Contravariant tensors

Bidual space V^{**} (dual space of the dual space)

Canonical isomorphism: $V^{**} \simeq V$ (in finite dimensions)

$$v \in V \longmapsto \varphi_v : V^* \to \mathbb{R} \quad \varphi_v(\alpha) = \alpha(v)$$

tensor product of two vectors:

$$(u \otimes v)(\alpha, \beta) = \alpha(u) \cdot \beta(v)$$

$\rightsquigarrow$ contravariant indices

$$S = S^{ij}\, b_i \otimes b_j$$

$$S^{ij} = S(\beta^i, \beta^j)$$

General tensors

$$T : (V^*)^p \times (V)^q \longrightarrow \mathbb{R} \qquad (p, q) \text{ tensors}$$

$(0, 0)$ scalar

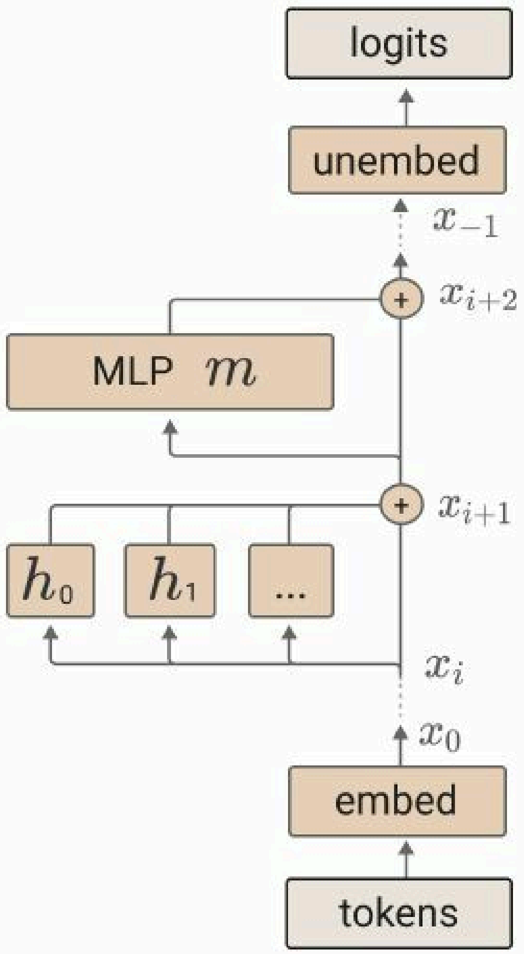
$(1, 0)$ covector α_i $\qquad$ $(0, 1)$ vector v^i

$(2, 0)$ bilinear forms g_{ij} $\qquad$ $(1, 1)$ linear transformation L^i_j $\qquad$ $(0, 2)$ β^{ij}

$\cdots$

$$T = T^{i_1, \ldots, i_p}_{j_1, \ldots, j_q} \beta^{j_1} \otimes \cdots \otimes \beta^{j_q} \otimes b_{i_1} \otimes \cdots \otimes b_{i_p}$$

Example: Transformers (cf. https://transformer-circuits.pub)



The final logits are produced by applying the the unembedding.

$$T(t) = W_U x_{-1}$$

An MLP layer, m, is run and added to the residual stream.

$$x_{i+2} = x_{i+1} + m(x_{i+1})$$

Each attention head, h, is run and added to the residual stream.

$$x_{i+1} = x_i + \sum_{h \in H_i} h(x_i)$$

Token embedding.

$$x_0 = W_E t$$

One residual block

Attention:

keys: $k_i = W_k x_i$

query: $q_i = W_Q x_i$

Value: $v_i = W_V x_0$

$$A = \mathrm{softmax}(q^T k)$$

$$r_i = \sum_j A_{ij} v_j \qquad h(x)_i = W_O r_i$$

multiple attention heads $\rightsquigarrow$ $W_O^H \begin{bmatrix} r^{h_1} \\ r^{h_2} \\ \vdots \end{bmatrix} = \begin{bmatrix} W_O^{h_1} & W_O^{h_2} & \dots \end{bmatrix} \begin{bmatrix} r^{h_1} \\ r^{h_2} \\ \vdots \end{bmatrix}$

$\rightsquigarrow$ linear map between matrices ~ (2,2) tensor

$$h(x) = (I \otimes W_O) \cdot (A \otimes I) \cdot (I \otimes W_V)$$

$$= (A \otimes W_O W_V)$$

$$A \otimes I \to \begin{pmatrix} a_{11} I & a_{12} I & \dots \\ a_{21} I & a_{22} I & \\ \vdots & & \ddots \end{pmatrix} \qquad I \otimes W_V \to \begin{pmatrix} W_V & & \\ & W_V & \\ & & \ddots & \\ & & & W_V \end{pmatrix}$$

Bonus: Unified product

vector of covectors

Vectors: $\begin{pmatrix} a \\ b \end{pmatrix}$ covectors: $\begin{pmatrix} a & b \end{pmatrix}$ matrices: $\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} a & b \end{pmatrix} \\ \begin{pmatrix} c & d \end{pmatrix} \end{pmatrix}$

$= \begin{pmatrix} \begin{pmatrix} a \\ c \end{pmatrix} \begin{pmatrix} b \\ d \end{pmatrix} \end{pmatrix}$

↳ covector of vectors

inner product rule: $\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = ax + by$

outer product rule $\begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} x & y \end{pmatrix} = \begin{pmatrix} ax & ay \\ bx & by \end{pmatrix}$ $\quad u \otimes \alpha = u^i \alpha_j \, b_i \otimes \beta^j$

tensor product rules $\begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a \begin{pmatrix} x \\ y \end{pmatrix} \\ b \begin{pmatrix} x \\ y \end{pmatrix} \end{pmatrix}$ $\quad u \otimes v = u^i v^j \, b_i \otimes b_j$

$\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} x & y \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} a & b \end{pmatrix} x & \begin{pmatrix} a & b \end{pmatrix} y \end{pmatrix}$ $\quad \lambda \otimes \mu = \lambda_i \mu_j \, \beta^i \otimes \beta^j$

Consistent with matrix-vector product.

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} a & b \end{pmatrix} \\ \begin{pmatrix} c & d \end{pmatrix} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} a & b \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \\ \begin{pmatrix} c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} \end{pmatrix} = \begin{pmatrix} ax+by \\ cx+dy \end{pmatrix}$$

$$= \left(\begin{pmatrix} a \\ c \end{pmatrix} \begin{pmatrix} b \\ d \end{pmatrix}\right)\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix} x + \begin{pmatrix} b \\ d \end{pmatrix} y$$

$$\begin{pmatrix} a & b \end{pmatrix}\begin{pmatrix} x & y \\ w & z \end{pmatrix} = \begin{pmatrix} a & b \end{pmatrix}\left(\begin{pmatrix} x \\ w \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix}\right) = \left(\begin{pmatrix} a & b \end{pmatrix}\begin{pmatrix} x \\ w \end{pmatrix} \quad \begin{pmatrix} a & b \end{pmatrix}\begin{pmatrix} y \\ z \end{pmatrix}\right)$$

$$= \begin{pmatrix} a & b \end{pmatrix}\begin{pmatrix} \begin{pmatrix} x & y \end{pmatrix} \\ \begin{pmatrix} w & z \end{pmatrix} \end{pmatrix} = a\begin{pmatrix} x & y \end{pmatrix} + b\begin{pmatrix} w & z \end{pmatrix}$$

Tensor Product of Vector Spaces

Given V, W vector spaces $\begin{pmatrix} \dim V = n \\ \dim W = m \end{pmatrix}$

$V \otimes W$ is the $(n \cdot m)$ dimensional vector space

with basis $\{b_i \otimes a_j\}_{i=1:n,\, j=1:m}$

$\{b_i\}$ basis of V

$\{a_j\}$ basis of W

$$v \otimes w = (v^i b_i) \otimes (w^j a_j) = v^i w^j \, b_i \otimes a_j$$

Isomorphisms:

currying

$$\begin{aligned}
Bil(V \times W, \mathbb{R}) &\cong Lin(V, W^*) \quad (\mathbb{R}^{m \times n}) \\
&\cong Lin(W, V^*) \quad (\mathbb{R}^{n \times m}) \\
&\cong V^* \otimes W^* \\
&\cong (V \otimes W)^* \\
&\cong Lin(V \otimes W, \mathbb{R})
\end{aligned}$$

Tensor space $T^p_q(V) = \underbrace{V \otimes \dots \otimes V}_{p} \otimes \underbrace{V^* \otimes \dots \otimes V^*}_{q}$

$\rightsquigarrow$ tensor power $V^{\otimes p} = T^p_0$

Bonus: Tensor Algebra

$$T(V) = \underbrace{\bigoplus_{k=0}^{\infty}}_{\hookrightarrow \text{ direct sum}} V^{\otimes k} = \mathbb{K} \oplus V \oplus (V \otimes V) \oplus (V \otimes V \otimes V) \dots$$

$\rightsquigarrow$ "one vector space to hold them all"

+ tensor product $\longrightarrow$ <u>algebra</u>

example element: "$1 + \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$"

Weird products

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}_{1,1} \begin{pmatrix} x & y \end{pmatrix}_{1,0} = \left(\begin{pmatrix} a \\ c \end{pmatrix} \begin{pmatrix} b \\ d \end{pmatrix} \right) \begin{pmatrix} x & y \end{pmatrix} = \left(\begin{pmatrix} a & b \\ c & c \end{pmatrix} x \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} y \right)_{2,1}$$

$$= \begin{pmatrix} \begin{pmatrix} a & b \end{pmatrix} \\ \begin{pmatrix} c & d \end{pmatrix} \end{pmatrix} \begin{pmatrix} x & y \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} a & b \end{pmatrix} x & \begin{pmatrix} a & b \end{pmatrix} y \\ \begin{pmatrix} c & d \end{pmatrix} x & \begin{pmatrix} c & d \end{pmatrix} y \end{pmatrix}_{2,1}$$

$$\begin{pmatrix} a \\ b \end{pmatrix}_{0,1} \begin{pmatrix} x & y \\ w & z \end{pmatrix}_{1,1} = \begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} \begin{pmatrix} x & y \end{pmatrix} \\ \begin{pmatrix} w & z \end{pmatrix} \end{pmatrix} = \begin{pmatrix} a \begin{pmatrix} x & y \\ w & z \end{pmatrix} \\ b \begin{pmatrix} x & y \\ w & z \end{pmatrix} \end{pmatrix}_{1,2}$$

$$= \begin{pmatrix} a \\ b \end{pmatrix} \left(\begin{pmatrix} x \\ w \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} \right) = \begin{pmatrix} a \begin{pmatrix} x \\ w \end{pmatrix} & a \begin{pmatrix} y \\ z \end{pmatrix} \\ b \begin{pmatrix} x \\ w \end{pmatrix} & b \begin{pmatrix} y \\ z \end{pmatrix} \end{pmatrix}_{1,2}$$

EleutherAI Math Reading Group - 21/05/2023

Multilinear Algebra

and

Tensors

$$\begin{bmatrix} a & b & e \\ d & e & f \\ g & h & i \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} a \\ d \\ g \end{bmatrix} . x + \begin{bmatrix} b \\ e \\ h \end{bmatrix} y + \begin{bmatrix} c \\ f \\ i \end{bmatrix} z = \begin{bmatrix} \begin{bmatrix} a & b & c \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} \\ \begin{bmatrix} d & e & f \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} \\ \begin{bmatrix} g & h & i \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} \end{bmatrix}$$

Background: Vectors and Vector Spaces

- Vector space $\big(\underbrace{V}_{\text{vectors}}, \underbrace{(\mathbb{K}, +, \cdot)}_{\text{field}}, \underbrace{+}_{\text{vector addition}}, \underbrace{\cdot}_{\text{scalar multiplication}}\big)$

- $(V, +)$ is a commutative group:

$$u + (v + w) = (u + v) + w$$

$$0 \in V \text{ s.t. } 0 + v = v + 0$$

$$\forall v \in V \; \exists \text{ "}-v\text{"} \in V \text{ s.t. } v + (-v) = 0$$

$$u + v = v + u$$

$$(\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v$$

$$\alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v$$

$$1 \cdot v = v$$

$$\alpha(\beta \cdot v) = (\alpha\beta) \cdot v$$

$$\left[\text{ "}-v\text{"} = -1 \cdot v \right]$$

Background: basis of a vector space

$$\mathcal{B} = \{b_i\}_{i\in I} \quad \text{s.t.} \quad \forall v \in V \; \exists \{v^i\}_{i\in I} \subset \mathbb{K}$$

$$v = \sum_{i\in I} v^i b_i \equiv v^i b_i = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}_{\mathcal{B}}$$

(Einstein notation)

Change of basis:

$$\tilde{b}_j = L^i_j b_i$$

↳ linear transformation

$$v = v^i b_i = v^i \underbrace{\Lambda^j_i L^i_j}_{} b_i = \tilde{v}^j \tilde{b}_j$$

↓

$$\Lambda = L^{-1}$$

$$\Rightarrow \tilde{v}^j = \Lambda^j_i v^i$$

basis coefficients are **contravariant**

Background: ① Dual Space

Linear form: $\alpha : V \to \mathbb{K}$ s.t. $\alpha(au + bv)$
$= a\,\alpha(u) + b.\alpha(v)$

$V^* = \{\alpha : V \to \mathbb{K}, \text{ linear}\}$ dual space

① Dual basis: Given $\mathcal{B} = \{b_i\}_{i\in I} \subset V$

$\rightsquigarrow \mathcal{B}^* = \{\beta^i\}_{i\in I}$ $\quad \beta^i(v) = v^i$ (coordinate forms)

NB: $\beta^i(b_j) = \delta^i_j$

$\alpha = \alpha_i \beta^i \quad [\alpha_i = \alpha(b_i)]$

$= [\alpha_1 \ \ldots \ \alpha_n]_{\mathcal{B}^*}$ (covector)

Change of basis $\quad \tilde{b}_j = L^i_j b_i$

$\tilde{\mathcal{B}} = \{\tilde{b}_i\}_{i\in I} \rightsquigarrow \tilde{\mathcal{B}}^* = \{\tilde{\beta}^i\}_{i\in I}$

$\tilde{\alpha}_j = \alpha(L^i_j b_i) = L^i_j \alpha(b_i) = \alpha_i L^i_j$

$\Rightarrow$ covectors are covariant

Recap Covariance & Contravariance

Vector space basis

$$\mathcal{B} = \{b_i\}_{i\in I} \qquad \tilde{b}_j = L_j^i\, b_i$$

$$(\tilde{b}_1 \;\ldots\; \tilde{b}_n) = (b_1 \;\ldots\; b_n)\, L$$

covariant basis

dual basis

$$\mathcal{B}^* = \{\beta^i\}_{i\in I} \qquad \tilde{\beta}^i = (L^{-1})^i_j\, \beta^j$$

$$\begin{pmatrix} \tilde{\beta}^1 \\ \vdots \\ \tilde{\beta}^n \end{pmatrix} = L^{-1} \begin{pmatrix} \beta^1 \\ \vdots \\ \beta^n \end{pmatrix}$$

contravariant basis

contravariant coordinates

$$v = v^i b_i = \tilde{v}^i \tilde{b}_i$$

$$\tilde{v}^i = (L^{-1})^i_j\, v^j$$

$$\begin{bmatrix} \tilde{v}^1 \\ \vdots \\ \tilde{v}^n \end{bmatrix} = L^{-1} \begin{bmatrix} v^1 \\ \vdots \\ v^n \end{bmatrix}$$

covariant coordinates

$$\alpha = \alpha_j \beta^j = \tilde{\alpha}_j \tilde{\beta}^j$$

$$\tilde{\alpha}_j = L_j^i\, \alpha_i$$

$$[\tilde{\alpha}_1 \;\ldots\; \tilde{\alpha}_n] = [\alpha_1 \;\ldots\; \alpha_n]\, L$$

Bilinear forms

$$\varphi : V \times V \to \mathbb{R}$$

$$\varphi(u, \lambda v + \mu w) = \lambda \varphi(u,v) + \mu \varphi(u,w)$$

$$\varphi(\lambda u + \mu v, w) = \lambda \varphi(u,w) + \mu \varphi(v,w)$$

Tensor product of 2 1-forms: $\alpha, \beta \in V^*$

$$(\alpha \otimes \beta)(u,v) = \alpha(u) \cdot \beta(v)$$

$\rightsquigarrow$ basis of bilinear forms: $\beta^i \otimes \beta^j \qquad i,j \in I$

$$\alpha = \alpha_{ij} (\beta^i \otimes \beta^j) \qquad \alpha_{ij} = \alpha(b_i, b_j)$$

$\hookrightarrow$ covariant indices

$$\tilde{\alpha}_{ij} = \alpha(\tilde{b}_i, \tilde{b}_j) = \alpha(L^k_i b_k, L^l_j b_l)$$

$$= L^k_i L^l_j \alpha_{kl}$$

Multilinear forms

k-linear form $\quad T : V^k \longrightarrow \mathbb{R}$ linear on each argument separately

basis $\quad \beta^{i_1} \otimes \dots \otimes \beta^{i_k} \qquad T = T_{i_1 \dots i_k} \left(\beta^{i_1} \otimes \dots \otimes \beta^{i_k} \right)$

Example:

determinant: $T : (\mathbb{R}^n)^n \longrightarrow \mathbb{R}\ ;\ (v_1, \dots, v_n) \longmapsto \det \begin{bmatrix} | & & | \\ v_1 & \dots & v_n \\ | & & | \end{bmatrix}$

$$T = \operatorname{sign}(i_1, \dots, i_n)\, \beta^{i_1} \otimes \dots \otimes \beta^{i_n}$$

$$\hookrightarrow \operatorname{sign}(i_1, \dots, i_n) = \begin{cases} 1 & \text{if } (i_1, \dots, i_n) \text{ even permutation} \\ -1 & \quad \text{"} \qquad\quad \text{odd} \\ 0 & \text{else} \end{cases}$$

Tensor product $\quad T : V^k \to \mathbb{R},\ U : V^\ell \to \mathbb{R} \qquad (T \otimes U) : V^{k+\ell} \to \mathbb{R}$

$$(T \otimes U)(v_1, \dots, v_{k+\ell}) = T(v_1, \dots, v_k)\, U(v_{k+1}, \dots, v_{k+\ell})$$

Inner products

$g : V^2 \to \mathbb{R}$ bilinear form that is

- symmetric : $g(u,v) = g(u,v)$
- positive definite : $g(v,v) \geq 0$ and $g(v,v) = 0 \Leftrightarrow v = 0$

Example : . inner product $g(u,v) = u^i v^i = \delta_{ij} u^i v^j$

. general form $g(u,v) = g_{ij} u^i v^j$

$\hookrightarrow$ Sdp matrix

$g_{ij} = g(b_i, b_j)$

NB. This is one of the main objects in Riemannian Geometry