Clifford group notes

Tobias Göbel

October 2023

1 Preliminary definitions

Definition 1.1 (Field). *A field F is a set that form a group under addition with identity element 0, and the nonzero elements also form a group under multiplication with identity element 1. The multiplication distributes over the addition.*

Definition 1.2 (Vector space). *A vector space over a field F is a non-empty set V together with 2 binary operations, addition and multiplication. The addition operator satisfies the following properties for $a, b, c \in V$.*

(i) $a + b = b + a$

(ii) $a + (b + c) = (a + b) + c$

(iii) There is an identity element 0 *with the property* $a + 0 = a$

(iv) Every element a has an inverse $-a$, where $-a + a = 0$

The multiplication operator is defined between the scalars and the vectors, such that for $\lambda \in F$ and $a \in V$, λa is also a member of the vector space. For $\mu, \nu \in F$ and $a, b \in V$ the following properties are satisfied.

(i) $\mu(a + b) = \mu a + \nu b$

(ii) $(\mu + \nu)a = \mu a + \nu a$

(iii) $(\mu\nu)a = \mu(\nu a)$

(iv) if $1\mu = \mu$ for all scalars μ then $1a = a$ for all vectors a.

Definition 1.3 (Algebra). *Let K be a field, and let A be a vector space over K equipped with an additional binary operator from $A \times A$ to A, denoted by $\cdot$. Then A is an algebra over K, if it holds for $x, y, z \in A$ and $a, b \in K$ that:*

(i) $(x + y) \cdot z = x \cdot z + y \cdot z$

(ii) $z \cdot (x + y) = z \cdot x + z \cdot y$

(iii) $(ax) \cdot (by) = (ab)(x \cdot y)$

Definition 1.4 (Unital Algebra). *An algebra is unital or unitary if it has a unit or identity element I with $Ix = x = xI$ for all x in the algebra.*

Definition 1.5 (Ring). *A set R with two binary compositions $(+)$ (addition) and $(\cdot)$ (multiplication), such that $(R, +)$ is a abelian group and $(R, \cdot)$ has the associative property. Furthermore, multiplication should distribute over addition and it should hold that $x \cdot 0 = 0 \cdot x = 0$ for all $x \in R$.*

Definition 1.6 (Ideal). *A non-empty subset J of a ring R is a (two-sided) ideal if, for all $x, y \in J$ and all $p, q \in R$ it holds that:*

(i) $x + y \in J$

(ii) $pxq \in J$

Definition 1.7 (Group homomorphism)**.** *Let* $(G, *)$ *and* $(H, \circ)$ *be two groups, and* $f : G \to H$ *be a map such that for* $x, y \in G$ *it holds that*

$$f(x * y) = f(x) \circ f(y) \tag{1}$$

then f *is a group homomorphism.*

Definition 1.8 (Algebra homomorphism)**.** *Let* A, B *be algebra over a field* F*.* $f : A \to B$ *is an algebra homomorphism if it holds for all* $\lambda \in F$ *and* $x, y \in A$ *that*

(i) $f(xy) = f(x)f(y)$

(ii) $f(x + y) = f(x) + f(y)$

(iii) $f(\lambda x) = \lambda f(x)$

Furthermore, if f *is a bijection,* f *is called an algebra isomorphism. And if* $A = B$*, such that* f *maps onto itself,* f *is said to be an algebra automorphism of* A *denoted by* $Aut_{\boldsymbol{alg}}(A)$*.*

Definition 1.9 (Group representation)**.** *A linear representation* (V, ϕ) *of a group* G *is a group homomorphism* $\phi : G \to GL(V, K)$*, where* $GL(V, K)$ *is the general linear group over* K*.*

Definition 1.10 (Orthogonal group)**.** *Let* V *be a vectorspace over* K*, associated with a bilinear form* $\mathfrak{b} : V \times V \to K$*. Then the orthogonal group for* $\mathfrak{b}$ *is denoted by* $O(V, \mathfrak{b})$ *and contains all invertible linear transformations* A*, such that for all* $v, w \in V$

$$\mathfrak{b}(v, w) = \mathfrak{b}(Av, Aw)$$

Definition 1.11 (Equivariant/Covariant function)**.**

Definition 1.12 (Invariant function)**.**

2 The Clifford algebra

2.1 Defining the Clifford algebra

Following [LS09] we e construct the Clifford algebra with the help of two objects: The tensor algebra of a vector space $T(V)$ and a quadratic form $\mathfrak{q}$, also on a vector space V.

Definition 2.1 (Tensor algebra)**.**

$$T(V) := \bigoplus_{k=0}^{\infty} V^{\otimes k} = span\{v_1 \otimes ... \otimes v_m | m \geq 0, v_i \in V\} \tag{2}$$

Where $V^{\otimes 0}$ is defined to be the ground set of scalars F.

Definition 2.2 (Quadratic form)**.** *A quadratic form Q on a vector space V is a map $\mathfrak{q} : V \to F$, such that for all $a, b \in V$ and all $\lambda \in F$*

(i) $\mathfrak{q}(\lambda a) = \lambda^2 q(a)$

(ii) the map $(a, b) \to \mathfrak{q}(a + b) - \mathfrak{q}(a) - \mathfrak{q}(b)$ is both linear in a and b

Now consider an ideal $I_\mathfrak{q}$ of $T(V)$ with a quadratic form $\mathfrak{q}$, defined as follows:

$$I_\mathfrak{q} = \left\{ \sum_k A_k \otimes (v_k \otimes v_k - \mathfrak{q}(v_k) \cdot 1_{T(V)}) \otimes B_k : v_i \in V, A_k, B_k \in T(V) \right\} \tag{3}$$

Note that $I_\mathfrak{q}$ satisfies the conditions to be an ideal of ring $T(V)$ by construction. We define the Clifford algebra $Cl(V, \mathfrak{q})$ by quoting out the ideal $I_\mathfrak{q}$ from the tensor algebra $T(V)$. In this way we enforce $v \otimes v$ to equal $\mathfrak{q}(v)$.

Definition 2.3 (Clifford Algebra)**.** *The Clifford algebra $Cl(V, \mathfrak{q})$ over the vector field V, with associated quadratic form $\mathfrak{q}$ is given by*

$$Cl(V, \mathfrak{q}) = T(V)/I_\mathfrak{q} \tag{4}$$

The product in the Clifford algebra is denoted with juxtaposition, such that

$$vv = v^2 = \mathfrak{q}(v) \tag{5}$$

Moreover, (if $char(F) \neq 2$) we can write

$$\mathfrak{q}(v + w) = (v + w)(v + w) = v^2 + vw + wv + w^2 \tag{6}$$

With defining $\mathfrak{b}(v, w) = \frac{1}{2}(\mathfrak{q}(v + w) - \mathfrak{q}(v) - \mathfrak{q}(w))$ we can identify

$$vw + wv = 2\mathfrak{b}(v, w) \tag{7}$$

known as the *the fundamental Clifford identity.*

2.2 Signatures

For now, let V be a n-dimensional vector space over the reals ($F = \mathbb{R}$) and let $E = \{e_1, ..., e_n\}$ be a corresponding orthogonal basis. We introduce E^+, E^-, E^0 as

$$E^+ = \{e \in E \quad : \quad \mathfrak{q}(e) > 0\} \tag{8}$$
$$E^- = \{e \in E \quad : \quad \mathfrak{q}(e) < 0\} \tag{9}$$
$$E^0 = \{e \in E \quad : \quad \mathfrak{q}(e) = 0\} \tag{10}$$

One can show that for another orthogonal basis H of $(V, \mathfrak{q})$ the sizes of the respective subspaces are maintained i.e. (For a proof see Theorem 2.4 of [LS09])

$$|E^+| = |H^+| \tag{11}$$
$$|E^-| = |H^-| \tag{12}$$
$$|E^0| = |H^0| \tag{13}$$

This tells us that there is a unique tuple $(p, q, r) = (|E^+|, |E^-|, |E^0|)$, called the *signature* of $(V, \mathfrak{q})$, which is independent of the chosen basis. Furthermore, a quadratic form for which $r = 0$ is said to be *nondegenerate*.

2.3 Bases for the Clifford algebra

If we take V to be finite dimension n, one can write a basis for $(V, \mathfrak{q})$ as $\{e_1, ..., e_n\}$. Such a basis is defined to be *orthogonal*, if it holds that for every $i \neq j$

$$\beta(e_i, e_j) = q(e_i + e_j) - q(e_i) - q(e_j) = 0 \tag{14}$$

Besides this property, an *orthonormal* basis also entails that $\mathfrak{q}(e_i) \in \{-1, 0, 1\}$ for every $i \in [n]$. Using equation 7 it is easily seen that two orthogonal basis vectors e_i, e_j anticommute:

$$e_i e_j + e_j e_i = 0 \tag{15}$$

If $\{e_1, ..., e_n\}$ is any basis for $(V, \mathfrak{q})$, a basis for $Cl(V, \mathfrak{q})$ is given by

$$\left\{ e_A \quad | \quad e_A = \prod_{i \in A} e_i, \quad A \subseteq \{1, ..., n\} \right\} \tag{16}$$

Where the product is taken in increasing order of i and $e_\emptyset = 1_{C(V,\mathfrak{q})}$.
For the sake of illustration, a typical element $v \in Cl(V, \mathfrak{q})$, with V being a 3-dimensional vectorspace over the reals, might look like

$$v = 2 + 5e_1 - 3.2e_2 + e_3 - 0.4e_1e_3 - e_1e_2e_3 \tag{17}$$

In order to intuitevely see that equation 16 is a basis for $Cl(V, \mathfrak{q})$, we recall that $Cl(V, \mathfrak{q}) = T(V)/I_\mathfrak{q}$, and by using the expression for the tensor algebra 2 we know:

$$Cl(V, \mathfrak{q}) = span\{e_{i1}...e_{im} | m \geq 0, i_j \in [n], j \in [m]\} \tag{18}$$

If we now choose $\{e_1, ..., e_n\}$ to be an orthogonal basis for $(V, \mathfrak{q})$, which one can always do (See Theorem 2.2 of [LS09]), we can transform products of the form $e_{i1}...e_{im}$ by commuting through elements at the cost of a minus sign, until we hit a similar element such that we can reduce the pair to a scalar with $e_i e_i = \mathfrak{q}(e_i)$. Repeating this operation, we can reach ordered products with at most n different indices. This shows us that $Cl(V, \mathfrak{q})$ is indeed spanned by $\{e_A | A \subseteq [n]\}$.
From this, we can additionally see that $dim(Cl(V, \mathfrak{q})) = 2^n$ with $n = dim(V)$, since there are precisely 2^n subsets $A \subseteq \{1, ..., n\}$.

In this part we will follow and refer to sections D.5 and D.6 of [RBF23]. Previously we saw that a basis for $Cl(V, \mathfrak{q})$ can be constructed using a basis for V but in order to define the notion of orthogonality on a Clifford algebra, we have to extend the bilinear form $\mathfrak{b} : V \times V \to F$ to $\bar{\mathfrak{b}} : Cl(V, \mathfrak{q}) \times Cl(V, \mathfrak{q}) \to F$. Therefore we define the *extended bilinear form* $\bar{(}\mathfrak{b})$ and the *extended quadratic form* $\bar{\mathfrak{q}} : Cl(V, \mathfrak{q}) \to F$ as follows:

$$\bar{\mathfrak{b}}(v_1, v_2) = \zeta(\beta(v_1)v_2) \quad \text{and} \quad \bar{\mathfrak{q}}(v) = \bar{\mathfrak{b}}(v, v) \quad \text{for} \quad v_1, v_2, v \in Cl(V, \mathfrak{q}) \tag{19}$$

Here $\zeta : Cl(V, \mathfrak{q}) \to F$ represents a projector on the scalar field F, and $\beta : Cl(V, \mathfrak{q}) \to Cl(V, \mathfrak{q})$ denotes the *main anti-involution*, which reverses the order of multiplication within Clifford algebra products. So let $v = x_0 + \sum_{i \in I} x_i \cdot v_{1i}...v_{ni}$ be an arbitrary element of $Cl(V, \mathfrak{q})$, then

$$\beta(v) = x_0 + \sum_{i \in I} x_i \cdot v_{ni}...v_{1i} \tag{20}$$

First, we claim that this extended bilinear form $\bar{\mathfrak{b}}$ behaves similarly to $\mathfrak{b}$ on the sub-vectorspace V within $Cl(V, q)$. So, for $v_1, v_2 \in V$ we have that $\bar{\mathfrak{b}}(v_1, v_2) = \mathfrak{b}(v1, v_2)$. (For a proof see Lemma D.25 of [RBF23]).
Thereafter, we state that if $\{e_1, ..., e_n\}$ is an orthogonal basis for $(V, \mathfrak{q})$, then $\{e_A | A \subseteq [n]\}$ is an orthogonal basis of $Cl(V, \mathfrak{q})$ with respect to the extended bilinear form as defined above. (See Theorem D.26 [RBF23])

2.4 Multivector grading

Let $\{e_1, ..., e_n\}$ be an orthogonal basis for $(V, \mathfrak{q})$, then we define a sub-vector space of $Cl(V, \mathfrak{q})$ for every *grade* $m \in \{0, ..., n\}$:

$$Cl(V, \mathfrak{q})^{(m)} = span\{e_A | A \subseteq [n], |A| = m\}, \quad Cl(V, \mathfrak{q})^{(0)} = F \tag{21}$$

These sub-vectorspaces are independent of the choice of basis. So for another orthogonal basis $\{\tilde{e}_1, ..., \tilde{e}_n\}$ of $(V, \mathfrak{q})$ we still have that

$$Cl(V, \mathfrak{q})^{(m)} = span\{e_A | A \subseteq [n], |A| = m\} = span\{\tilde{e}_A | A \subseteq [n], |A| = m\} \tag{22}$$

A proof for this claim can be found in section D.27 of [RBF23]. Note that the dimension of a sub-vectorspace $Cl(V, \mathfrak{q})^{(m)}$ is given by $\binom{n}{m}$. Now, we can write the total Clifford algebra $Cl(V, \mathfrak{q})$ as an orthogonal sum decomposition of the respective sub-vectorspaces

$$Cl(V, \mathfrak{q}) = \bigoplus_{m=0}^{n} Cl(V, \mathfrak{q})^{(m)} \tag{23}$$

Elements $x \in Cl(V, \mathfrak{q})^{(m)}$ are generally called *m-multivectors* or vectors of pure grade m. In particular, we refer to elements $x \in Cl(V, \mathfrak{q})^{(0)}$ as *scalars*, to $x \in Cl(V, \mathfrak{q})^{(1)}$ as *vectors*, to $x \in Cl(V, \mathfrak{q})^{(2)}$ as *bi-vectors* and to $x \in Cl(V, \mathfrak{q})^{(3)}$ as *tri-vectors*. Also, the basis element $e_1...e_n$ that spans the sub-vectorspace with maximal grade $Cl(V, \mathfrak{q})^{(n)}$ is called the *pseudoscalar* and is denoted as I.

In this way, we can express any element x of $Cl(V, \mathfrak{q})$ in terms of its grade decomposition

$$x = x^{(0)} + x^{(1)} + ... + x^{(n)} \tag{24}$$

Furthermore, we can define two subspaces $Cl(V, \mathfrak{q})^{[0]}$ and $Cl(V, \mathfrak{q})^{[1]}$ representing the direct sum of all even and respectively odd grade sub-vectorspaces

$$Cl(V, \mathfrak{q})^{[0]} = \bigoplus_{m=0}^{n} Cl(V, \mathfrak{q})^{(2m)}, \quad Cl(V, \mathfrak{q})^{[0]} = \bigoplus_{m=0}^{n} Cl(V, \mathfrak{q})^{(2m+1)} \tag{25}$$

With $Cl(V, \mathfrak{q})^{(m)} := 0$ for $m \notin \{0, ..., n\}$. Again, we can decompose any element of $Cl(V, \mathfrak{q})$ in terms of its *parity* decomposition

$$x = x^{[0]} + x^{[1]} \tag{26}$$

3 Expanding on the Clifford algebra

This section is inspired by chapter 4 of *Geometric algebra for physicists* [DL03]. The terms Geometric algebra and Clifford algebra are used interchangably in literature, although a Geometric algebra usually takes the ground field to equal the reals $F = \mathbb{R}$. In order to conveniently distinguish scalars, vectors and general elements of the Clifford algebra, we will use the Greek alphabet for scalars, lowercase Latin alphabet for vectors and uppercase Latin for general elements in $Cl(V, \mathfrak{q})$, such that

$$\lambda, \mu, \nu, .. \in F \quad v, w, ... \in V \quad A, B, C, .. \in Cl(V, \mathfrak{q})$$

3.1 Inner and outer products

We saw before that for a quadratic vectorspace $(V, \mathfrak{q})$ the bilinear form is given by

$$\mathfrak{b}(v, w) = \mathfrak{q}(v + w) - \mathfrak{q}(v) - \mathfrak{q}(w) \quad \text{for} \quad v, w \in V \tag{27}$$

For the sake of conciseness we adopt the notation of [DL03], such that $\mathfrak{b}(v, w) = v \cdot w = \frac{1}{2}(vw + wv)$. We will call this operation the *inner* product.

Now we also define the *outer* (or *exterior*) product as

$$v \wedge w = \frac{1}{2}(vw - wv) \tag{28}$$

From this expression, it is easy to see that the outer product between two vectors is antisymmetric, such that $a \wedge b = -b \wedge a$. Together, the inner product and the outer product, form the following identity for the geometric product of two vectors

$$ab = a \cdot b + a \wedge b \tag{29}$$

Both the innerproduct and the ouderproduct are associative and distributive (**proof?**). We can extend the definition of the outer products for a pair over a tuple of vectors $v_1, ...v_k$ as

$$v_1 \wedge v_2 \wedge ... \wedge v_k = \frac{1}{k!} \sum_{\sigma \in S_k} \mathrm{sign}(\sigma) \cdot v_{\sigma(1)} v_{\sigma(2)} ... v_{\sigma(k)} \tag{30}$$

Here, the sum is taken over all permutations σ and $\mathrm{sign}(\sigma)$ returns 1 if σ is an even permutation and -1 if odd. From this definition it follows that whenever two vectors v_i, v_j are linearly dependent, the outer product reduces to zero, since for every permutation there exists another permutation with v_i, v_j swapped, gaining opposite sign. If an element in the Clifford algebra can be written as the outerproduct of a set of vectors, this element called a *blade*. Every blade $v_1 \wedge v_2 \wedge ...v_k$ can always be expressed as the product of a set of orthogonal basisvectors $E = e_1 e_2 ... e_k$. Moreover, an arbitrary multivector of grade m, is always the sum of a set of blades with grade m.

Now consider a geometric product of the form $wv_1v_2...v_k$. Using $ab = 2a \cdot b - ba$ we get

$$\begin{aligned} wv_1v_2...v_k =& 2w \cdot v_1v_2...v_l - v_1wv_2...v_k \\ =& 2w \cdot v_1v_2...v_l - 2w \cdot v_2v_1wv_3...v_k + v_1v_2wv_3...v_k \\ =& (-1)^k v_1v_2...v_kw + 2\sum_{l=1}^{k} (-1)^{k+1}(w \cdot v_k)v_1v_2...\cancel{v_l}..v_k \end{aligned} \tag{31–33}$$

Where we adopt the notation $\cancel{v_i}$, which indicates that v_i is left out of the product. Suppose we replace $v_1..v_k$ by a set of orthogonal basisvectors $e_1...e_k$, such that we get

$$\frac{1}{2}(we_1...e_k - (-1)^k e_1...e_kw) = \sum_{l=1}^{k} (-1)^{k+1} w \cdot e_ke_1...\cancel{e_l}..e_k$$

Define the lefthand side to be $w \cdot E$, the innerproduct of vector w with blade E. Since any multivector A_r of grade r, can be written as the sum of a set of r-blades, we can also define the innerproduct of a vector with a grade r multivector A_r

$$w \cdot A_r = \frac{1}{2}(wA_r - (-1)^k A_r w)$$

Next, we wish to define the outerproduct between vectors and grade r multivectors. Starting from a general r-blade $e_1...e_r$ and a vector w we can write out the full antisymmetric product as in Equation 30

$$w \wedge e_1 \wedge ... \wedge e_r = \frac{1}{(r+1)!} \sum_{j=1}^{r} \sum_{\sigma \in S_r} \mathrm{sign}(\sigma) \cdot (-1)^j \cdot e_{\sigma(1)}...e_{\sigma(j)} w e_{\sigma(j+1)}...e_{\sigma(r)}$$

Collecting all terms that start with the same vector, we obtain

$$\frac{1}{(r+1)} we_1e_2...e_r + \frac{1}{(r+1)} \sum_{j=0}^{r} (-1)^j \cdot e_j (w \wedge e_1 \wedge ... \wedge \cancel{e_j} \wedge ... \wedge e_r) \tag{34}$$

Note that we have used that $e_1 \wedge ... \wedge e_r = e_1...e_r$, which is justified since e_i, e_j are mutually orthogonal.

If we assume that for an r-blade B_r it holds that $w \wedge B_r = \frac{1}{2}(wB_r + (-1)^r B_r w)$, and subtitute this expression for $w \wedge e_1 \wedge ... \wedge \cancel{e_j} \wedge ... \wedge e_r$, we get

$$\sum_{j=0}^{r} (-1)^j \cdot e_j (w \wedge e_1 \wedge ... \wedge \cancel{e_j} \wedge ... \wedge e_r) \tag{35}$$

$$= \sum_{j=0}^{r} (-1)^j \cdot \frac{1}{2} e_j (we_1...\cancel{e_j}...e_r + (-1)^{r-1} e_1...\cancel{e_j}...e_r) \tag{36}$$

$$= \frac{r}{2}(-1)^r e_1...e_r w + \frac{1}{2} \sum_{j=0}^{r} (-1)^j \cdot (w \cdot e_j) e_1...\cancel{e_j}...e_r + we_1...e_r \tag{37}$$

$$= \frac{r}{2}(we_1...e_r + (-1)^r e_1...e_r w) + \sum_{j=0}^{r} (-1)^j \cdot \frac{1}{2}(w \cdot e_j) e_1...\cancel{e_j}...e_r \tag{38}$$

Using equation 31 this becomes

$$= \frac{r}{2}(we_1...e_r + (-1)^r e_1...e_r w) + \frac{1}{2}((-1)^r e_1...e_r w - we_1...e_r) \tag{39}$$

$$= \frac{r-1}{2} we_1...e_r + \frac{r+1}{2}(-1)^r e_1...e_r w \tag{40}$$

Substituting this back into the right term of 34 yields

$$w \wedge e_1 \wedge ... \wedge e_r = \frac{1}{2}(we_1...e_r + (-1)^r e_1...e_r w) \tag{41}$$

In this derivation we assumed the existence of this identity for a blade of grade $r-1$, and since it holds by definition for a 1-blade, we can conclude that the identity is valid for arbitrary grade by induction.

Having defined both the inner product and outer product between a vector w and a r-blade A_r, the respective geometric product is easily expressed as

$$wA_r = w \cdot A_r + w \wedge A_r \tag{42}$$

Moreover, by inspecting the grades of the the inner and outer product we find

$$wA_r = (wA_r)^{(r-1)} + (wA_r)^{(r+1)} \tag{43}$$

Since each homogeneous multivector can be written as a sum of blades, we can iteravely apply the identity above for two homogeneous multivectors W_r, V_s

$$\begin{aligned}
W_r V_s &= \sum_m e_{1,m}...e_{r,m} V_s && (44)\\
&= \sum_m e_{1m}...e_{r-1,m}((e_{r,m}V_s)^{(s-1)} + (e_{r,m}V_s)^{(s+1)}) && (45)\\
&\vdots && (46)\\
&= (W_r V_s)^{|r-s|} + (W_r V_s)^{|r-s|+2} + \ldots + (W_r V_s)^{(r+s)} && (47)
\end{aligned}$$

This result tells us that the geometric product of two homogeneous multivectors only produces terms of grade m such that $|r-s| \leq m \leq r+s$ and $(r+s-m) \bmod 2 = 0$.

3.2 Pseudoscalars

Up to scalar multiplication, there exists a unique *pseudoscalar* for each Clifford algebra. This element is given by the blade of highest grade and is denoted with $I = e_1 \wedge \ldots \wedge e_n$, with e_i, e_j mutually orthogonal for $i \neq j$ and n being the dimension of the underlying vectorspace.

The pseudoscalar squares either to 1 or -1, since for an (non-degenerate) orthonormal basis, we can commute through elements at the cost of a minus sign and contract matching terms as $e_i e_i \in \{-1, 1\}$. Therefore, every ordering of an orthogonal set of vectors corresponds to the pseudoscalar with an associated sign, called the *orientation*.

Multiplying a blade of grade r by the pseudoscalar, returns a blade of grade $n-r$ formed by the space of vectors not in the original blade, referred to as the *orthogonal complement.*

In a space of even dimension, the pseudoscalar anticommutes with vectors, and thus with all oddly graded multivectors. In odd dimensional spaces, the pseudoscalar commutes with any element of the algebra, such that we can formulate for an homogeneous multivector A_r

$$IA_r = (-1)^{r(n-1)} A_r I \tag{48}$$

A practicality arises, as we can now transform inner products into outer products and vice versa. Consider a vector w and a blade A_r, such that

$$\begin{aligned} w \cdot (A_r I) &= \frac{1}{2}(wA_r I - (-1)^{n-r} A_r I w) \\ &= \frac{1}{2}(wA_r I - (-1)^{n-r}(-1)^{n-1} A_r w I) \\ &= \frac{1}{2}(wA_r I + (-1)^{r} A_r w I) \\ &= (w \wedge A_r) I \end{aligned}$$

3.3 Rotations

Suppose we want to reflect a vector a in the blade orthogonal to a unit vector n. Intuitively, we would do this by taking the component of a orthogonal to n and substracting the component parallel to n.

$$a = a_\perp + a_\parallel \to a_\perp - a_\parallel \tag{49}$$

Where $a_\parallel = (a \cdot n)n$ and $a_\perp = n(n \wedge a)$. Rewriting the reflection gives

$$\begin{aligned} a_\perp - a_\parallel &= n(n \wedge a) - (a \cdot n)n && (50) \\ &= \frac{1}{2}n(na - an) - \frac{1}{2}(na + an)n && (51) \\ &= -nan && (52) \end{aligned}$$

Two consecutive reflections in hyperplanes perpendicular to unit vectors m, n induce a rotation in the plane spanned by m and n.

$$a \to -n(-mam)n = nmamn$$

Next, consider a *rotor* R, in the $n \wedge m$ hyperplane, where $n \cdot m = \cos(\theta/2)$

$$R = \exp{(-\frac{\theta}{2}\hat{B})} = \exp{(-\frac{\theta}{2}\frac{m \wedge n}{\sin{(\theta/2)}})} \tag{53}$$

$$= 1 - \frac{\theta}{2}\frac{m \wedge n}{\sin{(\theta/2)}} + \frac{\theta^2}{8} - \frac{\theta^3}{48}\frac{m \wedge n}{\sin(\theta/2)} + \dots \tag{54}$$

$$= \cos(\theta/2) - \frac{\sin(\theta/2)}{\sin(\theta/2)} m \wedge n \tag{55}$$

$$= n \cdot m - m \wedge n = nm \tag{56}$$

Where we have used that $(m \wedge n)^2 = \sin^2(\theta/2)$. Such a rotor generates a rotates a vector a as follows

$$a \to a' = RaR^\dagger \tag{57}$$

Where $R^\dagger$ denotes the main anti-involution, previously denoted by $\zeta(R)$. We recognise that the rotor induces the exact same transformation as the sequential reflections in the hyperplanes perpendicular to m and n.

4 Towards Clifford group equivariant layers

Let us define the subgroup $Cl(V, \mathfrak{q})^\times$ of $Cl(V, \mathfrak{q})$, for which every $x \in Cl(V, \mathfrak{q})^\times$ has an inverse element x^{-1} such that $xx^{-1} = x^{-1}x = 1$. Note that this subspace forms a group under the Clifford algebra multiplication, since products of elements vw with an inverse, also have an inverse: $(vw)^{-1} = w^{-1}v^{-1}$.

We define the *twisted conjungation* for $v \in Cl(V, \mathfrak{q})^\times$ as follows:

$$\rho(v) : Cl(V, \mathfrak{q}) \to Cl(V, \mathfrak{q}) \quad \rho(v)(x) = vx^{[0]}v^{-1} + \alpha(v)x^{[1]}v^{-1}$$

Here α denotes the *main involution*, given by $\alpha(x) = x^{[0]} - x^{[1]}$.

This form of the twisted conjungation was chosen such that for a vector $x \in V$, $\rho(v)(x)$ reduces to a reflection in the hyperplane normal to v: $\rho(v)(x) = -vxv^{-1}$. Let us consider a particular subgroup $\Gamma(V, \mathfrak{q})$ of $Cl(V, \mathfrak{q})^\times$, called the *Clifford Group*, defined by

$$\Gamma(V, \mathfrak{q}) = \left\{ w \in Cl(V, \mathfrak{q})^\times \cap (Cl(V, \mathfrak{q})^{[0]} \cup Cl(V, \mathfrak{q})^{[1]} \middle| v \in V, \rho(w)(v) \in V \right\} \tag{58}$$

In words, this group contains all invertible elements of $Cl(V, \mathfrak{q})$ of *homogeneous parity* and for which the twisted conjungation maps vectors to vectors.

(i) $\rho(w)(x + y) = \rho(w)(x) + \rho(w)(y)$

(ii) $\rho(w)(xy) = \rho(w)(x)\rho(w)(y)$

(iii) $\rho(w^{-1})(x) = \rho(w)^{-1}(x)$

(iv) $\rho(w)(\rho(v)(x)) = (\rho(w) \circ \rho(v))(x) = \rho(wv)(x)$

(v) $\bar{\mathfrak{b}}(\rho(w)(x), \rho(w)(y)) = \bar{\mathfrak{b}}(x, y)$

(vi) $\rho(w)(x^{(m)}) = (\rho(w)(x))^{(m)}$

The first two identities make $\rho(w)$ into an algebra homomorphism (See Def. 1.8). Point (iii) shows that there exists an inverse such that $\rho(w)$ becomes a bijection. So, $\rho(w) : Cl(V, \mathfrak{q}) \to Cl(V, \mathfrak{q})$ is an algebra isomorphism onto itself, which is called an *algebra automorphism*.
Property (iv) states that a composition of two twisted conjungations $\rho(v), \rho(w)$ is again a twisted conjungation $\rho(vw)$. Therefore ρ is a group homomorphism to the group of algebra automorphisms $\mathrm{Aut}_{\mathbf{alg}}(Cl(V, \mathfrak{q}))$.
Identity (v) says that the extended bilinear form is maintained under transformation of $\rho(w)$. Intuitively this means that distances (and extended versions) are preserved. which implies identity (vi) following Theorem E.16 of [RBF23]. A twisted conjungation, followed by a projection on grade m, written as $\rho^{(m)}$, therefore maps the Clifford group to the orthogonal group (see definition 1.10) $O(Cl(V, \mathfrak{q})^{(m)})$:

$$\rho^{(m)} : \Gamma(V, \mathfrak{q}) \to O(Cl(V, \mathfrak{q})^{(m)}, \bar{\mathfrak{b}}) \tag{59}$$

Using these properties we see that polynomials of $P = F[T_1, ..., T_k]$ with coefficients in F, elements $x_1, ..x_k \in Cl(V, \mathfrak{q})$, behave covariantly with respect to $\rho(w)$ for $w \in \Gamma(V, \mathfrak{q})$

$$\rho(w)(P(x_1, ..., x_k)) = P(\rho(w)(x_1), ..., \rho(w)(x_k)) \tag{60}$$

Now, restating Theorem 3.5 of [RBF23], we say that when $(V, \mathfrak{q})$ is non-degenerate, the map

$$\bar{\rho}^{(1)} : \Gamma(V, \mathfrak{q})/F^{\times} \to O(Cl(V, \mathfrak{q})^{(1)}, \bar{\mathfrak{b}}) = O(V, \mathfrak{q})$$

where $\bar{\rho}^{(1)}([w]) := \rho(w)|_V$, is a group isomorphism. In words, the twisted conjungation restricted to vectors of a Clifford group element, up to non-zero scalar multiplication, is uniquely associated with an orthogonal transformation of $(V, \mathfrak{q})$. More specifically, this yields that for an element x in the Clifford algebra of the form $x = \sum_{i \in I} x_i \cdot v_{1i} ... v_{ni}$ with $x_i \in F$ and $v_{ij} \in V$ we have that for a $\bar{\rho}^{(1)}([w]) = \phi \in O(V, \mathfrak{q})$

$$\rho(w)(x) = \sum_{i \in I} x_i \cdot \rho(w)(v_{1i})...\rho(w)(v_{ni}) = \sum_{i \in I} x_i \cdot \phi(v_{1i})...\phi(v_{ni})$$

The key result here is that a function $f : Cl(V, \mathfrak{q})^{c_{\text{in}}} \to Cl(V, \mathfrak{q})^{c_{\text{out}}}$ is equivariant to the orthogonal group $O(V, \mathfrak{q})$ **iff** it is equivariant to the Clifford group $\Gamma(V, \mathfrak{q})$.
Now we can construct Clifford group equivariant network layers, and know they will be $O(V, \mathfrak{q})$-equivariant as well. In this manner, we introduce a *linear layer* for l multivectors $(x_1, ..., x_l) \in Cl(V, \mathfrak{q})^l$ as input:

$$y^{(k)}_{c_{\text{out}}} = \sum_{c_{\text{in}}=1}^{l} \phi_{c_{\text{in}} c_{\text{out}} k} \cdot x^{(k)}_{c_{\text{in}}} \tag{61}$$

Here, the $\phi_{c_{\text{in}} c_{\text{out}} k}$ are learnable scalar parameters. Essentially, this layer aplies a linear transformation on each grading subalgebra. Note that the map is equivariant with respect to the Clifford group by considering that for every $w \in \Gamma(V, \mathfrak{q})$

$$\begin{aligned}
\rho(w)(y^{(k)}_{c_{\text{out}}}) &= \rho(w)(\sum_{c_{\text{in}}=1}^{l} \phi_{c_{\text{in}} c_{\text{out}} k} \cdot x^{(k)}_{c_{\text{in}}}) \\
&= \sum_{c_{\text{in}}=1}^{l} \rho(w)(\phi_{c_{\text{in}} c_{\text{out}} k} \cdot x^{(k)}_{c_{\text{in}}}) \\
&= \sum_{c_{\text{in}}=1}^{l} \phi_{c_{\text{in}} c_{\text{out}} k} \cdot \rho(w)(x^{(k)}_{c_{\text{in}}}) \\
&= \sum_{c_{\text{in}}=1}^{l} \phi_{c_{\text{in}} c_{\text{out}} k} \cdot \rho(w)(x_{c_{\text{in}}})^{(k)}
\end{aligned}$$

In the last step we made use of the fact that grade projections are equivariant to the Clifford group. We see that when the input is tranformed by an element of the Clifford group, the output of this linear layer transforms accordingly. Consequently, this linear layer is also $O(V, \mathfrak{q})$-equivariant.

Since we saw that the geometric product is also preserved under actions of the Clifford group, we can construct another type of layer, called the *Geometric Product Layer*. First, consider two elements $x_1, x_2 \in Cl(V, \mathfrak{q})$, and their grading decomposition $x_1 = \sum_i^n x_1^{(i)}, x_2 = \sum_j^n x_2^{(j)}$. The map $P_\phi : Cl(V, \mathfrak{q}) \times Cl(V, \mathfrak{q}) \to Cl(V, \mathfrak{q})$ is then given by

$$P^{(k)}_\phi(x_1, x_2) = \sum_{i,j \in [n]} \phi_{ijk} \cdot \left(x_1^{(i)} x_2^{(j)}\right)^{(k)} \quad \text{with} \quad i, j, k \in [n]$$

This map provides a pairwise geometric product, where particular gradings are weighted and can be learned. Again, note that this map is equivariant with respect to the Clifford group, using $\rho(w)(x_1 x_2) = \rho(w)(x_1)\rho(w)(x_2)$ together with the reasons above. When one has an input tuple of l multivectors, one could consider a layer that computes all pairwise products, but for large l this becomes infeasable quickly. Therefore, an alternative aproach is to apply a linear layer to input $(x_1, ..., x_l) \in Cl(V, \mathfrak{q})^l$, to obtain $(y_1, ..., y_l) \in Cl(V, \mathfrak{q})^l$, and perform an elementwise product, such that [1]

[1] In the equations below, upper indices don't indicate Einstein summation, but are used for space considerations.

$$z_{c_{\text{out}}}^{(k)} = \sum_{i,j\in[n]} \phi_{ijk}^{c_{\text{in}}} \cdot \delta(c_{\text{in}}, c_{\text{out}}) \cdot \left(x_{c_{\text{in}}}^{(i)} y_{c_{\text{in}}}^{(j)}\right)^{(k)} \tag{62}$$

which is called the *elementwise geometric product layer*. One can also construct a generalization of this layer by learning a linear combination of these products

$$z_{c_{\text{out}}}^{(k)} = \sum_{c_{\text{in}}} \sum_{i,j\in[n]} \phi_{ijk}^{c_{\text{in}} c_{\text{out}}} \cdot \left(x_{c_{\text{in}}}^{(i)} y_{c_{\text{in}}}^{(j)}\right)^{(k)} \tag{63}$$

This *fully connected* geometric layer is more expressive, at the cost of requiring more parameters.

5 Optimal transport

In this chapter we will mainly follow the book by [PC19].

5.1 Measures

In the study of optimal transport, the goal to find an optimal map from one *measure* to another. A *measure* can be intuitively be thought of as a density function, where each point in some space $\mathcal{X}$ carries some amount of mass. A discrete measure α, with a set of n point masses can be written as:

$$\alpha = \sum_{i=1}^{n} \mathbf{a}_i \delta(x - x_i), \quad \mathbf{a} \in \Sigma \tag{64}$$

Where $\mathbf{a}$ is a *probability vector* defined to be in the probability simplex Σ, given by

$$\Sigma = \{\mathbf{a} \in \mathbb{R}_+^n | \sum_i \mathbf{a}_i = 1\} \tag{65}$$

In the context of machine learning, discrete measures are practically all one needs to consider. In a more general setting however, a measure on $\mathcal{X}$, can be expressed as a density $d\alpha(x) = \rho_\alpha(x)dx$, such that for all continuous functions f over $\mathcal{X}$ we have

$$\int_{\mathcal{X}} f(x) d\alpha(x) = \int_{\mathcal{X}} f(x) \rho_\alpha(x) dx \in \mathbb{R} \tag{66}$$

In the case $\mathcal{X}$ is non-compact, f should either have compact support, or should vanish in the limit. Furthermore, α is bound to be non-negative at each point and should integrate to one: $\int_{\mathcal{X}} d\alpha(x) = 1$.

5.2 Assignment problem

Suppose we have two discrete measures α and β, both consisting of n equally weighted point masses.

$$\alpha = \sum_{i=1}^{n} \frac{1}{n} \delta(x - x_i), \quad \beta = \sum_{j=1}^{n} \frac{1}{n} \delta(x - x_j) \tag{67}$$

Let $\mathbf{C} \in \mathbb{R}^{n \times n}$ be a cost matrix, where element $\mathbf{C}_{i,j}$ measures a cost between positions x_i and x_j. The goal is now to assign a pairing, or a bijective map, such that the total cost is minimized:

$$\min_{\sigma \in S_n} \sum_{i=1}^{n} \frac{1}{n} \mathbf{C}_{i,\sigma(i)} \tag{68}$$

Note that a solution to this problem is not neccesarily unique, since multiple permutations can yield the same total cost. Consider for example α and β both consisting of two point masses, with mutually equivalent costs. Then, every permutation yields the same (minimal) cost.

5.3 Monge problem

The Monge problem is a generalization of this problem, and does not constrain the point masses to be equally weighted, nor does the map need to be bijective. Instead, given the measures α and β with

$$\alpha(x) = \sum_{i=1}^{n} \mathbf{a}_i \delta(x - x_i), \quad \beta(y) = \sum_{j=1}^{m} \mathbf{b}_j \delta(y - y_j) \tag{69}$$

one seeks a *Push-forward* function $T : \{x_1, .., x_n\} \to \{y_1, .., y_m\}$ that maps each point x_i to a single point y_i, such that for all $j \in [m]$, T satisfies

$$\mathbf{b}_j = \sum_{i \in \{i : T(x_i) = y_j\}} \mathbf{a}_i \tag{70}$$

and requires minimal transportation cost in relation to a cost function $c(x, y) : \mathcal{X}, \mathcal{Y} \to \mathbb{R}$. So, the aim is to find a Monge map T, such that

$$\min_{T} \Big\{ \sum_{i} c(x_i, T(x_i)) \Big\} \tag{71}$$

Again, since it forms a generalization of the assignment problem, there might not be a unique solution. In this setting however, there might not even exist a valid Monge map. Take m to be larger than n for example. Since a Monge map T takes each point x_i only to single point y_j, there will be a point mass in β not gaining any mass, so equation 70 cannot be met.

For continuous measures $d\alpha(x), d\beta(y)$ the notion of the push-forward function can naturally be extendend to a function $T : \mathcal{X} \to \mathcal{Y}$, such that for any continuous function f on $\mathcal{Y}$ it holds that

$$\int_{\mathcal{Y}} f(y) d\beta(y) = \int_{\mathcal{X}} f(T(x)) d\alpha(x) \tag{72}$$

The Monge problem is then stated by finding such a function that minimizes the total cost.

$$\min_{T} \int_{\mathcal{X}} c(x, T(x)) d\alpha(x) \tag{73}$$

5.4 Kantorovich relaxation

We saw before the formulation of the assignment problem and the Monge problem. The assignment problem only being defined for permutations between discrete uniform distributions of the same size. The less restrictive Monge problem allowed for non-uniform measures, but is difficult to solve due to its *non-convexity*[PC19], besides the fact a solution is not assured.

Kantorovich introduced a deviation to the Monge problem by relaxing accepting *non-deterministic* mass transportation. Whereas a Monge map T can only push mass from a point mass x_i to a **single** position y_j, Kantorovich allows for *mass splitting*, meaning that mass from a source can be spread among different targets. For discrete measures, this idea can be captured in a coupling matrix $\mathbf{P} \in \mathbb{R}_+^{n\times m}$, where element $\mathbf{P}_{ij}$ represent the amount of mass moving from position x_i to y_j. Given two probability vectors $\mathbf{a} \in \mathbb{R}^n, \mathbf{b} \in \mathbb{R}^m$, the set of valid couplings $\mathbf{U}(\mathbf{a}, \mathbf{b})$ transform $\mathbf{a}$ into $\mathbf{b}$, such that

$$\mathbf{U}(\mathbf{a}, \mathbf{b}) = \left\{ \mathbf{P} \in \mathbb{R}_+^{n\times m} | \quad \sum_{j=1}^{m} \mathbf{P}_{ij} = \mathbf{a}_i \ , \ \sum_{i=1}^{n} \mathbf{P}_{ij} = \mathbf{b}_j \right\} \tag{74}$$

The Kantorovich relaxed problem for a cost matrix $\mathbf{C}$ is then formulated as

$$L_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) = \min_{\mathbf{P}\in\mathbf{U}(\mathbf{a},\mathbf{b})} \sum_{i,j} P_{ij} C_{ij} \tag{75}$$

For continuous problems, the coupling matrix becomes a product measure, often denoted by π. In order to encode the mass conservation, we require that π satisfies

$$\int_{\mathcal{X}} d\pi(x, y) = d\beta(y), \quad \int_{\mathcal{Y}} d\pi(x, y) = d\alpha(x) \tag{76}$$

We define the set of admissable product measures $\mathcal{U}(\alpha, \beta)$ as all product measures on $\mathcal{X} \times \mathcal{Y}$ such that equation 76 holds. Note that a product measure of the form $\pi(x, y) = \alpha(x)\beta(y)$ automatically satisfies the required conditions, which implies that $\mathcal{U}(\alpha, \beta)$ is nonempty. Therefore the generalized Kantorovich problem, given by

$$\mathcal{L}_c(\alpha, \beta) = \min_{\pi\in\mathcal{U}(\alpha,\beta)} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y) d\pi(x, y) \tag{77}$$

will always have a solution. blue

5.5 Wasserstein metric

Solutions to the Optimal transport problem quantify the amount of cost needed to move one distribution into another. Therefore it provides an alternative measure of *distance* between distributions to more conventional methods such as Kullbeck-Leibler divergence. It turns out however, that for $\mathcal{L}_c(\alpha, \beta)$ to be a reasonable distance, meaning that it

(i) is symmetric in it's arguments: $\mathcal{L}_c(\alpha, \beta) = \mathcal{L}_c(\beta, \alpha)$

(ii) is non-negative $\mathcal{L}_c(\alpha, \beta) \geq 0$

(iii) $\mathcal{L}_c(\alpha, \beta) = 0$ if and only if $\alpha = \beta$

(iv) satisfies the triangle inequality: $\mathcal{L}_c(\alpha, \gamma) \leq \mathcal{L}_c(\alpha, \beta) + \mathcal{L}_c(\beta, \alpha)$

, we have to make a particular choice of cost function. If we assume that $\mathcal{X} = \mathcal{Y}$, and impose that for $c(x, y) = d(x, y)^p$ for $p \geq 1$, where $d(x, y)$ satisfies the conditions

(i) $d(x, y) = d(y, x) \geq 0$

(ii) $d(x, y) = 0$ iff $x = y$

(iii) $d(x, z) \leq d(x, y) + d(y, z)$

, then $\mathcal{L}_{d^p}(\alpha, \beta)^{1/p}$ will be a proper distance referred to as the *p-Wasserstein* metric. The non-negativity, symmetry condition and $\mathcal{L}_c(\alpha, \beta) = 0 \iff \alpha = \beta$ follow trivially from the monotonicity on the *ground-metric* (or cost function). For a proof of the triangle inequality see Propostion 2.2 of [PC19].

5.6 Dual problem

to be written

5.7 Entropic regularization

Consider a discrete Kantorovich problem with a coupling matrix $\mathbf{P} \in \mathbb{R}_+^{n \times m}$ together with it's entropy $H(\mathbf{P})$ given by

$$H(\mathbf{P}) = -\sum_{i,j} \mathbf{P}_{ij}(\log(P_{ij}) - 1), \tag{78}$$

which gives a measure of how much information is encoded in $\mathbf{P}$. Note that the entropy of a coupling matrix is a strongly concave function since $\mathbf{P}_{ij} \leq 1$ for all i, j, and the Hessian is given by $\partial^2 H(\mathbf{P}) = -\mathrm{diag}(1/\mathbf{P}_{ij}) \leq 0$. The idea is now to use the negative Hessian as a regulizer for the Kantorovich problem as stated in equation 75. Such that it becomes

$$L_{\mathbf{C}}^{\epsilon}(\mathbf{a}, \mathbf{b}) = \min_{\mathbf{P} \in \mathbf{U}(\mathbf{a}, \mathbf{b})} \sum_{i,j} P_{ij} C_{ij} - \epsilon H(\mathbf{P}), \tag{79}$$

where ϵ is a real, non-negative regulization parameter. In this way we give preference to coupling matrices of higher entropy, intuitively understood by the amount of *mixing* that is induced between the distributions. Recall that the standard Kantorovich problem did not necessarily have a unique solution. But when we decrease ϵ towards zero, a unique solution arises that solves the Kantorovich problem with maximal entropy. The uniqueness of this solution is due to the convexity of the entropic term.

On the other hand, when we take ϵ to be infinitely large, such that problem effectively determines the coupling matrix $\mathbf{P} \in \mathbf{U}(\mathbf{a}, \mathbf{b})$ with maximal entropy, one can show this resolves to $\mathbf{P} = \mathbf{a}\mathbf{b}^T$. So, for $\epsilon \to 0$, the optimal solution is a solution to the Kantorovich problem with maximal entropy, while for $\epsilon \to \infty$ we have that the optimal solution consist of the independent joint probability distribution $\mathbf{P} = \mathbf{a}\mathbf{b}^T$.

We can restate the regularized Kantorovich problem in terms of the Kullbeck-Leibler divergence, defined as

$$\mathrm{KL}(\mathbf{P}|\mathbf{K}) = \sum_{i,j} P_{ij} \log\left(\frac{\mathbf{P}_{ij}}{\mathbf{K}_{ij}}\right) - \mathbf{P}_{ij} + \mathbf{K}_{ij}, \tag{80}$$

where $\mathbf{K}$ is a Gibbs kernel with respect to the cost matrix $\mathbf{C}$, such that $\mathbf{K}_{ij} = \exp\left(-\mathbf{C}_{ij}/\epsilon\right)$. The equivalent problem is now expressed as

$$\min_{\mathbf{P} \in \mathbf{U}(\mathbf{a},\mathbf{b})} \mathrm{KL}(\mathbf{P}|\mathbf{K}). \tag{81}$$

This is easily verified by plugging the definition of $\mathbf{K}$ into the the KL-divergence and comparing to equation 79.

5.8 Sinkhorn's algorithm

In equation 81 we seek to find the coupling matrix $\mathbf{P}$ that minimizes the KL-divergens, subject to the constraint that the marginals satisfy

$$\sum_j \mathbf{P}_{ij} = \mathbf{a}_i, \quad \sum_i \mathbf{P}^T_{ij} = \mathbf{b}_j.$$

This constrained minimization problem is then naturally formulated using a *Lagrange function* as follows

$$\mathfrak{L}(\mathbf{P}, \mathbf{u}, \mathbf{v}) = \mathrm{KL}(\mathbf{P}|\mathbf{K}) - \sum_i \mathbf{u}_i \Big(\sum_j \mathbf{P}_{ij} - \mathbf{a}_i\Big) - \sum_j \mathbf{v}_j \Big(\sum_i \mathbf{P}_{ij} - \mathbf{b}_j\Big), \tag{82}$$

where the elements of $\mathbf{u} \in \mathbb{R}^n$ and $\mathbf{v} \in \mathbb{R}^m$ serve as Lagrange multipliers. Taking the gradient with respect to an element $\mathbf{P}_{kl}$ then yields the condition

$$\frac{\partial \mathfrak{L}(\mathbf{P}, \mathbf{u}, \mathbf{v})}{\partial \mathbf{P}_{kl}} = \epsilon \log(\mathbf{P}_{kl}) + \mathbf{C}_{kl} - \mathbf{u}_k - \mathbf{v}_l = 0. \tag{83}$$

Solving for $\mathbf{P}_{kl}$ gives

$$\begin{aligned}
\mathbf{P}_{kl} &= \exp\left(\frac{\mathbf{v}_l + \mathbf{u}_k - \mathbf{C}_{kl}}{\epsilon}\right) && (84)\\
&= \exp(\mathbf{u}_k/\epsilon)\mathbf{K}_{kl} \exp(\mathbf{v}_l/\epsilon) && (85)\\
&= \mathbf{s}_k \mathbf{K}_{kl} \mathbf{t}_l && (86)
\end{aligned}$$

in which we recognized $\exp(-\mathbf{C}_{kl}/\epsilon)$ as the Gibbs kernel, and defined $\mathbf{s} = \exp(\mathbf{u}/\epsilon) \in \mathbb{R}^n_+$ and $\mathbf{t} = \exp(\mathbf{v}/\epsilon) \in \mathbb{R}^m_+$.

Therefore the unique solution of equation 81 is compactly expressed in matrix form as

$$\mathbf{P} = \mathrm{diag}(\mathbf{s})\mathbf{K}\mathrm{diag}(\mathbf{t}), \tag{87}$$

conditioned by the marginalizations

$$\sum_j \Big(\mathrm{diag}(\mathbf{s})\mathbf{K}\mathrm{diag}(\mathbf{t})\Big)_{ij} = \Big(\mathrm{diag}(\mathbf{s})\mathbf{K}\mathbf{t}\Big)_i = \mathbf{a}_i, \tag{88}$$

$$\sum_i \Big(\mathrm{diag}(\mathbf{s})\mathbf{K}\mathrm{diag}(\mathbf{t})\Big)_{ij} = \Big(\mathbf{s}\mathbf{K}\mathrm{diag}(\mathbf{t})\Big)_j = \mathbf{b}_j \tag{89}$$

In short, we get

$$\mathbf{s} \odot \mathbf{K}\mathbf{t} = \mathbf{a}, \quad \mathbf{s}\mathbf{K} \odot \mathbf{t} = \mathbf{b} \tag{90}$$

where $\odot$ represents the elementwise product. From here, Sinkhorn's algorithm provides a iterative method for evaluating $\mathbf{s}$ and $\mathbf{t}$. Starting at a certain initalization $\mathbf{s}^{(0))}, \mathbf{t}^{(0))}$ we apply the following update rule:

$$\mathbf{s}^{(k)} = \mathbf{a}/\mathbf{K}\mathbf{t}, \quad \mathbf{t}^{(k)} = \mathbf{b}/\mathbf{s}\mathbf{K} \tag{91}$$

where elementwise division is implied.

5.9 Next

1. Dual problem
2. Tensor transport
3. Multivector transport?

6 Multivector transport

Consider a discrete measure μ on a space $\mathcal{X}$ over a Clifford algebra

$$\mu(x) = \sum_{i=1}^{n} \mathbf{m}_i \delta(x - x_i), \quad \mathbf{m}_j \in Cl(V, \mathfrak{q}) \quad \forall j \in [n], \tag{92}$$

and a measure ν on $\mathcal{Y}$

$$\nu(y) = \sum_{k=1}^{m} \mathbf{n}_k \delta(y - y_k), \quad \mathbf{n}_l \in Cl(V, \mathfrak{q}) \quad \forall l \in [m]. \tag{93}$$

Similar to the Kantarovich problem, we could define a coupling matrix $\mathbf{P} \in Cl(V, \mathfrak{q})^{n \times m}$, and require that

$$\sum_i \mathbf{P}_{ik} = \nu_k = \nu(y_k), \quad \sum_k \mathbf{P}_{ik} = \mu_i = \mu(x_i). \tag{94}$$

However, this is likely to be too restrictive, since there generally doesn't exist a coupling matrix that satisfies these constraints. For the sake of illustration, let $\mu = e_1 \delta(x - x_1)$ and $\nu = e_2 \delta(x - x_1)$ be two measures, then a coupling matrix $\mathbf{P} \in Cl(V, \mathfrak{q})^{1 \times 1}$, consisting of only one element, can never satisfy $\mathbf{P}_{11} = e_1 = e_2$.

More generally, every element in $\mathbf{P}$ has 2^N degrees of freedom (with N being the dimensionality of the underlying vector field). So, in total $\mathbf{P}$ has $nm2^N$ degrees of freedom. Whereas there are $(n + m)2^N$ constraints, captured in equation 94. So the set of valid coupling matrices is found by solving for this system of linear equations. But, since $span\{\hat{e}\}$, with $\hat{e}$ being a basis element of $Cl(V, \mathfrak{q})$, is closed under addition, we can treat the system as 2^N parallel systems.

$$\sum_i \hat{e} \cdot \mathbf{P}_{ik} = \hat{e} \cdot \nu_k, \quad \sum_k \hat{e} \cdot \mathbf{P}_{ik} = \hat{e} \cdot \mu_i \quad \forall \hat{e} \in E. \tag{95}$$

These are however, contrary to the scalar transport problems, defined with measures that are both *signed* an *unbalanced*. Meaning mass is allowed to be negative and is generally not conserved. A violation of mass conservation would imply that $\sum_i \mu_i \neq \sum_k \nu_k$, which is problematic since the set of constraints in 94 becomes inconsistent,

$$\sum_{i,k} \mathbf{P}_{ik} = \sum_i \mu_i = \sum_k \nu_k. \tag{96}$$

We could decide to constrain the measures to have non-negative masses, adding up to 1 in each dimension, but this seems rather artificial. Other methods that circumvent this problem include the use of dual norms, such as in [PRT19], or relaxing the constraints altogether in exchange for a penalty term. The latter is perhaps the most logical in the context of imaging and ML, as the data that is dealt with is not exact anyways.

Such a penalty term would probably consist of some divergence D_ϕ, that measures the mass discrepancy and the rotation discrepancy between the found and exact marginalization. So, the optimization problem would take the form of

$$OMT_C(\mu,\nu)^\tau = \min_{\mathbf{P}\in Cl(V,\mathfrak{q})^{n\times m}} \mathbf{C}(\mathbf{P}) + \tau_1 D_\phi(\mathbf{P}\mathbb{1}_m|\mu) + \tau_1 D_\phi(\mathbf{P}^T\mathbb{1}_n|\nu), \quad (97)$$

where τ_1, τ_2 are parameters that control the weight of the penalty and $\mathbf{C}(\mathbf{P})$ is a transportation cost function, which is most naturally formulated as

$$\mathbf{C}(\mathbf{P}) = \sum_{i,j} \bar{\mathfrak{q}}(\mathbf{P}_{ij}) \cdot c(x_i, y_j). \quad (98)$$

Here $c : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ is a function that serves as the *ground metric*, relating tuples of points in $\mathcal{X}$ and $\mathcal{Y}$ to distances. And we interpret the extended quadratic form of element $\mathbf{P}_{ij}$, given by

$$\bar{\mathfrak{q}}(\mathbf{P}_{ij}) = \zeta(\beta(\mathbf{P}_{ij})\mathbf{P}_{ij}) = (\beta(\mathbf{P}_{ij})\mathbf{P}_{ij})^{(0)}, \quad (99)$$

as the amount of *mass* moved from x_i to y_j.

For Patrick:

The exact formulation of the divergence D_ϕ is still to be chosen, and probably decides how straightforward this can be extended to a Sinkhorn-like algorithm. It will probably consist of a term that penalizes depending on whether too much or too little mass is moved somewhere, and a term that measures how well the marginal is aligned with the target.
Apart from including that last term, I don't yet see how the structure of the Clifford algebra comes in useful. Is my thinking in the right direction, or did you had something else in mind?

7 Questions

- Should the ideal used to define Clifford algebra also have an index?
- Proof of associativity of outer product?
- Has is it been tried to use a linear layer like this:

$$y_{c_{\text{out}}}^{(k)} = \sum_{c_{\text{in}}=1}^{l} \phi_{c_{\text{in}} c_{\text{out}} k} \cdot (x_{c_{\text{in}}})^{(k)} + \tilde{\phi}_{c_{\text{in}} c_{\text{out}} k} \cdot (I x_{c_{\text{in}}})^{(k)}$$

- Tensor transport: similar idea. captures directional data and rotations.

  Defining transport between tensor measures. Unlike scalar valued measures, exact maps do not always exist.

  This paper extends the product measure to a tensor product and requires partial traces give the correct marginal. `https://arxiv.org/pdf/1304.3931.pdf`

  This one proposes a relaxation based on the von neumann entropy. `https://arxiv.org/pdf/1612.08731.pdf`

- Multivector transport?
  Exact maps don't exist.

  Define a way to transport multivectors, and define a cost for translation/rotation.

  Mathematically interesting: Fluid dynamics formulation of optimal transport. In the case of multivector this would possibly look very much like non-relativistic wave equations for free fermionic fields.

- What imaging method to apply on and how would the Clifford group equivariant NN come in?
- Second reader. Leo dorst

Lemma 7.1. *Let $u, v, w \in V$, it holds that*

$$u \wedge (v \wedge w) = (u \wedge v) \wedge w$$

Proof.

$$\begin{aligned} u \wedge (v \wedge w) &= u \wedge \frac{1}{2}(vw - wv) \\ &= \frac{1}{4}(uvw - vwu - uwv + wvu) \end{aligned}$$

and

$$\begin{aligned} (u \wedge v) \wedge w &= \frac{1}{2}(uv - vu) \wedge w \\ &= u \wedge \frac{1}{2}(vw - wv) \\ &= \frac{1}{4}(uvw - vuw - wuv + wvu) \end{aligned}$$

Matching terms, we see that the only thing left to show is that

$$vwu + uwv = vuw + wuv$$

Now consider a decomposition $u = u_{\parallel} + u_{\perp}$, such that $u_{\parallel} w = w u_{\parallel}$ and $u_{\perp} w = -w u_{\perp}$. Since parallel terms commmute, the expression reduces to

$$vwu_{\perp} + u_{\perp}wv = vu_{\perp}w + wu_{\perp}v$$

Decomposing v in similar fashion, gives

$$(v_{\parallel} + v_{\perp})wu_{\perp} + u_{\perp}w(v_{\parallel} + v_{\perp}) = (v_{\parallel} + v_{\perp})u_{\perp}w + wu_{\perp}(v_{\parallel} + v_{\perp})$$

Commuting and anticommuting parallel and respectively perpendicular vectors to the right, we obtain

$$-wu_{\perp}(v_{\parallel} + v_{\perp}) + u_{\perp}w(v_{\parallel} + v_{\perp}) = -u_{\perp}w(v_{\parallel} + v_{\perp}) + wu_{\perp}(v_{\parallel} + v_{\perp})$$

Finally, anticommuting $u_{\perp}$ and w gives

$$2u_{\perp}w(v_{\parallel} + v_{\perp}) = -u_{\perp}w(v_{\parallel} + v_{\perp}) + wu_{\perp}(v_{\parallel} + v_{\perp})$$

$\square$

References

[DL03] Chris Doran and Anthony Lasenby. *Geometric Algebra for Physicists.* Cambridge University Press, 2003. DOI: 10.1017/CBO9780511807497.

[LS09] Douglas Lundholm and Lars Svensson. *Clifford algebra, geometric algebra, and applications.* 2009. arXiv: 0907.5356 [math-ph].

[PC19] Gabriel Peyre and Marco Cuturi. "Computational Optimal Transport". In: *Foundations and Trends in Machine Learning* 11.5-6 (2019), pp. 355–607.

[PRT19] Benedetto Piccoli, Francesco Rossi, and Magali Tournus. *A Wasserstein norm for signed measures, with application to nonlocal transport equation with source term.* 2019. arXiv: 1910.05105 [math.AP].

[RBF23] David Ruhe, Johannes Brandstetter, and Patrick Forré. *Clifford Group Equivariant Neural Networks.* 2023. arXiv: 2305.11141 [cs.LG].