集智学园范畴论讲义 v1.7.0809

第二季第 1 课：用箭头构造矩阵 单点集就像一个探针，通过共变 Hom 函子去试探集合范畴的特定对象，通过态射探出了集合中的特定元素。这是范畴论的思维方式，把任何集合都用态射集也就是箭头的集合来替代，对集合中的元素的研究变成了对箭头的研究。范畴论就是箭头的科学，它可以脱离集合论来讨论元素。元素是静态的，集合是静态的，态射是动态的，函子是动态的。范畴论更关注动态的联系而非静态的孤立的概念。

箭头有起点和终点，在范畴上对应域和余域的对象。箭头可以被动地描述系统的演化，也可以主动地描述信号处理的系统。集合映射就是信号处理系统的一种实现。类似的有逻辑信号处理、概率信号处理、线性信号处理等等。范畴论所可以囊括的是各式各样的用带有方向的箭头所描述的信号处理系统。

既然可以用箭头描述集合的元素，进一步可以到其它的范畴中，例如用线性映射描述向量。这样能够用动态的线性映射来描述静态的向量。这种思想和集合范畴是一脉相承的。从这里展开线性代数的讨论，向量、线性映射、矩阵，都可以通过箭头来表述。

本课中用箭头方式重构了基础的线性代数，解释了矩阵作为有限维度线性映射的本质特点，即线性映射的乘法性质和线性叠加的加法性质。通过范畴论的态射更深刻地理解矩阵的意义。并且可以用类似的方法把线性代数的研究方法推广到其它范畴，构成图论中的邻接矩阵、关系中的逻辑矩阵、随机过程中的概率矩阵，等等。

元素 过去的数学都是从集合论开始展开具体问题，为了把集合这样的基础概念纳入到范畴论，需要讨论集合元素的个数．本书严格区分集合的映射和集合元素的映射，分别用普通箭头和带竖线的箭头表示，通常合并写为：

$$\begin{aligned} f: X &\to Y \\ x &\mapsto f(x) \end{aligned} \tag{0.0.1}$$

由于可数、不可数等无限大的个数比较问题，数学上用势来严格定义集合中元素：

定义 0.0.1: 势

集合 $X, Y \in \mathbf{Set}$ 之间若存在**双射 (bijection)**，也就是通常理解的一一对应，则称集合具有相同的**势 (cardinality)**，集合 X 的势记为 $|X|$，这样构造了一种等价关系 $X \simeq Y \Leftrightarrow |X| = |Y|$。显然非空有限集 X 的势 $|X| \in \mathbb{N}$ 是自然数.

范畴论是关于箭头的科学。范畴论中的箭头可以很简单，也可以有深厚的内涵，囊括科学思维的方方面面。下面借助单点集这种特殊的集合，用范畴论的观点来重新定义集合中的元素：

定义 0.0.2: 单点集

一个元素构成的集合称为**单点集 (singleton set)**，集合范畴中所有单点集具有相同的势，可以建立双射的等价关系，在等价的意义下相等，记为 $\mathbb{1} = \{1\}$。

从单点集 $\mathbb{1} = \{1\}$ 到任意集合 X 的映射是 $\varphi: 1 \to X$ 的箭头，这样的箭头的全体记为 $h^{\mathbb{1}}X = \{\varphi: 1 \to X\}$，单点在映射下的像 $\varphi(1) \in X$ 可以区分映射箭头本身，即集合的元素 $x \in X$ 和集合范畴的态射 $\varphi_x \in h^{\mathbb{1}}X$ 在 $\varphi_x(1) = x$ 的约束下有一一对应 $x \leftrightarrow \varphi_x$，于是有相等的势 $|X| = |h^{\mathbb{1}}X|$。在同构的意义下可以把任意集合视为单点集出发的映射的集合

$$X = h^{\mathbb{1}}X \tag{0.0.2}$$

如下图：

$$\begin{tikzcd}
x_1 = \varphi_1(1) & \\
x_2 = \varphi_2(1) & 1 \in \mathbb{1} \arrow[ul, mapsto, "\varphi_1"'] \arrow[l, mapsto, "\varphi_2"'] \arrow[dl, mapsto, "\varphi_3"] \\
x_3 = \varphi_3(1) &
\end{tikzcd}$$

这一过程让静态的集合元素变成了动态的箭头，这种描述方式正是范畴论所偏好的。范畴论关注箭头 φ_x 超过关注元素 x，对于任意的集合映射 $f: X \to Y$，复合映射 $(f \circ \varphi_x): \mathbb{1} \to Y$ 相当于求值：

$$(f \circ \varphi_x)(1) = f(x)$$

见下图 [1]

$$\begin{tikzcd} (f\circ\varphi_x)(1) = f(x) \in Y & \varphi_x(1) = x \in X \arrow[l, mapsto, "f"] & 1 \in \mathbb{1} \arrow[l, mapsto, "\varphi_x"] \arrow[ll, bend right, mapsto, "f\circ\varphi_x"'] \end{tikzcd}$$

以上讨论启发我们把传统意义上集合中的元素，定义为映射的箭头：

定义 0.0.3: 集合中的元素

令 X 为集合，单点集 $\mathbb{1} = \{1\}$ 为单点集，$\mathbb{1} \to X$ 中的映射：

$$\begin{aligned} \varphi_x &: \mathbb{1} \to X \\ 1 &\mapsto \varphi_x(1) = x \end{aligned}$$

称为集合 X 中的**元素 (element)**。

范畴论中箭头又叫**态射 (morphism)**以暗示某些代数结构。经过如上构造，集合的元素视为映射，映射就是一种箭头，因而集合视为箭头的集合，在范畴论中称为**态射集 (hom-set)**。这样整个静态的集合论的基础转到了对动态的箭头的研究。

SISO 信号处理系统

例 0.0.1: 整数乘法

$\mathbb{Z}$ 为整数的集合，一个固定的整数 $n \in \mathbb{Z}$ 可以产生由 n 的倍数，即可以被 n 整除的整数所构成的集合：

$$n\mathbb{Z} = \{nx \mid x \in \mathbb{Z}\}$$

整数 n 可以视为集合的映射也就是箭头：

$$\begin{aligned} \varphi_n &: \mathbb{Z} \to n\mathbb{Z} \\ x &\mapsto \varphi_n(x) = nx \end{aligned}$$

它把输入信号乘以整数输。下面考虑两个整数 $f = 2$ 和 $g = 3$ 作为箭头的串联：

$$\begin{tikzcd} 6\mathbb{Z} & 3\mathbb{Z} \arrow[l, "f=2"'] & \mathbb{Z} \arrow[l, "g=3"'] \arrow[ll, bend left, "f\circ g"] \end{tikzcd} \qquad \begin{tikzcd} f(g(x)) = 6x & g(x) = 3x \arrow[l, mapsto, "f=2"'] & x \arrow[l, mapsto, "g=3"'] \arrow[ll, bend left, mapsto, "f\circ g"] \end{tikzcd}$$

复合箭头 $f \circ g$ 相当于直接把 $x \in \mathbb{Z}$ 映射为 $(f \circ g)(x) = 6x$，这就是整数的乘法。

例: 信号处理系统

信号处理系统有一个输入口 X 和一个输出口 Y，称为**单输入单输出 (SISO, single input, single ouput)**信号处理系统，信号处理视为箭头 $f: X \to Y$，它的域 $\mathrm{Dom}(f) = X$ 是输入，余域 $\mathrm{Cod}(f) = Y$

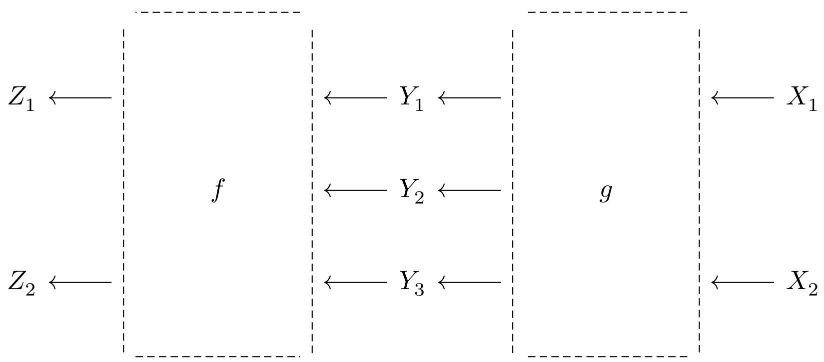
[1] 习惯从右往左描述复合映射 $f \circ \varphi_x$，这样和乘积有相同的顺序

是输出。对于输入信号 $x \in X$ 有：

$$\begin{aligned} f : X &\to Y \\ x &\mapsto f(x) \end{aligned}$$

信号处理系统有多个输入和输出，称为**多输入多输出 (MIMO, multiple input, multiple ouput)**信号处理系统。信号的某个输出可能受到不止一个输入的影响，因此从其内部实现而言可能十分复杂，但从总体上可以视为一个 SISO 信号处理系统。

信号处理系统可以通过**串联 (series connection)**的方式构成多级信号处理系统，相当于箭头的复合，SISO 的串联是显然的。MIMO 的信号处理系统也可以串联，但需要满足式 (**??**)，即前一级信号处理的输出数等于后一级信号处理的输入数。



$$\begin{tikzcd} Z & Y \arrow[l, "f"'] & X \arrow[l, "g"'] \arrow[ll, bend left, "f\circ g"] \end{tikzcd}$$

这是一个 2 输入 3 输出的系统 $g : X \to Y$ 和一个 3 输入 2 输入系统 $f : Y \to Z$ 的串联，总体效果相当于 2 输入 2 输出的系统 $f \circ g : X \to Z$。

信号处理系统的含义很广，无论是一个数学上的映射，一块电信号处理的电路或者装置，乃至计算机程序，都可以视为信号处理系统：

例: 程序

X 和 Y 是两种数据类型，以 X 为输入类型以 Y 为输出类型的函数可以表示为箭头 $f : X \to Y$。若前一级程序的输出类型等于后一级程序的输入类型，则两级程序的处理可以串联。

实数的集合记为 $\mathbb{R}$，它是数域，实线性空间就是数域 $\mathbb{R}$ 上的线性空间，简称 $\mathbb{R}$ 线性空间。$\mathbb{R}$ 线性映射相当于满足线性性质的信号处理系统：

例: 实线性空间范畴 $\mathbf{Vct}_{\mathbb{R}}$

实线性空间范畴 $\mathbf{Vct}_{\mathbb{R}}$ 中箭头也就是态射是 $\mathbb{R}$ 线性映射 $f : V \to W$，其域 V 与余域 W 对象是实线性空间。前一级线性映射的余域若等于后一级线性映射的域，则两级线性映射可以复合，复合出来的箭头仍是线性映射。

例 (0.0.1) 把整数乘法描述为箭头。类似的：

例: 线性信号处理系统

实数 $a \in \mathbb{R}$ 决定了如下线性 SISO 系统:

$$\begin{aligned}\varphi_a : \mathbb{R} &\to \mathbb{R}\\ t &\mapsto \varphi_a(t) = at\end{aligned}$$

注意到当 $t = 1$ 时 $\varphi_a(t) = at = a$,而线性映射集合 $\mathrm{End}(\mathbb{R})$ 中的任意线性映射必有如上乘法形式,根据 $a_1 = a_2 \Leftrightarrow \varphi_{a_1} = \varphi_{a_2}$,实数 a 和线性映射 φ_a 互相决定,实数集合 $\mathbb{R}$ 和线性映射集合 $\mathrm{End}(\mathbb{R})$ 一一对应。它们是同一个事物的两种体现形式,不混淆时都用 a 表示,根据上下文视 $a \in \mathbb{R}$ 或者 $a \in \mathrm{End}(\mathbb{R})$。

例: 关系信号处理系统

关系 $R \subseteq X \times Y$ 决定了 X 到 Y 的关系。若 $X = \{x\}$ 和 $Y = \{y\}$ 是单点集,即有集合的势 $|X| = |Y| = 1$,则 X 到 Y 的关系 g 可能的情况构成幂集 $2^{X \times Y} = \{\emptyset, \{(x,y)\}\}$,记为 $g \in \Omega = \{0,1\}$,它描述了 $x \sim_g y$ 是否成立的逻辑命题。下图箭头上的 $g = 1$ 表示关系成立:

$$Y = \{y\} \xleftarrow[g]{} X = \{x\}$$

引入另一个单点集 $Z = \{z\}$ 和关系 $f \in \Omega$,考虑复合关系:

$$\begin{tikzcd} Z = \{z\} & Y = \{y\} \arrow[l, "f"'] & X = \{x\} \arrow[l, "g"'] \arrow[ll, bend left, "f\circ g"] \end{tikzcd}$$

$f \circ g$ 也取值于 Ω,这就是逻辑与 (AND)。

箭头的复合体现为各种乘法,那么自然可以考虑加法:

例: 线性信号处理系统

实数的加法对应为线性映射的加法,易于理解实数的和 $(a_1 + a_2)$ 可以视为如下的线性映射:

$$\begin{aligned}&\mathbb{R} \to \mathbb{R}\\ &t \mapsto (a_1 + a_2)(t) = a_1(t) + a_2(t) = a_1 t + a_2 t = (a_1 + a_2)t\end{aligned}$$

这样借助实数集 $\mathbb{R}$ 上的加法结构,构造了线性映射集 $\mathrm{End}(\mathbb{R})$ 上的加法结构,实数上加法所具有的结合律、交换律、单位元 0 一样可以对应到线性映射的加法:

$$\begin{tikzcd} \mathbb{R} & \mathbb{R} \arrow[l, bend right, "a_1"'] \arrow[l, "a_1+a_2"'] \arrow[l, bend left, "a_2"] \end{tikzcd} \tag{0.0.3}$$

例: 关系信号处理系统

$$\begin{tikzcd} Y \arrow[r, bend left, "f_1"] \arrow[r, "f_1+f_2"] \arrow[r, bend right, "f_2"'] & X \end{tikzcd}$$

$f_1 + f_2$ 也取值于 Ω,这就是逻辑或 (OR)。

现在可以同时考虑加法和乘法。用一个简单的例子揭示加法和乘法的对偶性,它的本质体现在过去讨论过的范畴极限理论中:

例: 有限集范畴

有限集范畴 **FinSet** 中的对象是有限集，可以用自然数表示，取 $M = \{1, \cdots, m\}, N = \{1, \cdots, n\} \in$ **FinSet**. 直积 $M \times N$ 就是 Cartes 积，直和 $M + N$ 是不交并，注意到：

$$M \times N = \{1, \cdots, m\} \times \{1, \cdots, n\} \simeq \{1, \cdots, mn\}, \quad M + N = \{1, \cdots, m\} + \{1, \cdots, n\} \simeq \{1, \cdots, m+n\}$$

以及

$$|M \times N| = |M||N| = mn, \quad |M + N| = |M| + |N| = m + n$$

线性性 在箭头图示中，乘法代表箭头的复合，加法代表箭头的收束。

定义 0.0.4: 半环

合 R 上赋予加法 $+$ 和乘法 $\cdot$，且有元素 $0, 1 \in R$，若：

1. $(R, +, 0)$ 构成交换幺半群；
2. $(R, \cdot, 1)$ 构成幺半群；
3. 乘法对加法有左右结合律；
4. $\forall r \in R : 0 \cdot r = r \cdot 0 = 0$

则称 $R = (R, +, 0, \cdot, 1)$ 为**半环 (semiring)**.

线性代数中如此定义线性性：

定义: 线性性

K 线性空间 $V, W \in \mathbf{Vct}_K$ 有映射 $f : V \to W$，若对于 $\forall v, v_1, v_2 \in V, \forall \lambda \in K$ 满足：

1. **可加性 (additivity)**：$f(v_1 + v_2) = f(v_1) + f(v_2)$
2. **齐次性 (homogeneity)**：$f(cv) = cf(v)$

则称 f 为 K 线性映射。

现在用范畴和半环来讨论这个定义。可加性

$$\begin{tikzcd} f(v_1+v_2) & v_1+v_2 \arrow[l, "f"'] & {} \arrow[l, bend right, "v_1"'] \arrow[l, bend left, "v_2"] \end{tikzcd} \quad = \quad \begin{tikzcd} & {} \arrow[dl, "f"'] & \\ f(v_1)+f(v_2) & & {} \arrow[ul, "v_1"'] \arrow[dl, "v_2"] \\ & {} \arrow[ul, "f"] & \end{tikzcd}$$

齐次性

$$\begin{tikzcd} f(cv) & {} \arrow[l, "f"] & {} \arrow[l, "c"] & {} \arrow[l, "v"] \end{tikzcd} \quad = \quad \begin{tikzcd} cf(v) & {} \arrow[l, "c"] & {} \arrow[l, "f"] & {} \arrow[l, "v"] \end{tikzcd}$$

线性性的条件也可以合并为，对于 $\forall v_1, v_2 \in V, \forall \alpha, \beta \in K$ 满足：

$$f(\alpha v_1 + \beta v_2) = \alpha f(v_1) + \beta f(v_2)$$

图示为：

$$
\begin{tikzcd}
 & & {} \arrow[dl, "\alpha"'] & \\
f(\alpha v_1+\beta v_2) & \alpha v_1+\beta v_2 \arrow[l, "f"'] & & {} \arrow[ul, "v_1"'] \arrow[dl, "v_2"] \\
 & & {} \arrow[ul, "\beta"] &
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
 & {} \arrow[dl, "\alpha"'] & {} \arrow[l, "f"'] & \\
\alpha f(v_1)+\beta f(v_2) & & & {} \arrow[ul, "v_1"'] \arrow[dl, "v_2"] \\
 & {} \arrow[ul, "\beta"] & {} \arrow[l, "f"] &
\end{tikzcd}
$$

矩阵　加法和乘法共同出现时产生了分配律

$(b_1+b_2)\cdot(a_1+a_2)=b_1\cdot(a_1+a_2)+b_2\cdot(a_1+a_2)+(b_1+b_2)\cdot a_1+(b_1+b_2)\cdot a_2=b_1\cdot a_1+b_1\cdot a_2+b_2\cdot a_1+b_2\cdot a_2$。

这里出现的 4 项，对应下图中的 4 种复合线性映射的和：

$$
\begin{tikzcd}
\mathbb{R} \arrow[r, "a_1"'] \arrow[rr, bend left, "b_1\cdot a_1"] & \mathbb{R} \arrow[r, "b_1"'] & \mathbb{R} & \mathbb{R} \arrow[r, "a_1"'] \arrow[rr, bend left, "b_2\cdot a_1"] & \mathbb{R} \arrow[r, "b_2"'] & \mathbb{R} & \cdots & \mathbb{R} \arrow[r, "a_1"'] \arrow[rr, bend left, "(b_1+b_2)\cdot a_1"] & \mathbb{R} \arrow[r, "b_1+b_2"'] & \mathbb{R} \\
\mathbb{R} \arrow[r, "a_2"'] \arrow[rr, bend left, "b_1\cdot a_2"] & \mathbb{R} \arrow[r, "b_1"'] & \mathbb{R} & \mathbb{R} \arrow[r, "a_2"'] \arrow[rr, bend left, "b_2\cdot a_2"] & \mathbb{R} \arrow[r, "b_2"'] & \mathbb{R} & \cdots & \mathbb{R} \arrow[r, "a_2"'] \arrow[rr, bend left, "(b_1+b_2)\cdot a_2"] & \mathbb{R} \arrow[r, "b_1+b_2"'] & \mathbb{R} \\
 & \vdots & & & \vdots & & & & \vdots & \\
\mathbb{R} \arrow[r, "a_1+a_2"'] \arrow[rr, bend left, "b_1\cdot(a_1+a_2)"] & \mathbb{R} \arrow[r, "b_1"'] & \mathbb{R} & \mathbb{R} \arrow[r, "a_1+a_2"'] \arrow[rr, bend left, "b_2\cdot(a_1+a_2)"] & \mathbb{R} \arrow[r, "b_2"'] & \mathbb{R} & \cdots & \mathbb{R} \arrow[rr, bend left, "b_1\cdot(a_1+a_2)+b_2\cdot(a_1+a_2)"] \arrow[rr, bend right, "(b_1+b_2)\cdot a_1+(b_1+b_2)\cdot a_2"'] & & \mathbb{R}
\end{tikzcd}
$$

以上的构造把实数从静态的实数集中的元素 $a\in\mathbb{R}$，发展为线性映射 $a:\mathbb{R}\to\mathbb{R}$，易于理解从数到矩阵的发展脉络。

例 0.0.2: 点餐

餐厅点拌面和溏心蛋这 2 种菜，价格写为 2-维行向量

$$p=\begin{bmatrix}p_1 & p_2\end{bmatrix}=\begin{bmatrix}38 & 7\end{bmatrix}$$

数量写为 2-维列向量

$$q=\begin{bmatrix}q_1\\ q_2\end{bmatrix}=\begin{bmatrix}2\\ 3\end{bmatrix}$$

总花费为行列向量的矩阵乘积

$$
\begin{aligned}
s=pq=\begin{bmatrix}p_1 & p_2\end{bmatrix}\begin{bmatrix}q_1\\ q_2\end{bmatrix}&=p_1q_1+p_2q_2\\
=\begin{bmatrix}38 & 7\end{bmatrix}\begin{bmatrix}1\\ 3\end{bmatrix}&=38\times 2+7\times 3=76+21=97
\end{aligned}
$$

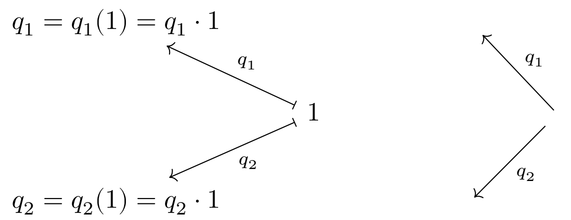
不同的菜计价过程独立，总价是各自小计项的叠加，下图中平行的箭头上的数字求和：

$$
\begin{tikzcd}
 & {} \arrow[dl, "p_1=38"'] & \\
{} & & {} \arrow[ul, "q_1=2"'] \arrow[dl, "q_2=3"] \\
 & {} \arrow[ul, "p_2=7"] &
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
{} & {} \arrow[l, bend right, "p_1\cdot q_1=38\times 2=76"'] \arrow[l, bend left, "p_2\cdot q_2=7\times 3=21"]
\end{tikzcd}
\quad = \quad
\begin{tikzcd}
{} & {} \arrow[l, "p_1q_1+p_2q_2=38\times 2+7\times 3=76+21=97"']
\end{tikzcd}
$$

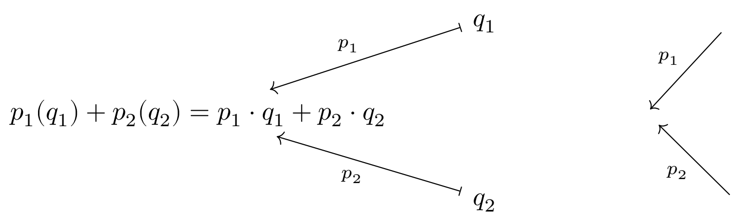
这个简单的生活实例体现了线性代数和范畴论中许多深刻的问题，图中给出了三种等效的箭头表示：

1. 运算过程从右到左
2. 列向量用一组发散箭头表示，行向量用一组收束箭头表示

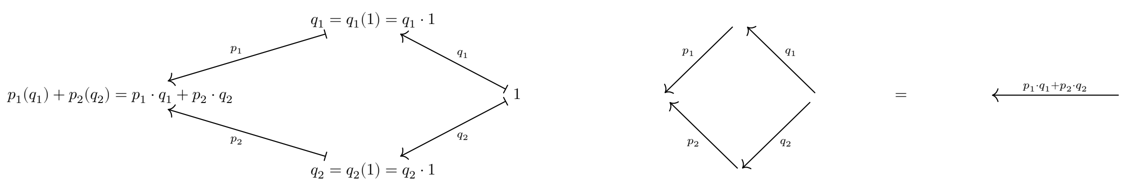
列向量 $\begin{bmatrix} q_1 \\ q_2 \end{bmatrix}$ 可以用下图方式表示，左图体现了 1 经过线性映射得到了列向量的分量，熟悉记号后用右图的简化记法：



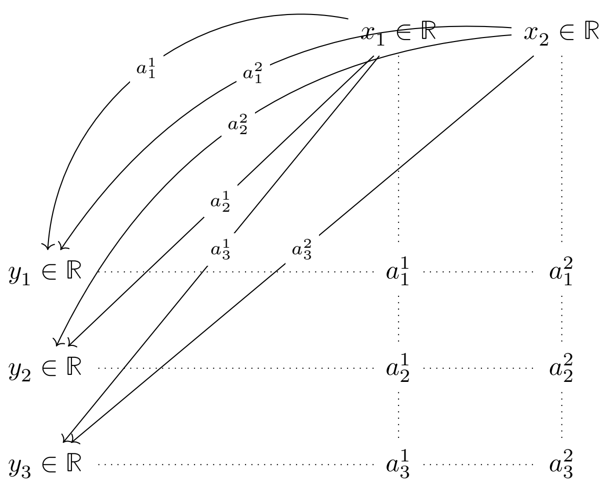
对偶地，行向量 $[p_1 \quad p_2]$ 可以用下图方式表示，它接受一组输入，各自按照箭头上的数字进行线性映射后，收束到一起求和。左图体现了具体的输入产生的结果，熟悉记号后用右图的简化记法：



求总消费金额的过程也分别在下图左右表示：



目前讨论的线性映射 $\mathbb{R} \to \mathbb{R}$ 是 SISO 线性映射，进而考虑 MIMO 即**多输入多输出 (MIMO, multiple input, multiple ouput)**线性映射，举例说明，把 2 元输入写在上侧，把 3 元输出写在左侧：

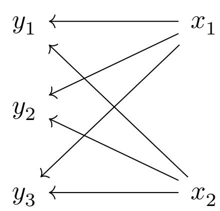


图中的逐个线性映射按行列排列，行对应的输出习惯用下标表示行指标，列对应的输入习惯用上标表示列指标。$m \times n$ 维矩阵可以建立 n 个输入和 m 个输出的联系，矩阵中第 i 行第 j 列的元素 a_i^j 既视为实数也视为 $\mathbb{R} \to \mathbb{R}$ 的线性映射。通常把取值于实数域 $\mathbb{R}$ 中的 $(m \times n)$-维矩阵记为

$$a = [a_i^j] = \begin{bmatrix} a_1^1 & \cdots & a_1^n \\ \vdots & \ddots & \vdots \\ a_m^1 & \cdots & a_m^n \end{bmatrix}$$

这类 m-行 n-列矩阵的全体记为 $\mathbb{R}_m^n = \left\{[a_i^j] \mid a_i^j \in \mathbb{R}\right\}$ 。行向量和列向量都可以视为矩阵．n-行向量空间是一行 n 列的矩阵空间，简记为 $\mathbb{R}^n = \mathbb{R}_1^n$，$m$-列向量空间是一列 m 行的矩阵空间，简记为 $\mathbb{R}_m = \mathbb{R}_m^1$．上图也可简略表示为：

$$\mathbb{R}_3 \xleftarrow{A \in \mathbb{R}_2^3} \mathbb{R}_2$$



例: 点餐

考虑 n 种商品的经济系统，价格用 p_k 表示，每次消费的数量用 q_k 表示。每次消费的总金额是以数量 q_k 为输入的线性函数，各个价格 p_k 则对应了线性映射，所有的输入经过线性映射后求和得到总消费金额：

$$f(q_1, \cdots, q) = p^k q_k = \sum_k p^k q_k$$

这里使用了 Einstein 求和约定。[a] 总金额作为线性函数由 n 个商品小计的求和，不同商品的计算过程之间互不影响，分开买不同的商品求和和一起买所有的商品总费用相同，这是**叠加 (superposition)**规律的体现：

$$\begin{tikzcd}
 & \vdots & q_1 \in \mathbb{R} \arrow[dll, mapsto, "p_1"'] \\
f(q_1, \cdots, q) = p^k q_k \in \mathbb{R} & & \vdots \arrow[ll, mapsto] \\
 & & q_2 \in \mathbb{R} \arrow[ull, mapsto, "p_2"]
\end{tikzcd}$$

[a] 张量计算、微分几何中有繁复的指标求和，Einstein 将其简化为：

$$a^k b_k = \sum_k a^k b_k$$

即行列指标若相同则自动求和，称为 **Einstein 求和约定 (Einstein's summation convention)**.

更复杂的如下，见 (**??**) 节，注意只有 j 指标会自动求和：

$$a_i^j b_j^k = \sum_{j=1}^m a_i^j b_j^k = a_i^0 b_0^k + a_i^1 b_1^k + \cdots + a_i^m b_m^k$$

初学者需要适应用上标 a^i 表示分量，而不要误解为指数．上下标在微分几何和张量中往往是对偶的．范畴论中大量出现对偶的概念，常常用上下标区分．如在拓扑和同调代数中，同调和上同调是对偶的体系，同调中的概念一般用下标 $Z_\bullet, B_\bullet, H_\bullet$，作为对偶的上同调 (cohomology) 就用上标 $Z^\bullet, B^\bullet, H^\bullet$．这就是上同调"上"字的来由．上同调的英文 cohomology 中的前缀

co-表示对偶，通常翻译为“余”，如正弦 (sine) 余弦 (cosine)，正切 (tangent) 余切 (cotangent)，同调 (homology) 上同调 (cohomology) 等。从范畴论的角度看，对偶无非是有向箭头的反向．本书中我们将会学到更多的这类对偶概念.

n 个输入 $\{x_1, \cdots, x_n\}$ 经过 n 个线性映射后叠加得到一个数，把这个过程独立做 m 次，产生 m 个不同的输出 $\{a_1^k x_k, \cdots, a_m^k x_k\}$，其中每个输出都通过线性映射后求和叠加而成。多元输入输出都可以用多个实数构成的列向量来表示，$\mathbb{R}$-矩阵 $A \in \mathbb{R}_m^n$ 对 n-列向量 $x \in \mathbb{R}_n$ 的左乘作用产生了 m-列向量 $Ax \in \mathbb{R}_m$，它相当于 m 个行向量以 $\mathbb{R}$-线性映射的方式并行作用于 n-列向量，并将其结果拼接成新的 m-列向量：

$$\begin{aligned} A : \mathbb{R}_n &\to \mathbb{R}_m \\ x &\mapsto A(x) = Ax = \begin{bmatrix} a_1^k x_k \\ \vdots \\ a_m^k x_k \end{bmatrix} \end{aligned}$$

这里已经出现了有限维线性映射的基本构造。[2] 这里的 $m \times n$ 维矩阵 $A \in \mathbb{R}_m^n$ 决定了如何从 n-列线性地映射到 m-列，虽然 n-列的输入可以任意变化，但线性映射的规则是固定的 $\mathbb{R}_n \to \mathbb{R}_m$ 的形式。显然表达线性映射的矩阵也是可加的：

$$(A_1 + A_2)x = A_1 x + A_2 x$$

这是实数集 $\mathbb{R}$ 以及线性映射 $\mathrm{End}(\mathbb{R})$ 上的可加性的直接推广，易知矩阵的这种可加性满足结合律、交换律、单位元等性质。式 (0.0.3) 推广为：

$$\begin{tikzcd} \mathbb{R}_n \arrow[r, bend left, "A_1 \in \mathbb{R}_m^n"] \arrow[r, "A_1 + A_2 \in \mathbb{R}_m^n"] \arrow[r, bend right, "A_2 \in \mathbb{R}_m^n"'] & \mathbb{R}_m \end{tikzcd}$$

自然希望能够把前面讨论的实数乘法和线性映射的复合规律推广到多元，因此需要考察矩阵的乘法。

矩阵乘法　希望代表两种多元线性映射的矩阵 A 和 B 可以通过适当的方式复合，复合得到的 $B \circ A$ 也是一个矩阵，满足如下的规律：

$$\begin{tikzcd} \mathbb{R}_2 & \mathbb{R}_3 \arrow[l, "B \in \mathbb{R}_2^3"'] & \mathbb{R}_2 \arrow[l, "A \in \mathbb{R}_3^2"'] \arrow[ll, bend right, "B \circ A \in \mathbb{R}_2^2"'] \end{tikzcd}$$

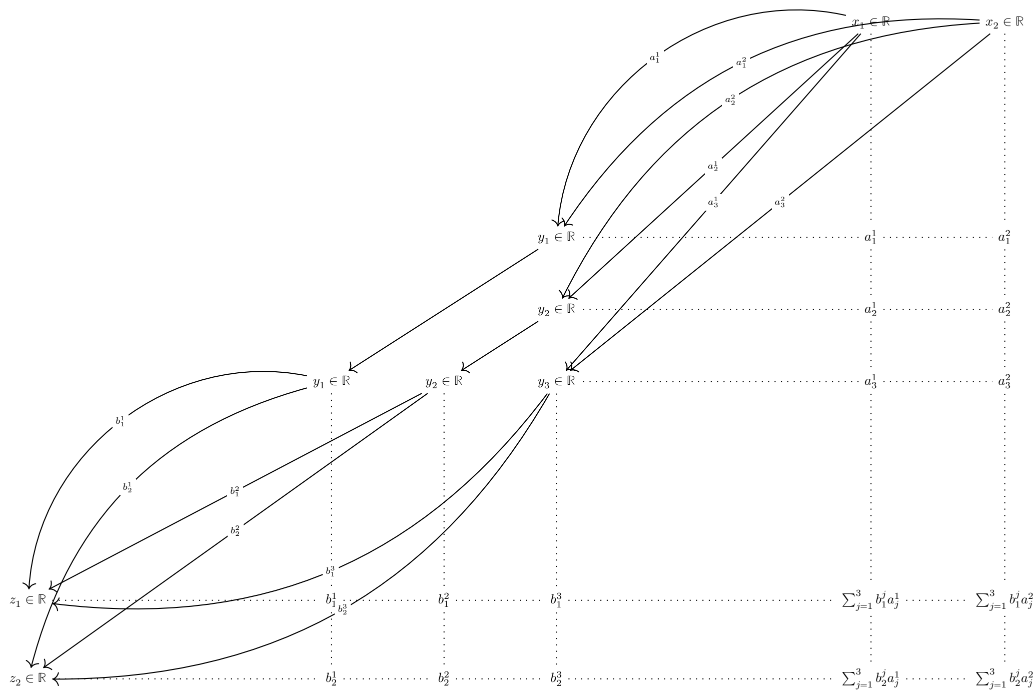
$$\begin{tikzcd} z_1 & y_1 \arrow[l] \arrow[ddl] & x_1 \arrow[l] \arrow[dl] \arrow[ddl] \\ & y_2 \arrow[ul] \arrow[dl] & \\ z_2 & y_3 \arrow[l] \arrow[uul] & x_2 \arrow[l] \arrow[ul] \arrow[uul] \end{tikzcd}$$

[2]根据线性代数的基本常识，在有限维线性空间中选定了基，矩阵和线性映射便一一对应．上式构造了两个不同维度的 $\mathbb{R}$-列向量空间的 $\mathbb{R}$-线性映射 $A : \mathbb{R}_n \to \mathbb{R}_m$.

这种复合规则相当于实数乘法和单线性映射复合规律的推广：[3]

$$\begin{aligned} &\mathbb{R}^m_l \times \mathbb{R}^n_m \to \mathbb{R}^n_l \\ &(b, a) \mapsto ba = [b^j_i] \circ [a^k_j] = [b^j_i a^k_j] = \left[\sum_{j=1}^m b^j_i a^k_j\right] \end{aligned} \tag{0.0.4}$$

矩阵乘法中出现的各分量的乘法，以及按照指标的加法，都可以在下图中找到意义：



例：平面线性映射的合成

平面上的旋转和伸缩都是线性映射，矩阵表示为：

$$Ax = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1\cos\theta - x_2\sin\theta \\ x_1\sin\theta + x_2\cos\theta \end{bmatrix}, \quad By = \begin{bmatrix} r & 0 \\ 0 & s \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} ry_1 \\ sy_2 \end{bmatrix}$$

先旋转再伸缩的合成运算有矩阵表示：

$$(B \circ A)x = \begin{bmatrix} r & 0 \\ 0 & s \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} r\cos\theta & -s\sin\theta \\ r\sin\theta & s\cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

如下图所示，矩阵的乘法体现了多元线性映射的复合：

[3] 用 $\circ$ 符号强调矩阵乘法，也可忽略不用. $\circ$ 符号有映射复合的意义，矩阵乘法实际上也有线性映射复合的意义.

$$
\begin{tikzcd}
\mathbb{R}_2 & \mathbb{R}_2 \arrow[l, "B\in\mathbb{R}_2^2"'] & \mathbb{R}_2 \arrow[l, "A\in\mathbb{R}_2^2"'] & & \mathbb{R}_2 & \mathbb{R}_2 \arrow[l, "B\circ A\in\mathbb{R}_2^2"'] \\
z_1 & y_1 \arrow[l, "r"'] \arrow[dl, "0"] & x_1 \arrow[l, "\cos\theta"'] \arrow[dl, "\sin\theta"] & = & z_1 & x_1 \arrow[l, "r\cos\theta"'] \arrow[dl, "r\sin\theta"] \\
z_2 & y_2 \arrow[l, "s"] \arrow[ul, "0"'] & x_2 \arrow[l, "\cos\theta"] \arrow[ul, "-\sin\theta"'] & & z_2 & x_2 \arrow[l, "s\cos\theta"] \arrow[ul, "-s\sin\theta"']
\end{tikzcd}
$$

首尾相接的一系列线性映射就像并行处理多元数据的**流水线 (pipeline)**，式0.0.4的规则要求行列维度的吻合，这体现了首尾相接的意义。对角矩阵 $\operatorname{diag}(\lambda_1, \cdots, \lambda_n)$ 相当于多个单线性映射的并行处理，下图忽略了值为 0 的箭头：

$$
\begin{tikzcd}
\mathbb{R}_2 & \mathbb{R}_2 \arrow[l, "\operatorname{diag}(a,b)"'] \\
\mathbb{R} & \mathbb{R} \arrow[l, "a"'] \\
\mathbb{R} & \mathbb{R} \arrow[l, "b"]
\end{tikzcd}
$$

n-单位矩阵 I_n 是对角矩阵的特例，从这里的讨论易于理解，单位矩阵把每个输入原封不动地并行输出到流水线的下一个环节。