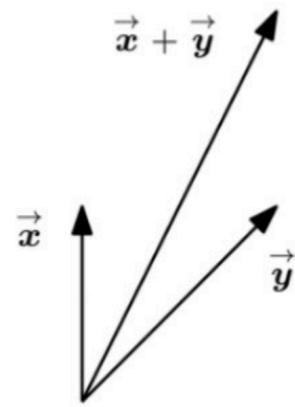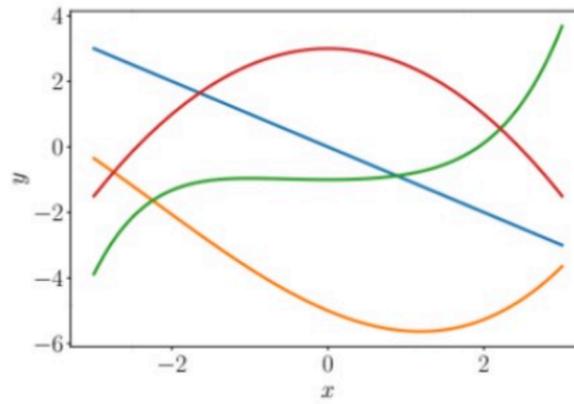Notes on MML

2. Polynomials are also vectors; see Figure 2.1(b): Two polynomials can

(a) Geometric vectors.

(b) Polynomials.

Figure 2.1
Different types of
vectors. Vectors can
be surprising
objects, including
(a) geometric
vectors
and (b) polynomials.

向量可以是特别的类型属于走生常识了。但按我的认知，深度学习涉及的向量仍然符合传统认知，我确实好奇 Embedding 模型转化成的向量是什么形式了。

3. Audio signals are vectors. Audio signals are represented as a series of numbers. We can add audio signals together, and their sum is a new audio signal. If we scale an audio signal, we also obtain an audio signal. Therefore, audio signals are a type of vector, too.

2.3.3 The Minus-1 Trick

In the following, we introduce a practical trick for reading out the solutions x of a homogeneous system of linear equations $Ax = 0$, where $A \in \mathbb{R}^{k \times n}, x \in \mathbb{R}^n$.

To start, we assume that A is in reduced row-echelon form without any rows that just contain zeros, i.e.,

$$A = \begin{bmatrix} 0 & \cdots & 0 & \mathbf{1} & * & \cdots & * & 0 & * & \cdots & * & 0 & * & \cdots & * \\ \vdots & & \vdots & 0 & 0 & \cdots & 0 & \mathbf{1} & * & \cdots & * & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & 0 & \vdots & & & \vdots & \vdots & & \vdots \\ \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & 0 & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 & \mathbf{1} & * & \cdots & * \end{bmatrix},$$

(2.51)

where $*$ can be an arbitrary real number, with the constraints that the first nonzero entry per row must be 1 and all other entries in the corresponding column must be 0. The columns $j_1, \ldots, j_k$ with the pivots (marked in **bold**) are the standard unit vectors $e_1, \ldots, e_k \in \mathbb{R}^k$. We extend this matrix to an $n \times n$-matrix $\tilde{A}$ by adding $n - k$ rows of the form

$$\begin{bmatrix} 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \end{bmatrix}$$

(2.52)

so that the diagonal of the augmented matrix $\tilde{A}$ contains either 1 or -1.

the homogeneous equation system $Ax = 0$. To be more precise, these columns form a basis (Section 2.6.1) of the solution space of $Ax = 0$, which we will later call the *kernel* or *null space* (see Section 2.7.3).

kernel
null space

> **Example 2.8 (Minus-1 Trick)**
>
> Let us revisit the matrix in (2.49), which is already in reduced REF:
>
> $$A = \begin{bmatrix} 1 & 3 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 9 \\ 0 & 0 & 0 & 1 & -4 \end{bmatrix}. \qquad (2.53)$$
>
> We now augment this matrix to a 5×5 matrix by adding rows of the form (2.52) at the places where the pivots on the diagonal are missing and obtain
>
> $$\tilde{A} = \begin{bmatrix} 1 & 3 & 0 & 0 & 3 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 9 \\ 0 & 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 & -1 \end{bmatrix}. \qquad (2.54)$$
>
> From this form, we can immediately read out the solutions of $Ax = 0$ by taking the columns of $\tilde{A}$, which contain -1 on the diagonal:
>
> $$\left\{ x \in \mathbb{R}^5 : x = \lambda_1 \begin{bmatrix} 3 \\ -1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 3 \\ 0 \\ 9 \\ -4 \\ -1 \end{bmatrix}, \quad \lambda_1, \lambda_2 \in \mathbb{R} \right\}, \qquad (2.55)$$
>
> which is identical to the solution in (2.50) that we obtained by "insight".

这也是傅代课上讲过的技巧。不饶都忘了。

Definition 2.15 (Linear Mapping). For vector spaces V, W, a mapping $\Phi : V \to W$ is called a *linear mapping* (or *vector space homomorphism/ linear transformation*) if

$$\forall x, y \in V \; \forall \lambda, \psi \in \mathbb{R} : \Phi(\lambda x + \psi y) = \lambda \Phi(x) + \psi \Phi(y). \qquad (2.87)$$

It turns out that we can represent linear mappings as matrices (Section 2.7.1). Recall that we can also collect a set of vectors as columns of a matrix. When working with matrices, we have to keep in mind what the matrix represents: a linear mapping or a collection of vectors. We will see more about linear mappings in Chapter 4. Before we continue, we will briefly introduce special mappings.

With these definitions, we introduce the following special cases of linear mappings between vector spaces V and W:

- *Isomorphism*: $\Phi : V \to W$ linear and bijective 同构
- *Endomorphism*: $\Phi : V \to V$ linear 自同态
- *Automorphism*: $\Phi : V \to V$ linear and bijective 自同构
- We define $\mathrm{id}_V : V \to V$, $\boldsymbol{x} \mapsto \boldsymbol{x}$ as the *identity mapping* or *identity automorphism* in V.

Example 2.19 (Homomorphism)

The mapping $\Phi : \mathbb{R}^2 \to \mathbb{C}$, $\Phi(\boldsymbol{x}) = x_1 + ix_2$, is a homomorphism: 同态

$$\Phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right) = (x_1 + y_1) + i(x_2 + y_2) = x_1 + ix_2 + y_1 + iy_2$$

$$= \Phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) + \Phi\left(\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right)$$

$$\Phi\left(\lambda \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \lambda x_1 + \lambda ix_2 = \lambda(x_1 + ix_2) = \lambda\Phi\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right).$$
(2.88)

This also justifies why complex numbers can be represented as tuples in $\mathbb{R}^2$: There is a bijective linear mapping that converts the elementwise addition of tuples in $\mathbb{R}^2$ into the set of complex numbers with the corresponding addition. Note that we only showed linearity, but not the bijection.

群、环的相关概念可以对比。向量空间元素的运算是加法，所以第一个条件是加法群同态，同时需要保持数乘，也就是额外的域作用来保证齐次性。对应到群，环就是保持各自运算，比如从加法群到乘法群同态还有：
$\varphi(a) \cdot \varphi(b) = \varphi(a+b)$.
同理类比到环的两个运算。
但是环同态的幺元不是自动保持的。如果 $\varphi(1_S) = 1_R$，那么为含幺环同态，不要求则为一般环同态。
此外还有 ker 也不同：
群同态的 ker：正规子群
环同态的 ker：理想，不是子环
线性映射的 ker：子向量空间
关于 ker 以后有时间再看。

Theorem 2.17 (Theorem 3.59 in Axler (2015)). *Finite-dimensional vector spaces V and W are isomorphic if and only if* $\dim(V) = \dim(W)$.

2.7.2 Basis Change

In the following, we will have a closer look at how transformation matrices of a linear mapping $\Phi : V \to W$ change if we change the bases in V and W. Consider two ordered bases

$$B = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n), \quad \tilde{B} = (\tilde{\boldsymbol{b}}_1, \ldots, \tilde{\boldsymbol{b}}_n)$$
(2.98)

of V and two ordered bases

$$C = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m), \quad \tilde{C} = (\tilde{\boldsymbol{c}}_1, \ldots, \tilde{\boldsymbol{c}}_m)$$
(2.99)

of W. Moreover, $\boldsymbol{A}_\Phi \in \mathbb{R}^{m \times n}$ is the transformation matrix of the linear mapping $\Phi : V \to W$ with respect to the bases B and C, and $\tilde{\boldsymbol{A}}_\Phi \in \mathbb{R}^{m \times n}$ is the corresponding transformation mapping with respect to $\tilde{B}$ and $\tilde{C}$. In the following, we will investigate how $\boldsymbol{A}$ and $\tilde{\boldsymbol{A}}$ are related, i.e., how/ whether we can transform $\boldsymbol{A}_\Phi$ into $\tilde{\boldsymbol{A}}_\Phi$ if we choose to perform a basis change from B, C to $\tilde{B}, \tilde{C}$.

Theorem 2.20 (Basis Change). *For a linear mapping* $\Phi : V \to W$, *ordered bases*

$$B = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n), \quad \tilde{B} = (\tilde{\boldsymbol{b}}_1, \ldots, \tilde{\boldsymbol{b}}_n) \tag{2.103}$$

of V *and*

$$C = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m), \quad \tilde{C} = (\tilde{\boldsymbol{c}}_1, \ldots, \tilde{\boldsymbol{c}}_m) \tag{2.104}$$

of W, *and a transformation matrix* $\boldsymbol{A}_\Phi$ *of* Φ *with respect to* B *and* C, *the corresponding transformation matrix* $\tilde{\boldsymbol{A}}_\Phi$ *with respect to the bases* $\tilde{B}$ *and* $\tilde{C}$ *is given as*

$$\tilde{\boldsymbol{A}}_\Phi = \boldsymbol{T}^{-1} \boldsymbol{A}_\Phi \boldsymbol{S}. \tag{2.105}$$

Here, $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ *is the transformation matrix of* id_V *that maps coordinates with respect to* $\tilde{B}$ *onto coordinates with respect to* B, *and* $\boldsymbol{T} \in \mathbb{R}^{m \times m}$ *is the transformation matrix of* id_W *that maps coordinates with respect to* $\tilde{C}$ *onto coordinates with respect to* C.

Vector spaces

$$V \xrightarrow{\Phi} W \qquad\qquad V \xrightarrow{\Phi} W$$

Ordered bases

$$
\begin{array}{ccc}
B & \xrightarrow[\boldsymbol{A}_\Phi]{\Phi_{CB}} & C \\
\Psi_{B\tilde{B}} \Big\uparrow \boldsymbol{S} & & \boldsymbol{T} \Big\uparrow \Xi_{C\tilde{C}} \\
\tilde{B} & \xrightarrow[\Phi_{\tilde{C}\tilde{B}}]{\tilde{\boldsymbol{A}}_\Phi} & \tilde{C}
\end{array}
\qquad
\begin{array}{ccc}
B & \xrightarrow[\boldsymbol{A}_\Phi]{\Phi_{CB}} & C \\
\Psi_{B\tilde{B}} \Big\uparrow \boldsymbol{S} & & \boldsymbol{T}^{-1} \Big\downarrow \Xi_{\tilde{C}C} = \Xi_{C\tilde{C}}^{-1} \\
\tilde{B} & \xrightarrow[\Phi_{\tilde{C}\tilde{B}}]{\tilde{\boldsymbol{A}}_\Phi} & \tilde{C}
\end{array}
$$
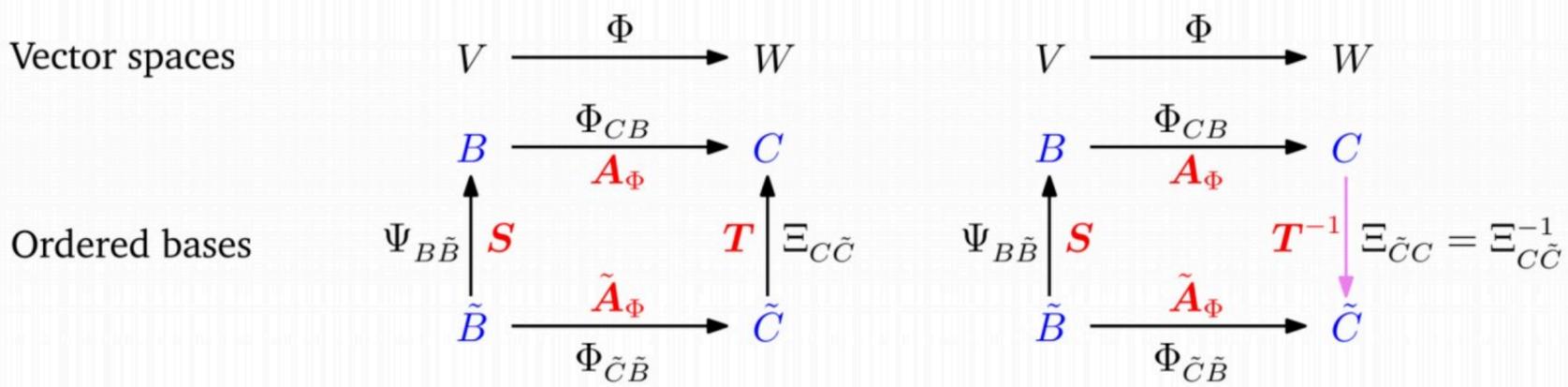
Theorem 2.20 tells us that with a basis change in V (B is replaced with $\tilde{B}$) and W (C is replaced with $\tilde{C}$), the transformation matrix $\boldsymbol{A}_\Phi$ of a linear mapping $\Phi : V \to W$ is replaced by an equivalent matrix $\tilde{\boldsymbol{A}}_\Phi$ with

$$\tilde{\boldsymbol{A}}_\Phi = \boldsymbol{T}^{-1} \boldsymbol{A}_\Phi \boldsymbol{S}. \tag{2.113}$$

Figure 2.11 illustrates this relation: Consider a homomorphism $\Phi : V \to W$ and ordered bases $B, \tilde{B}$ of V and $C, \tilde{C}$ of W. The mapping Φ_{CB} is an instantiation of Φ and maps basis vectors of B onto linear combinations of basis vectors of C. Assume that we know the transformation matrix $\boldsymbol{A}_\Phi$ of Φ_{CB} with respect to the ordered bases B, C. When we perform a basis change from B to $\tilde{B}$ in V and from C to $\tilde{C}$ in W, we can determine the corresponding transformation matrix $\tilde{\boldsymbol{A}}_\Phi$ as follows: First, we find the matrix representation of the linear mapping $\Psi_{B\tilde{B}} : V \to V$ that maps coordinates with respect to the new basis $\tilde{B}$ onto the (unique) coordinates with

respect to the "old" basis B (in V). Then, we use the transformation matrix $\boldsymbol{A}_\Phi$ of $\Phi_{CB}: V \to W$ to map these coordinates onto the coordinates with respect to C in W. Finally, we use a linear mapping $\Xi_{\tilde{C}C}: W \to W$ to map the coordinates with respect to C onto coordinates with respect to $\tilde{C}$. Therefore, we can express the linear mapping $\Phi_{\tilde{C}\tilde{B}}$ as a composition of linear mappings that involve the "old" basis:

$$\Phi_{\tilde{C}\tilde{B}} = \Xi_{\tilde{C}C} \circ \Phi_{CB} \circ \Psi_{B\tilde{B}} = \Xi_{C\tilde{C}}^{-1} \circ \Phi_{CB} \circ \Psi_{B\tilde{B}}. \qquad (2.114)$$

Concretely, we use $\Psi_{B\tilde{B}} = \mathrm{id}_V$ and $\Xi_{C\tilde{C}} = \mathrm{id}_W$, i.e., the identity mappings that map vectors onto themselves, but with respect to a different basis.

Definition 2.21 (Equivalence). Two matrices $\boldsymbol{A}, \tilde{\boldsymbol{A}} \in \mathbb{R}^{m \times n}$ are *equivalent* if there exist regular matrices $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{T} \in \mathbb{R}^{m \times m}$, such that $\tilde{\boldsymbol{A}} = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{S}$.

Definition 2.22 (Similarity). Two matrices $\boldsymbol{A}, \tilde{\boldsymbol{A}} \in \mathbb{R}^{n \times n}$ are *similar* if there exists a regular matrix $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ with $\tilde{\boldsymbol{A}} = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$

特性	相似	等价
映射类型	线性变换 $(V \to V)$	线性映射 $(V \to W)$
空间关系	同一个空间到自身	两个空间之间（可不同）
基的变换	输入和输出空间使用同一套新基。	输入空间和输出空间使用两套独立的新基。
核心不变性	特征值、行列式、迹、特征多项式等。这些是变换固有的、内在的动力学属性。	秩。这是映射结构的核心，代表了信息的"有效维度"或映射的"非平凡性"。
几何比喻	观察同一个物体，只是旋转了观察角度或换了测量单位。物体本身没变。	执行同一个翻译任务，比如将英文（输入）翻译成中文（输出），现在改为将法文（新输入）翻译成德文（新输出）。核心语义（秩）不变。
关系	相似是等价的一个特例：当 $V = W$ 且我们强制要求 $T = S$ 时，等价就退化为了相似。	

Definition 2.23 (Image and Kernel).

For $\Phi : V \to W$, we define the *kernel/null space*

$$\ker(\Phi) := \Phi^{-1}(\mathbf{0}_W) = \{v \in V : \Phi(v) = \mathbf{0}_W\} \qquad (2.122)$$

and the *image/range*

$$\mathrm{Im}(\Phi) := \Phi(V) = \{w \in W | \exists v \in V : \Phi(v) = w\}. \qquad (2.123)$$

We also call V and W also the *domain* and *codomain* of Φ, respectively.

Intuitively, the kernel is the set of vectors $v \in V$ that Φ maps onto the neutral element $\mathbf{0}_W \in W$. The image is the set of vectors $w \in W$ that can be "reached" by Φ from any vector in V. An illustration is given in Figure 2.12.

Remark. Consider a linear mapping $\Phi : V \to W$, where V, W are vector spaces.

- It always holds that $\Phi(\mathbf{0}_V) = \mathbf{0}_W$ and, therefore, $\mathbf{0}_V \in \ker(\Phi)$. In particular, the null space is never empty.
- $\mathrm{Im}(\Phi) \subseteq W$ is a subspace of W, and $\ker(\Phi) \subseteq V$ is a subspace of V.

$$\Phi : V \to W$$

- Φ is injective (one-to-one) if and only if $\ker(\Phi) = \{0\}$.
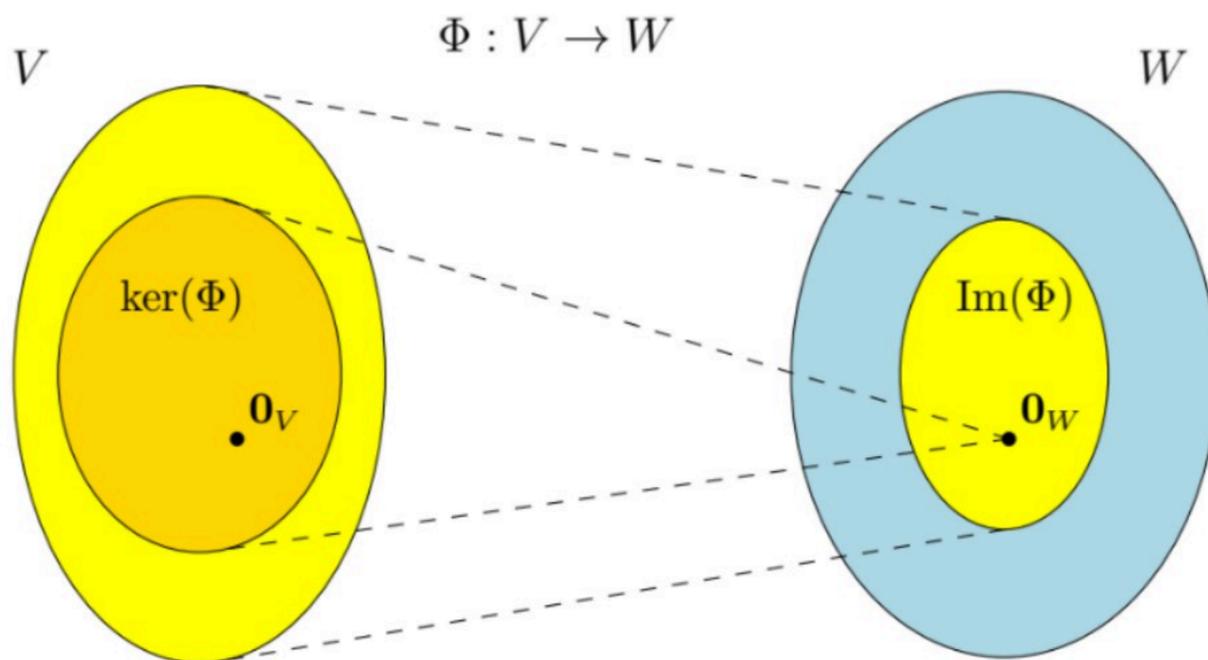
关于从范畴论角度对比这些概念的内容就放在下次的笔记里了。

概念	描述方向	与 f 的关系	范畴论角色
Ker(f)	指向 **Domain (A)**	`f ∘ ker(f) = 0`	子对象
Im(f)	来自 **Codomain (B)**	`A -> Im(f) -> B`	子对象
Coker(f)	来自 **Codomain (B)**	`coker(f) ∘ f = 0`	商对象