (\mathrm{ad}_A)^k$ 就是指数函数的泰勒级数表达式，可以写成 $e^{-\mathrm{ad}_A}$。因此，我们有：

$$\begin{aligned} \frac{1}{\mathrm{ad}_A} \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} (\mathrm{ad}_A)^{k+1} &= \frac{1}{\mathrm{ad}_A} \left(1 - 1 + \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} (\mathrm{ad}_A)^{k+1} \right) \\ &= \frac{1}{\mathrm{ad}_A} \left(1 - \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} (\mathrm{ad}_A)^k \right) \\ &= \frac{1}{\mathrm{ad}_A} \left(1 - e^{-\mathrm{ad}_A} \right). \end{aligned}$$

- 最后，因为 1 就是单位映射 id。所以，我们得到：

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{(k+1)!} (\mathrm{ad}_A)^k = \frac{\mathrm{id} - e^{-\mathrm{ad}_A}}{\mathrm{ad}_A}.$$

但是,这也可证明,如果 A 的特征值为 $\lambda_1, \ldots, \lambda_n$,则 ad_A 的特征值为 $\lambda_i - \lambda_j$,其中 $1 \leq i, j \leq n$。

总之，$d(\exp)_A$ 是可逆的，当且仅当对于所有的 i, j，我们有

$$\lambda_i - \lambda_j \neq k2\pi i, \quad k \in \mathbb{Z} - \{0\}. \tag{1}$$

此建议定义 $\mathrm{M}_n(\mathbb{R})$ 的以下子集 $\mathcal{E}(n)$。

集合 $\mathcal{E}(n)$ 由所有矩阵 $A \in \mathrm{M}_n(\mathbb{R})$ 组成，$A(\lambda, \mu \in \mathbb{R})$ 的特征值 $\lambda + i\mu$ 位于由条件 $-\pi < \mu < \pi$ 确定的水平带形中。

则，这很清晰，$\mathcal{E}(n)$ 中的矩阵满足条件方程 (1)，因此 $d(\exp)_A$ 对于所有的 $A \in \mathcal{E}(n)$ 是可逆的。

根据反函数定理，指数映射是一个在 $\mathcal{E}(n)$ 与 $\exp(\mathcal{E}(n))$ 之间的局部微分同胚映射。

值得注意的是，更多为真是的：指数映射是在 $\mathcal{E}(n)$ 与 $\exp(\mathcal{E}(n))$ 之间的微分同胚映射 (特别是，它是一个双射)。

这需要大量的工作来证明。例如，参见 Mnemné 和 Testard [36]。我们得到以下结果。

定理 1. *指数映射对 $\mathcal{E}(n)$ 的约束是 $\mathcal{E}(n)$ 在其象 $\exp(\mathcal{E}(n))$ 上是一个微分同胚映射。此外，$\exp(\mathcal{E}(n))$ 由所有不具有实负特征值的可逆矩阵组成；它是 $\mathbf{GL}(n, \mathbb{R})$ 的一个开子集；它包含开球 $B(I, 1) = \{A \in \mathbf{GL}(n, \mathbb{R}) \mid \|A - I\| < 1\}$，对于在 $n \times n$ 矩阵上的每一个矩阵范数 $\|\ \|$ 都成立。*

定理 1 有一些实际应用，因为有一些算法可以找到一个矩阵的实对数，其不具有实负特征值；有关定理 1 在医学成像中应用的更多信息，请参见第 19 章。