KRUSKAL'S THEOREM

J.M. LANDSBERG

Abstract. This is just a short proof of Kruskal's theorem regarding uniqueness of expressions for tensors, phrased in geometric language.

Let A, B, C be complex vector spaces of dimensions $\mathbf{a}, \mathbf{b}, \mathbf{c}$. Consider a tensor $T \in A \otimes B \otimes C$ and say we have an expression

$$T = u_1 \otimes v_1 \otimes w_1 + \cdots + u_r \otimes v_r \otimes w_r \tag{1}$$

where $u_j \in A, v_j \in B, w_j \in C$, and we want to know if the expression is unique up to re-ordering the factors (call this *essentially unique*). The *rank* of T is by definition the smallest such r such that T admits an expression of the form (1). For the tensor product of two vector spaces, an expression as a sum of r elements is never unique unless $r = 1$. Thus an obvious necessary condition for uniqueness is that we cannot be reduced to a two factor situation. For example, an expression of the form

$$T = a_1 \otimes b_1 \otimes c_1 + a_1 \otimes b_2 \otimes c_2 + a_3 \otimes b_3 \otimes c_3 + \ldots + a_r \otimes b_r \otimes c_r$$

where each of the sets $\{a_i\}, \{b_j\}, \{c_k\}$ are linearly independent is not unique because of the first two terms. In other words if we consider for (1) the sets $\mathcal{S}_A = \{[u_i]\} \subset \mathbb{P}A$, $\mathcal{S}_B = \{[v_i]\} \subset \mathbb{P}B$, $\mathcal{S}_C = \{[w_i]\} \subset \mathbb{P}C$ each of the sets must consist of r distinct points.

We recall the classical fact:

Proposition 1. *Let $n > 2$. Let $T \in A_1 \otimes \cdots \otimes A_n$ have rank r. Say $T \in A_1' \otimes \cdots \otimes A_n'$, where $A_j' \subseteq A_j$, with at least one inclusion proper. Then any expression $T = \sum_{i=1}^{\rho} u_i^1 \otimes \cdots \otimes u_i^n$ with some $u_j^s \notin A_s'$ has $\rho > r$.*

Proof. Choose complements A_t'' so $A_t = A_t' \oplus A_t''$. Assume $\rho = r$ and write $u_j^t = u_j^{t\prime} + u_j^{t\prime\prime}$ with $u_j^{t\prime} \in A_t'$, $u_j^{t\prime\prime} \in A_t''$. Then $T = \sum_{i=1}^{\rho} u_i^{1\prime} \otimes \cdots \otimes u_i^{n\prime}$ so all the other terms must cancel. Assume $\rho = r$, and say, e.g., some $u_{j_0}^{1\prime\prime} \neq 0$. Then $\sum_{j=1}^{r} u_j^{1\prime\prime} \otimes (u_j^{2\prime} \otimes \cdots \otimes u_j^{n\prime}) = 0$, but all the terms $(u_j^{2\prime} \otimes \cdots \otimes u_j^{n\prime})$ must be linearly independent in $A_2' \otimes \cdots \otimes A_n'$ otherwise r would not be minimal, thus all the $u_j^{1\prime\prime}$ must all be zero, a contradiction. □

Definition 2. Let $\mathcal{S} = \{x_1, ..., x_p\} \subset \mathbb{P}W$ be a set of points. We say the points of $\mathcal{S}$ are in 2-general linear position if no two points coincide, they are in 3-general linear position if no three lie on a line and more generally they are in *r-general linear position* if no $r - 1$ of them lie in a $\mathbb{P}^{r-2}$. We let the *Kruskal rank* of $\mathcal{S}$, $k_{\mathcal{S}}$, be the maximum number r such that the points of $\mathcal{S}$ are in r-general linear position.

If one chooses a basis for W so that the points of $\mathcal{S}$ can be written as columns of a matrix (well defined up to rescaling columns), then $k_{\mathcal{S}}$ will be the maximum number r such that all subsets of r column vectors of the corresponding matrix are linearly independent. (This was Kruskal's original definition.)

Supported by NSF grant DMS-DMS-0805782.

Theorem 3 (Kruskal,[1])**.** *Let* $T \in A\otimes B\otimes C$. *Say* T *admits an expression* $T=\sum_{i=1}^r u_i\otimes v_i\otimes w_i$. *Let* $\mathcal{S}_A=\{[u_i]\}$,$\mathcal{S}_B=\{[v_i]\}$,$\mathcal{S}_C=\{[w_i]\}$. *If*

$$r\leq \frac{1}{2}(k_{\mathcal{S}_A}+k_{\mathcal{S}_B}+k_{\mathcal{S}_C})-1 \tag{2}$$

then T *has rank* r *and its expression as a rank* r *tensor is essentially unique.*

Above, we saw a necessary condition for uniqueness is that $k_{\mathcal{S}_A},k_{\mathcal{S}_B},k_{\mathcal{S}_C}\geq 2$ and it is an easy exercise to show that if (2) holds, then $k_{\mathcal{S}_A},k_{\mathcal{S}_B},k_{\mathcal{S}_C}\geq 2$. (Hint: *a priori* $k_{\mathcal{S}_A}\leq r$.)

Note that if $\mathbf{a}=\mathbf{b}=\mathbf{c}$ and $T:(A\otimes B)^*\to C$ and similar permutations are surjective, then it is very easy to see such an expression is unique when $r=\mathbf{a}$. Kruskal's Theorem extends the uniqueness to $\mathbf{a}\leq r\leq \frac{3}{2}\mathbf{a}-1$.

The key to the proof of Kruskal's theorem is the following lemma:

Lemma 4 (Permutation lemma)**.** *Let* W *be a complex vector space and let* $\mathcal{S}=\{p_1,...,p_r\}$, $\tilde{\mathcal{S}}=\{q_1,...,q_r\}$ *be sets of points in* $\mathbb{P}W$ *and assume no two points of* $\mathcal{S}$ *coincide (i.e., that* $k_{\mathcal{S}}\geq 2$*) and that* $\langle\tilde{\mathcal{S}}\rangle=W$. *If all hyperplanes* $H\subset\mathbb{P}W$ *that have the property that they contain at least* $\dim(H)+1$ *points of* $\tilde{\mathcal{S}}$ *also have the property that* $\#(\mathcal{S}\cap H)\geq\#(\tilde{\mathcal{S}}\cap H)$, *then* $\mathcal{S}=\tilde{\mathcal{S}}$.

If one chooses a basis for $W=\mathbb{C}^n$ and writes the two sets of points as matrices $M,\tilde{M}$, then the hypothesis can be rephrased (in fact this was the original phrasing) as to say that for all $x\in\mathbb{C}^n$ such that the number of nonzero elements of the vector ${}^t\tilde{M}x$ is less than $r-\mathrm{rank}(\tilde{M})+1$ also has the property that the number of nonzero elements of the vector ${}^t\tilde{M}x$ is at most the number of nonzero elements of the vector ${}^tMx$. To see the correspondence, the vector x should be thought of as point of W^* giving an equation of H, zero elements of the vector ${}^t\tilde{M}x$ correspond to columns that pair with x to be zero, i.e., that satisfy an equation of H, i.e., points that are contained in H.

Note a slight discrepancy with the original formulation: we have assumed $\langle\tilde{\mathcal{S}}\rangle=W$ so $\mathrm{rank}(\tilde{M})=n$. Our hypothesis is slightly different, but it is all that is needed by Proposition 1. Had we not assumed this, there would be trivial cases to eliminate at each step of our proof.

Proof. First note that if one replaces "hyperplane" by "point" in the hypotheses of the lemma, then it follows immediately as the points of $\mathcal{S}$ are distinct. The proof will proceed by induction going from hyperplanes to points. Assume $(k+1)$-planes M that have the property that they contain at least $k+2$ points of $\tilde{\mathcal{S}}$ also have the property that $\#(\mathcal{S}\cap M)\geq\#(\tilde{\mathcal{S}}\cap M)$ and we will show the same holds for k-planes. Fix a k-plane L containing $\mu\geq k+1$ points of $\tilde{\mathcal{S}}$, and let $\{M_\alpha\}$ denote the set of $k+1$ planes containing L and at least $\mu+1$ elements of $\tilde{\mathcal{S}}$. We have

$$\#(\tilde{\mathcal{S}}\cap L)+\sum_\alpha\#(\tilde{\mathcal{S}}\cap(M_\alpha\backslash L))=R$$

$$\#(\mathcal{S}\cap L)+\sum_\alpha\#(\mathcal{S}\cap(M_\alpha\backslash L))\leq R$$

the first line because every point of $\tilde{\mathcal{S}}$ not in L is in exactly one M_α and the second because every point of $\mathcal{S}$ not in L is in at most one M_α. Rewrite these as

$$(\#M_\alpha-1)\#(\tilde{\mathcal{S}}\cap L)-\sum_\alpha\#(\tilde{\mathcal{S}}\cap M_\alpha)=-R$$

$$(\#M_\alpha-1)\#(\mathcal{S}\cap L)-\sum_\alpha\#(\mathcal{S}\cap M_\alpha)\geq -R$$

But by our induction hypothesis $\sum_\alpha\#(\mathcal{S}\cap M_\alpha)\geq\#(\tilde{\mathcal{S}}\cap M_\alpha)$ so putting the two lines together, we obtain the result for L. □

Proof of theorem. Given decompositions $\phi = \sum_{j=1}^r u_j \otimes v_j \otimes w_j, \tilde{\phi} = \sum_{j=1}^r \tilde{u}_j \otimes \tilde{v}_j \otimes \tilde{w}_j$ of length r we want to show they are essentially the same. (Note that if there were a decomposition $\tilde{\phi}$ of length e.g., $r-1$, we could construct from it a decomposition of length r by replacing $\tilde{u}_1 \otimes \tilde{v}_1 \otimes \tilde{w}_1$ by $\frac{1}{2}\tilde{u}_1 \otimes \tilde{v}_1 \otimes \tilde{w}_1 + \frac{1}{2}\tilde{u}_1 \otimes \tilde{v}_1 \otimes \tilde{w}_1$, so uniqueness of the length r decomposition implies the rank is r.) We first show $\mathcal{S}_A = \tilde{\mathcal{S}}_A, \mathcal{S}_B = \tilde{\mathcal{S}}_B, \mathcal{S}_C = \tilde{\mathcal{S}}_C$. By symmetry it is sufficient to prove the last statement. By the permutation lemma it is sufficient to show that if $H \subset \mathbb{P}C$ is a hyperplane such that $\#(\tilde{\mathcal{S}}_C \cap H) \geq \mathbf{c} - 1$ then $\#(\mathcal{S}_C \cap H) \geq \#(\tilde{\mathcal{S}}_C \cap H)$ because we already know $k_{\mathcal{S}_C} \geq 2$.

Recall the classical fact about matrices (due to Sylvester): if $M \in A \otimes B$ and $U \subset A$, $V \subset B$, then

$$\mathrm{rank}(M) \geq \mathrm{rank}(M|_{U^\perp \times B^*}) + \mathrm{rank}(M|_{A^* \times V^\perp}) - \mathrm{rank}(M|_{U^\perp \times V^\perp}).$$

Let $A_H := \langle u_j \mid [w_j] \notin H \rangle$, $B_H := \langle v_j \mid [w_j] \notin H \rangle$

$$\begin{aligned}
\#(\tilde{\mathcal{S}}_c \not\subset H) &\geq \mathrm{rank}(T(H^\perp)) \\
&\geq \mathrm{rank}(T(H^\perp)|_{A_H{}^\perp \times B^*}) + \mathrm{rank}(T(H^\perp)|_{A^* \times B_H{}^\perp}) - \mathrm{rank}(T(H^\perp)|_{A_H{}^\perp \times B_H{}^\perp}) \\
&\geq \min(k_A, \#(\mathcal{S}_C \not\subset H)) + \min(k_B, \#(\mathcal{S}_C \not\subset H)) - \#(\mathcal{S}_C \not\subset H)
\end{aligned}$$

where the last line follows by the definition of Kruskal rank. Finally we need to show that $\#(\mathcal{S}_C \not\subset H) \leq \min(k_A, k_B)$. But this follows because

$$r - \#(\mathcal{S}_C \not\subset H) = \#(\mathcal{S}_C \subset H) \geq \mathbf{c} - 1 \geq k_C - 1 \geq 2r - k_A - k_B + 1$$

i.e., $k_A + k_B - \#(\mathcal{S}_C \not\subset H) \geq r + 1$, which can only hold if $\#(\mathcal{S}_C \not\subset H) \leq \min(k_A, k_B)$.

Now that we have $\mathcal{S}_A = \tilde{\mathcal{S}}_A$ etc.. , say we have two expressions

$$\begin{aligned}
T &= u_1 \otimes v_1 \otimes w_1 + \cdots + u_r \otimes v_r \otimes w_r \\
T &= u_1 \otimes v_{\sigma(1)} \otimes w_{\tau(1)} + \cdots + u_r \otimes v_{\sigma(r)} \otimes w_{\tau(r)}
\end{aligned}$$

for some $\sigma, \tau \in \mathfrak{S}_r$. First observe that if $\sigma = \tau$ then we are reduced to the two factor case which is easy, i.e., if $T \in A \otimes B$ of rank r has expressions $T = a_1 \otimes b_1 + \cdots + a_r \otimes b_r$ and $T = a_1 \otimes b_{\sigma(1)} + \cdots + a_r \otimes b_{\sigma(r)}$, then it is easy to see that $\sigma = Id$.

So assume $\sigma \neq \tau$, then there exists a smallest $j_0 \in \{1, ..., r\}$ such that $\sigma(j_0) =: s_0 \neq t_0 := \tau(j_0)$. We claim there exist subsets $S, T \subset \{1, ..., r\}$ with the properties

- $s_0 \in S$, $t_0 \in T$,
- $S \cap T = \emptyset$,
- $\#(S) \leq r - k_{\mathcal{S}_B} + 1$, $\#(T) \leq r - k_{\mathcal{S}_C} + 1$ and
- $\langle v_j \mid j \in S^c \rangle =: H_S \subset B$, $\langle w_j \mid j \in T^c \rangle =: H_T \subset C$ are hyperplanes.

Here $S^c = \{1, ..., r\} \backslash S$.

To prove the claim take a hyperplane $H_T \subset C$ containing w_{s_0} but not containing w_{t_0}, and let T^v be the set of indices of the w_j contained in H_T, so in particular $\#(T^c) \geq k_{\mathcal{S}_C} - 1$ insuring the cardinality bound for T. Now consider the linear space $\langle v_t \mid t \in T \rangle \subset B$. Since $\#(T) \leq r - k_{\mathcal{S}_C} + 1 \leq k_{\mathcal{S}_B} - 1$ (the last inequality because $k_{\mathcal{S}_A} \leq r$), adding any vector of $\mathcal{S}_B$ to $\langle v_t \mid t \in T \rangle$ would increase its dimension, in particular, $v_{s_0} \notin \langle v_t \mid t \in T \rangle$. Thus there exists a hyperplane $H_S \subset B$ containing $\langle v_t \mid t \in T \rangle$ and not containing v_{s_0}. Let S be the set of indices of the v_j contained in H_S. Then S, T have the desired properties.

Now by construction $T|_{H_S{}^\perp \times H_T{}^\perp} = 0$, which implies there is a nontrivial linear relation among the u_j for the j appearing in $S \cap T$, but this number is at most $\min(r - k_{\mathcal{S}_B} + 1, r - k_{\mathcal{S}_C} + 1)$ which is less than $k_{\mathcal{S}_A}$. □

Remark 5. There were several inequalities used in the proof that were far from sharp. In fact, Kruskal proves versions of his theorem with weaker hypotheses designed to be more efficient regarding the use of the inequalities.

Remark 6. The proof above is essentially Kruskal's. The reduction from a 16 page proof to the 2 page proof above is mostly due to writing statements invariantly rather than in coordinates.

More generally, Kruskal shows that for d factors, if $\sum_{i=1}^{d} \mathcal{S}_{k_i} \geq 2r + d - 1$ then uniqueness holds.

Acknowledgments. This short note is an outgrowth of the AIM workshop *Geometry and representation theory of tensors for computer science, statistics and other areas* July 21-25, 2008, and the author gratefully thanks AIM and the other participants of the workshop, in particular L. De Lathauwer and P. Comon who encouraged the writeup. It will appear in the forthcoming book *Geometry of Tensors: Applications to complexity, statistics and engineering* with J. Morton.

References

1. Joseph B. Kruskal, *Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra and Appl. **18** (1977), no. 2, 95–138. MR MR0444690 (56 #3040)

E-mail address: jml@math.tamu.edu