International Journal of Mathematics and Statistics Studies
Vol.10, No.5, pp.1-14, 2022
Print ISSN: 2053-2229 (Print),
Online ISSN: 2053-2210 (Online)

A Mathematical Proof to the HKLam Theory by Linear/Tensor Algebra and Analysis

Carson, Lam Kai Shun
University of Hong Kong, Hong Kong

Citation: Carson, Lam Kai Shun (2022) A Mathematical Proof to the HKLam Theory by Linear/Tensor Algebra and Analysis, *International Journal of Mathematics and Statistics Studies*, Vol.10, No.5, pp.1-14

ABSTRACT: *In the previous papers, I have mentioned several times of HKLam Theory and their everyday usage but without the abstract mathematical proof. In order to remediate the flaws, I am now trying to proof the theory through both Tensor Algebra and Analysis as well as the statistical inference in this present paper. Indeed, people always say that mathematicians are linear animals or participate much in the subject of linear algebra while the British Scientist Newton observed a falling apple and discovered the gravity together with the development of calculus. In a similar case, my proof in the part of tensor algebra will be an analogy to the linear mapping, transformation etc while there are the corresponding corollary real physical life cases – 2 to 3 dimensional vectors calculus or even higher dimension of tensor analysis. Indeed, my proof will be based on the order two tensor but the HKLam theory may be extended up to nth order tensor but NOT applicable to the topic of the planned politics or even economics etc.The main aim is to show the proof of HKLam Theory by linear/Tensor algebra together with some applications in fluid dynamic and stress tensor field etc.*

KEYWORDS: mathematical proof, HKLam Theory, linear/tensor algebra , analysis

PROOF OF THE HKLAM THEORY

1.1 Linear Mapping (Transformation)

Theorem 1.1 [1] Let V and W be vector spaces. Let $\{\underline{\mathbf{v}}_1,\ldots,\underline{\mathbf{v}}_n\}$ be a basis of V, and let $\underline{\mathbf{w}}_1,\ldots,\underline{\mathbf{w}}_n$ be any arbitrary elements of W. Then there exists a unique linear mapping T: V → W such that

$$T(\underline{\mathbf{v}}_1) = \underline{\mathbf{w}}_1,\ldots,T(\underline{\mathbf{v}}_n) = \underline{\mathbf{w}}_n.$$

If furthermore, $\alpha_1,\ldots, \alpha_n$ are scalars, then

$$T(\alpha_1 \underline{\mathbf{v}}_1+ \ldots + \alpha_n \underline{\mathbf{v}}_n) = \alpha_1 \underline{\mathbf{w}}_1 + \ldots + \alpha_n \underline{\mathbf{w}}_n.$$

Proof:

Part A. Existence of the (linear) map – "T" (Assume the linearity first as I will prove immediately in the following (part B)).

Let V and W be vector spaces. Let $\{\underline{\mathbf{v}}_1,\ldots,\underline{\mathbf{v}}_n\}$ be a basis of V, and let $\{\underline{\mathbf{w'}}_1,\ldots,\underline{\mathbf{w'}}_n\}$ be the basis of W. Let T:V → W be a linear map that

$$T(\underline{\mathbf{v}}_1) = \underline{\mathbf{w'}}_1,\ T(\underline{\mathbf{v}}_2) = \underline{\mathbf{w'}}_2,\ \ldots\ ,\ T(\underline{\mathbf{v}}_n) = \underline{\mathbf{w'}}_n$$

Then $\alpha_1\,\underline{\mathbf{v}}_1 + \ldots + \alpha_n\,\underline{\mathbf{v}}_n = \underline{\mathbf{0}}$ iff $\alpha_i = 0$ for i = 1,2, …, n

$$\begin{aligned} T(\alpha_1\,\underline{\mathbf{v}}_1 + \ldots + \alpha_n\,\underline{\mathbf{v}}_n) = T(\underline{\mathbf{0}}) \\ &= \alpha_1 T(\underline{\mathbf{v}}_1) + \ldots + \alpha_n T(\underline{\mathbf{v}}_n) \\ &= 0T(\underline{\mathbf{v}}_1) + \ldots + 0T(\underline{\mathbf{v}}_n) \\ &= 0\underline{\mathbf{w'}}_1 + \ldots + 0\underline{\mathbf{w'}}_n \\ &= \underline{\mathbf{0}} \end{aligned}$$

(N.B. [2] 1. For if $\{\underline{\mathbf{w'}}_1, \ldots , \underline{\mathbf{w'}}_n\}$ are linearly dependent, then by definition, one may find scalars $\lambda_1 ,\ldots, \lambda_i$ that is not equal to zero such that

$$\lambda_1\underline{\mathbf{w'}}_1 + \ldots + \lambda_i\underline{\mathbf{w'}}_i + \lambda_{i+1}\underline{\mathbf{w'}}_{i+1}\ldots + \lambda_n\underline{\mathbf{w'}}_n = \underline{\mathbf{0}}$$

Then $T(\lambda_1\underline{\mathbf{v}}_1 + \ldots + \lambda_i\underline{\mathbf{v}}_i + \lambda_{i+1}\underline{\mathbf{v}}_{i+1}+ \lambda_n\underline{\mathbf{v}}_n) = \underline{\mathbf{0}}$ for $\lambda_1 ,\ldots, \lambda_i$ not equal to zero and otherwise equals zero. Hence, we have non-zero elements in the kernel, says $\underline{\mathbf{v}} = \lambda_1\underline{\mathbf{v}}_1 + \ldots + \lambda_i\underline{\mathbf{v}}_i$ not equals to zero as $\lambda_1 ,\ldots, \lambda_i$ not equal to zero.

(But $\{\underline{\mathbf{v}}_1, \ldots, \underline{\mathbf{v}}_n\}$ is the base of V or if

$$\lambda_1\,\underline{\mathbf{v}}_1 + \ldots + \lambda_n\,\underline{\mathbf{v}}_n = \underline{\mathbf{0}}$$

then all λ_i must be zero by the property that all $\underline{\mathbf{v}}_i$ s are linear independent for i = 1,2, …, n which obviously induce a contradiction with previous assumption $\lambda_1 ,\ldots, \lambda_i$ not equal to zero). Thus, if kernel of T is zero, then the image vector $\{T(\underline{\mathbf{v}}_1), \ldots , T(\underline{\mathbf{v}}_n)\}$ can form a basis.

Conversely, if $\{T(\underline{\mathbf{v}}_1), \ldots , T(\underline{\mathbf{v}}_n)\}$ forms the basis of W, then $T(\underline{\mathbf{x}}) = \underline{\mathbf{0}}$ where $\underline{\mathbf{x}} = (x_1, \ldots, x_n)$ belongs to a vector space V and $\underline{\mathbf{x}} = x_1\,\underline{\mathbf{v}}_1 + \ldots + x_n\,\underline{\mathbf{v}}_n$ with $\{\underline{\mathbf{v}}_1, \ldots , \underline{\mathbf{v}}_n\}$ as the basis of V

$$x_1T(\underline{\mathbf{v}}_1) + \ldots + x_nT(\underline{\mathbf{v}}_n) = \underline{\mathbf{0}}$$

$$T(x_1\,\underline{\mathbf{v}}_1) + \ldots + T(x_n\,\underline{\mathbf{v}}_n) = \underline{\mathbf{0}}$$

$$T(x_1 \underline{\mathbf{v}}_1 + \ldots + x_n \underline{\mathbf{v}}_n) = \underline{\mathbf{0}}$$

$$T(\underline{\mathbf{x}}) = \underline{\mathbf{0}}.$$

Hence, we may have the following corollary:

Corollary 1.2 [3] The image vectors $\{T(\mathbf{v}_1), \ldots, T(\mathbf{v}_n)\}$ can form a basis of V (i.e. dim(V) = dim(W) from the fact that dim V = dim Ker T + dim Im T and Ker T = $\{\underline{\mathbf{0}}\}$) if only if the kernel of T is zero. (N.B. Detail proofs of the Kernel and Image of a Linear Map is out of the scope of the present paper and has been described on [1] p.59 – 63.)

Part B [1]: Linearity of the map – "T"

Suppose $\underline{\mathbf{v}}$ be an element of V and also $\alpha_1, \ldots, \alpha_n$ be the unique scalars such that

$$\underline{\mathbf{v}} = \alpha_1 \underline{\mathbf{v}}_1 + \ldots + \alpha_n \underline{\mathbf{v}}_n$$

Suppose further that,

$$T(\underline{\mathbf{v}}) = \alpha_1 \underline{\mathbf{w}}_1 + \ldots + \alpha_n \underline{\mathbf{w}}_n$$

and one may have additional element $\underline{\mathbf{v}'} = \gamma_1 \underline{\mathbf{v}}_1 + \ldots + \gamma_n \underline{\mathbf{v}}_n$, then

$$\underline{\mathbf{v}} + \underline{\mathbf{v}'} = (\alpha_1 + \gamma_1) \underline{\mathbf{v}}_1 + \ldots + (\alpha_n + \gamma_n) \underline{\mathbf{v}}_n.$$

By definition,

$$\begin{aligned} T(\underline{\mathbf{v}} + \underline{\mathbf{v}'}) &= (\alpha_1 + \gamma_1) \underline{\mathbf{w}}_1 + \ldots + (\alpha_n + \gamma_n) \underline{\mathbf{w}}_n \\ &= \alpha_1 \underline{\mathbf{w}}_1 + \gamma_1 \underline{\mathbf{w}}_1 + \ldots + \alpha_n \underline{\mathbf{w}}_n + \gamma_n \underline{\mathbf{w}}_n \\ &= (\alpha_1 \underline{\mathbf{w}}_1 + \ldots + \alpha_n \underline{\mathbf{w}}_n) + (\gamma_1 \underline{\mathbf{w}}_1 + \ldots + \gamma_n \underline{\mathbf{w}}_n) \\ &= T(\underline{\mathbf{v}}) + T(\underline{\mathbf{v}}') \end{aligned}$$

Let c be a scalar, then $c\underline{\mathbf{v}} = c\,(\alpha_1 \underline{\mathbf{v}}_1 + \ldots + \alpha_n \underline{\mathbf{v}}_n)$

Consider

$$\begin{aligned} T(cv) &= c\,\alpha_1 \underline{\mathbf{w}}_1 + \ldots + c\,\alpha_n \underline{\mathbf{w}}_n \\ &= c\,(\alpha_1 \underline{\mathbf{w}}_1 + \ldots + \alpha_n \underline{\mathbf{w}}_n) \\ &= c\,T(\underline{\mathbf{v}}) \end{aligned}$$

Hence, the map T is linear.

@ECRTD-UK: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

Part C [5]: Uniqueness of the linear map – "T"

Let T: V → W be a linear map which satisfies $T(\underline{\mathbf{v}}_j) = \underline{\mathbf{w}}_j$ for all j belongs to {1, … , n}. At the same time, suppose there is also another linear map S: V → W such that $S(\underline{\mathbf{v}}_j) = \underline{\mathbf{w}}_j$ for all j belongs to {1, …, n}. It suffixes to show that the linear maps T = S. Otherwise, the uniqueness may turn out to be the problem in change of basis [4].

Let $\underline{\mathbf{x}} = c_1\underline{\mathbf{v}}_1 + \ldots + c_n\underline{\mathbf{v}}_n$, then

$$
\begin{aligned}
T(\underline{\mathbf{x}}) &= T(c_1\underline{\mathbf{v}}_1 + \ldots + c_n\underline{\mathbf{v}}_n) \\
&= c_1T(\underline{\mathbf{v}}_1) + \ldots + c_nT(\underline{\mathbf{v}}_n) \\
&= c_1\underline{\mathbf{w}}_1 + \ldots + c_n\underline{\mathbf{w}}_n \\
&= c_1S(\underline{\mathbf{v}}_1) + \ldots + c_nS(\underline{\mathbf{v}}_n) \\
&= S(c_1\underline{\mathbf{v}}_1 + \ldots + c_n\underline{\mathbf{v}}_n) \\
&= S(\underline{\mathbf{x}})
\end{aligned}
$$

Thus, this author has shown that for every $\underline{\mathbf{x}}$ belongs to V, $T(\underline{\mathbf{x}}) = S(\underline{\mathbf{x}})$. Hence, T=S proving uniqueness of the original T that I have already established.

1.2 Linear Combination & Regression

Definition 1.2.1: Let V be an arbitary vector space, and let $\underline{\mathbf{v}}_1, \ldots, \underline{\mathbf{v}}_n$ be elements of V. Let $x_1, \ldots, x_n$ be scalars. Any vector $\underline{w} \in V$ that can be expressed of the type

$$\underline{\mathbf{w}} = x_1\,\underline{\mathbf{v}}_1 + \ldots + x_n\underline{\mathbf{v}}_n$$

is called a linear combination of vectors $\{\underline{\mathbf{v}}_1, \ldots, \underline{\mathbf{v}}_n\}$.

Now, consider a vector equation like the following:

$$x_1\,\underline{\mathbf{v}}_1 + \ldots + x_n\,\underline{\mathbf{v}}_k = \underline{\mathbf{b}}$$

where $\underline{\mathbf{v}}_1, \ldots, \underline{\mathbf{v}}_k, \underline{\mathbf{b}}$ are vectors in $\mathbb{R}^n$ and $x_1, x_2, \ldots, x_k$ are unknown scalars and has the same solution set as the linear system with augmented matrix

$$\left[\begin{array}{cccc|c} l & l & \dots & l & l \\ v_1 & v_2 & \dots & v_k & b \\ l & l & \dots & l & l \end{array}\right]$$

whose columns are the $\underline{\mathbf{v}}_i$'s and the $\underline{\mathbf{b}}$'s.

Corollary 1.2.2: For the vector **b** in $\mathbb{R}^n$, we may always express it as the linear combination of the basis $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \underline{\mathbf{e}}_n\}$through vector equation where $\underline{\mathbf{e}}_i = (0, \dots,1, \dots 0)$ with the i-th coordinate equals to 1.
Proof: It follows directly from the fact that vector $\underline{\mathbf{b}} = (b_1, \dots , b_n)$ for some scalars b_i

where i = 1,2,… ,n

Then $\underline{\mathbf{b}} = b_1\, \underline{\mathbf{e}}_1 + \dots + b_n\, \underline{\mathbf{e}}_n$.

Corollary 1.2.3 [6]: In particular, the linear regression may be considered as the approximation of the vector $\bar{b} \in W$ through a suitable selection of combination coefficients or projection in terms of vector equation.

Proof: By definition, $\underline{\mathbf{b}} = b_1\, \underline{\mathbf{y}}_1 + \dots + b_k\underline{\mathbf{y}}_k$ or $\underline{\mathbf{b}} = (b_1, b_2, \dots , b_k)\begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_k \end{pmatrix}$, where $\mathbf{y}_i = \begin{pmatrix} 0 \\ 0 \\ . \\ y_i \\ . \\ 0 \end{pmatrix}$ is the i-th vector with only the i-th coordinate not equals to zero for some basis $\{\underline{\mathbf{y}}_1, \underline{\mathbf{y}}_2,\dots,\underline{\mathbf{y}}_i,\dots,\underline{\mathbf{y}}_n\}$, hence by considering linear regression as a special case of linear combination, one may always approximate $\underline{\mathbf{b}}$ by the linear function like $\underline{\mathbf{y}} = \underline{\mathbf{m}}^T \underline{\mathbf{x}} + \underline{\mathbf{c}}$ or for multiple variables, one may have

$\underline{\mathbf{b}} = (b'_1, b'_2, \dots , b'_k)\begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_k \end{pmatrix} + \begin{pmatrix} c_1 \\ c_2 \\ . \\ . \\ . \\ c_k \end{pmatrix}$ for other approximated scalars b'i (N.B. one may need to choose these approximated scalars b'_i indeed) not equals to the original scalars b_i where I = 1, 2, … , k in the above $\underline{\mathbf{b}} = b'_1\underline{\mathbf{y}}_1 + b'_2\underline{\mathbf{y}}_2 + \dots + b'_k\underline{\mathbf{y}}_k$ with error terms $(c_1, c_2, \dots , c_k)$ is used to adjust for the fitting of the approximated one into the original b vector's expression. i.e.

$$\underline{\mathbf{b}} = b'_1\underline{\mathbf{y}}_1 + b'_2\underline{\mathbf{y}}_2 + \dots + b'_k\underline{\mathbf{y}}_k + c_1 + c_2 + \dots + c_k.$$

Or in general, for any resulted vector $\overline{w} \in W$ which is obtained from the space V through the linear mapping "T" to the space like W, we may have:

$\underline{\mathbf{w}} = \beta_0 + \beta_1\underline{\mathbf{x}}_1 + \ldots + \beta_n\underline{\mathbf{x}}_n + \underline{\boldsymbol{\varepsilon}}$ for some basis $\{\underline{\mathbf{x}}_1, \underline{\mathbf{x}}_2, \ldots, \underline{\mathbf{x}}_n\}$ in W and a vector $\underline{\boldsymbol{\varepsilon}}$ in W with some scalars β_i where i = 1,2, … ,n.

Indeed, one of the most famous case is the selection of dot product as the b'$_i$ s in the orthogonal base with positive definite scalar product and more general case. With reference to Gram-Schmidt orthogonalization process [1], p.103, p.123, for any given arbitrary basis $\{\underline{\mathbf{y}}_1, \underline{\mathbf{y}}_2,\ldots,\underline{\mathbf{y}}_k\}$ of V, one may always select $\frac{\langle b_{(i-1)}\underline{y}_{(i-1)}^T, b_i\underline{y}_i\rangle}{\langle b_i\underline{y}_i^T, b_i\underline{y}_i\rangle}$ as our b'$_i$ s such that

$$\underline{\mathbf{b}} = \underline{\mathbf{y}}_1 + \frac{\langle b_1\underline{y}_1^T, b_2\underline{y}_2\rangle}{\langle b_2\underline{y}_2^T, b_2\underline{y}_2\rangle}\underline{\mathbf{y}}_2 + \frac{\langle b_2\underline{y}_2^T, b_3\underline{y}_3\rangle}{\langle b_3^T\underline{y}_3, b_3\underline{y}_3\rangle}\underline{\mathbf{y}}_3 + \ldots + \frac{\langle b_{(k-1)}\underline{y}_{(k-1)}^T, b_k\underline{y}_k\rangle}{\langle b_k^T\underline{y}_k, b_k\underline{y}_k\rangle}\underline{\mathbf{y}}_k + \text{Error terms ---------- } (\ *\ )$$

where $<b_{(i-1)}\underline{\mathbf{y}}_{(i-1)}{}^T, b_i\underline{\mathbf{y}}_i>$ denotes the dot product between vectors $b_{(i-1)}\underline{\mathbf{y}}_{(i-1)}$ and $b_i\underline{\mathbf{y}}_i$ for I = 1,2,…,k and error terms will be defined as:

$$\mathbf{c}_i = b_i\,\underline{\mathbf{y}}_i - \frac{\langle b_{(i-1)}\underline{y}_{(i-1)}^T, b_i\underline{y}_i\rangle}{\langle b_i\underline{y}_i^T, b_i\underline{y}_i\rangle}\underline{\mathbf{y}}_i$$

The above result is defined as my dot product (of the orthogonal projection) regression. Certainly, one may develop some other kind of regression by using the ideas of geometric mean and ordinary least square like the prescribed dot product one etc. This author remarks that one may go a further step by using the recursive Gram-Schmidt orthogonalization process to obtain the orthogonal basis and perform another type of projection etc.

In Brief,

1. given known values of b_i s, one may find the corresponding b'_i s by applying a suitable projection method;

2. given values of b'_i s and a known method of projection, one may find the respective b_i s with the vice versa process of applying the projection in order to reconstruct those b_i s.

(N.B. Actually, you may consider the corollary 1.2.3 as a kind of projection etc as b'_i s are selected by the users which technically forms a geometric projection. i.e. projects part of the vector $\underline{\mathbf{b}}$'s suitable values with b'_i s or takes a proportion of the original values of b_i s. Conceptually, this is just like the famous orthogonal projection [7]. Practically, we may use the set.seed command in JASP and R or machine learning in guessing the level/degree of b'_i s etc required. They constitute a kind of philosophy. However, the computer stimulation of the vector projection or any computational tasks etc are NOT the focus of the present paper. The main aim of this paper has been described in the abstract.)

1.3 Matrix (or Array) as a Linear Mapping

Definition 1.3.1 [8] Let V' and W' be two vector spaces. Let $\{\underline{\mathbf{v}}_1, \underline{\mathbf{v}}_2, \ldots, \underline{\mathbf{v}}_n\}$ be a basis for V' and $\{\underline{\mathbf{w}}_1, \underline{\mathbf{w}}_2, \ldots, \underline{\mathbf{w}}_m\}$ be a basis for W'. For any $\underline{v} \in V'$ and any $\underline{w} \in W'$, denote by $[\underline{\mathbf{v}}]_{V'}$ and $[\underline{\mathbf{w}}]_{W'}$ their mX1 and nX1 coordinate vectors with respect to the two bases of V' and W' respectively. Let f: V' → W' be a linear map. An mxn (i.e. m rows and n columns) matrix $[f]_{V'W'}$ such that, for any $\underline{v} \in V'$,

$$[f(\mathrm{v})]_{W'} = [f]_{V'W'}[\underline{\mathbf{v}}]_{V'}$$

is called a matrix of the linear map f with respect to the bases V' and W'.

Moreover, the resulted $[f(\mathrm{v})]_{W'}$ is a vector of rank m.

Theorem 1.3.1.2 A map f is linear if only if it has a matrix A and transform coordinates.

"If" part: Assume that the mapping is linear, then according to [1], p.83 there exists a unique matrix A such that $[f]_{V'W'} = [f]_A$;

Print ISSN: 2053-2229 (Print),
Online ISSN: 2053-2210 (Online)

"Only if" part: Assume there is a mxn matrix A, then we may construct a map *f*: V' → W' with the basis of V' = n independent column vectors and the basis of W' = m independent row vectors such that $[f]_{V'W'} = [f]_A$. Then we may verify the map is linear like the case in the previous section. Obviously, the map also transforms coordinates.

Corollary 1.3.1.3 For any matrix A, according to Theorem 1.3.1.2, there exists a unique linear map L such that $L(\underline{\mathbf{x}}) = A\underline{\mathbf{x}}$ and results a coordinate transformed vector $\underline{\mathbf{b}}$ in the matrix equation form:

$$A\underline{\mathbf{x}} = \underline{\mathbf{b}}$$

By the theorems in section 1.1, the corollary 1.2.3 and the proof in Lam (March, 2020), one may have the following theory:

Theorem 1.3.2 (HKLam Theorem) For any matrix equation $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$, one may express it as a linear transformation such that T: V → W. Then one may approximate (or project) the resulted vector $\underline{b} \in W$ by selecting a suitable linear combination coefficients with some error terms. Furthermore, by Lam (29 March, 2020) [9], one may also expand the projection or approximation by an expression with a series of recursive linear regression substitution.

Proof: Follows directly from the

1. Theorem 1.1 – existence, linearity and uniqueness of linear transformation,

2. Corollary 1.2.3 – linear regression as an approximation/projection of the linear combination of the domain in linear map),

3. Definition 1.3.1 – matrix equation as a kind of linear transformation;

4. Theorem 1.3.1.2 – A map is linear iff it has a matrix A and transform coordinates;

5. Corollary 1.3.1.3 – Given any matrix A, one may construct its linear map together with its its matrix equation and results its transformed coordinate vector $\underline{\mathbf{b}}$, $\underline{\mathbf{b}}$ is then expressed as the linear combination or regression etc;

@ECRTD-UK: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

6. Lam (March, 2020) – the approximated regression of the domain in the matrix linear transformation can then be expressed as a series of recursive regression substitution.

N.B. The theorem may be proved by contrapositive: Assume that A**x** = **b** with $\underline{b} \neq b'_1\underline{y_1}+\ldots+b'_n\underline{y_n} + \underline{c_1}+\ldots+\underline{c_n}$. But by the assumption $b'_1\underline{y_1}+\ldots+b'_n\underline{y_n} + \underline{c_1}+\ldots+\underline{c_n} = b_1\underline{y_1}+\ldots+b_n\underline{y_n}$. This implies $\underline{b} \neq b_1\underline{y_1}+\ldots+b_n\underline{y_n}$ which obviously contradicts to the fact that A**x** = **b** = $b_1\mathbf{y}_1 + \ldots + b_n\mathbf{y}_n$!!!!!! Hence, one may always approximate or project vector "**b**" by the vector equation: $b'_1\mathbf{y}_1 + b'_2\mathbf{y}_2 + \ldots + b'_n\mathbf{y}_n + \mathbf{c}_1 + \ldots + \mathbf{c}_n$ with $\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ as the error terms or the linear regression to the combination. Suppose further that A**x** = **b**, then in matrix form, one may have [1]:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ . \\ . \\ . \\ . \\ b_m \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{j=1}^{n} a_{1j}x_j \\ . \\ . \\ . \\ . \\ \sum_{j=1}^{n} a_{mj}x_j \end{pmatrix} \quad \text{where } b_i = \sum_{j=1}^{n} a_{ij}\, x_j \text{ for I} = 1,2,\ldots,m$$

But **b** = $b_1\mathbf{y}_1 + b_2\mathbf{y}_2 + \ldots + b_m\mathbf{y}_m = (\sum_{j=1}^{n} a_{1j}\, x_j\,)\,\mathbf{y}_1 + (\sum_{j=1}^{n} a_{2j}\, x_j)\mathbf{y}_2 + \ldots + (\sum_{j=1}^{n} a_{mj}\, x_j)\mathbf{y}_m$ for some basis $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, y_m\}$ spans **b**. In other words,

$$\mathbf{b} = b_1\mathbf{y}_1 + b_2\mathbf{y}_2 + \ldots + b_m\mathbf{y}_m = (\sum_{j=1}^{n} a_{1j}\, x_j\,)\,\mathbf{y}_1 + (\sum_{j=1}^{n} a_{2j}\, x_j)\mathbf{y}_2 + \ldots + (\sum_{j=1}^{n} a_{mj}\, x_j)\mathbf{y}_m$$

$$= [\,(\sum_{j=1}^{n} a'_{1j}\, x_j)\;\mathbf{y}_1 + \mathbf{c}_1] + [(\sum_{j=1}^{n} a'_{2j}\, x_j)\mathbf{y}_2 + \mathbf{c}_2] + \ldots + [(\sum_{j=1}^{n} a'_{mj}\, x_j)\mathbf{y}_m + \mathbf{c}_m]$$

$$= b'_1\mathbf{y}_1 + b'_2\mathbf{y}_2 + \ldots + b'_m\mathbf{y}_m + \mathbf{c}_1 + \mathbf{c}_2 + \ldots + \mathbf{c}_m$$

= **b**' + error term where $(\sum_{j=1}^{n} a_{ij}\, x_j)\underline{\mathbf{y}}_i = (\sum_{j=1}^{n} a'_{ij}\, x_j)\underline{\mathbf{y}}_i + \underline{\mathbf{c}}_i$ for I = 1, …, m

Now, consider the approximated vector **b**', one may get:

$$\mathbf{b}' = (a'_{11}, a'_{12}, \ldots, a'_{1n})\begin{pmatrix} y_1 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix}\mathbf{x} + \ldots + (a'_{m1}, a'_{22}, \ldots, a'_{mn})\begin{pmatrix} 0 \\ 0 \\ . \\ . \\ . \\ y_m \end{pmatrix}\mathbf{x}$$

$$= [\ (a'_{11}, a'_{12}, \ldots, a'_{1n})\begin{pmatrix} y_1 \\ 0 \\ . \\ . \\ . \\ 0 \end{pmatrix} + \ldots + (a'_{m1}, a'_{22}, \ldots, a'_{mn})\begin{pmatrix} 0 \\ 0 \\ . \\ . \\ . \\ y_m \end{pmatrix}\ ]\ \underline{\mathbf{x}}$$

$$=[\begin{bmatrix} a'_{11}y_1 & a'_{12}y_1 & \ldots & a'_{1n}y_1 \\ . & . & \ldots & . \\ . & . & \ldots & . \\ . & . & \ldots & . \\ a'_{m1}y_m & a'_{m2}y_m & \ldots & a'_{mn}y_m \end{bmatrix}]\underline{\mathbf{x}}$$

Now consider the transpose of $\begin{bmatrix} a'_{11}y_1 & a'_{12}y_1 & \ldots & a'_{1n}y_1 \\ . & . & \ldots & . \\ . & . & \ldots & . \\ . & . & \ldots & . \\ a'_{m1}y_m & a'_{m2}y_m & \ldots & a'_{mn}y_m \end{bmatrix}$, i.e.

$$\begin{bmatrix} a'_{11} & a'_{21} & \ldots & a'_{m1} \\ . & . & \ldots & . \\ . & . & \ldots & . \\ . & . & \ldots & . \\ a'_{1n} & a'_{2n} & \ldots & a'_{mn} \end{bmatrix}\begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_m \end{pmatrix}\underline{\mathbf{x}}$$

$\underline{\mathbf{b}'^{T}} = (A'^{T}\underline{\mathbf{y}})\ \underline{\mathbf{x}} = W^{T}\underline{\mathbf{x}}$ where $W^{T} = (A'^{T}\underline{\mathbf{y}})$

Now, by adding the error vector $\underline{\mathbf{C}}$ to the above vector equation, one may further get:

$\underline{\mathbf{b}}'$ + error term = $(W^{T}\underline{\mathbf{x}})^{T} + \underline{\mathbf{C}} = W\underline{\mathbf{x}}^{T} + \underline{\mathbf{C}}$.

In terms of dot product projection (or equation (*)), we may get:

$$b_i' = \frac{\left\langle \sum_{i=1}^{n} a_{(i-1)j} x_j \underline{y_{i-1}}^T, \sum_{i=1}^{n} a_{ij} x_j \underline{y_i} \right\rangle}{\left\langle \sum_{i=1}^{n} a_{ij} x_j \underline{y_i^T}, \sum_{i=1}^{n} a_{ij} x_j \underline{y_i} \right\rangle}$$

Or one will get the wanted result that can connect the statistical model recursive vector-matrix equation as mentioned early in my paper named "Evaluation of the Weather-Influenza Pattern with a Regression Model Approximation to Casuality" etc. This completes the proof in my HKLam Theory or in general the so-called statistical modeling theory. From the above proof of linear regression is just a proportion of linear combination one may find that the converse is also true. This is because if one get a regression vector equation such as $\underline{\mathbf{b'}} = W^T \underline{\mathbf{x}} + \underline{\mathbf{c}}$ and a set of data, say (x_1,y_1), (x_2,y_2), … , (x_n,y_n), together with the error term, one may get back the vector $\underline{\mathbf{b}}$. Then one may revese the previous proof procedure and finally get the proportional ratio between the coefficients of linear regression and combination or (a'_i / a_i)'s for i = 1,2, … ,n so as the matrix A etc. This writer will omit the reverse of the proof and leave it to those interested readers.

N.B. Indeed, one may further compute the values of the error term **C** by considering the monic polynomials as a basis for the linear regression (/ combination) to approximate the vectors $\mathbf{y}_i$s. Practically, one may have:

$$\underline{\mathbf{c}}_i = \left(\sum_{j=1}^{n} a_{ij}\, x_j - \sum_{j=1}^{n} a'_{ij}\, x_j \right) \underline{\mathbf{y}}_i$$

$$= \left(\sum_{j=1}^{n} \left(a_{ij} - a'_{ij}\right) x_j \right) \underline{\mathbf{y}}_i$$

$$= \left(\sum_{j=1}^{n} \left(a_{ij} - a'_{ij}\right) x_j \right) \left(\min \sum_{i=1}^{n} \left(y_i - \hat{f}(x_i) \right)^2 \right)$$

where $\hat{f}(x_i)$ is an approximation to y_i

If one selects the set of monic polynomial $\{1, x, x^2, \dots, x^{n-1}\}$ or $\{f(x_i)\}$ for I =1, …, n as the basis for spanning $\hat{f}(x_i)$, then one may get the following linear regression or combination:

$$f(x_i) = \sum_{i=1}^{n}(w_i f(x_i))$$

To be precise,

$$\underline{\mathbf{c}}_i = (\sum_{j=1}^{n}(a_{ij} - a'_{ij})\, x_j)\, (\, min \left(\sum_{i=1}^{n}\left(\underline{y_i} - \sum_{k=1}^{l} w_{ik} f_{ik}(x)\right)\right)^2)$$

and our goal is to minimize

$$\left(\sum_{i=1}^{n}\left(\underline{y_i} - \sum_{k=1}^{l} w_{ik} f_{ik}(x)\right)\right)^2$$

The method of finding the minimum point may be by using gradient descent etc just as mentioned in my paper "A Saddle Point Finding Method for Lorenz Attractor through Business Machine Learning Algorithm" etc.

N.B. It is true that there is a Johnson Lindenstrauss Lemma for the random matrix which may be used for high dimensional space mapping to the lower one, however, this is not the focus of the present section as its aim is to prove my HKLam theory. The lemma may be related to the distance between data during the analysis in the topic of machine learning. This author will leave such lemma and the usage for those interested readers in their further study. I should also note that there is a need considering the order of those data points or vectors in the daily usage of the HKLam theory. Indeed, two different types of result may be given.

N.B. As we may have the Moore-Penrose inverse of a given matrix A belongs to the $\mathbb{R}^{mxn}$,one may get the corresponding singular value decomposition (SVD) and hence get the pseudo-inverse. Once we have get the inverse, we may find the solution x to the matrix equation A$\underline{\mathbf{x}}$ = $\underline{\mathbf{b}}$ with the linear least-squares [12]:

where the modified $\underline{\mathbf{x}}$ = A^+ (Regression_Approximation) + $(I – A^+A)w$

and (Regression_Approximation) is well defined in corollary 1.2.3.

Theorem 1.3.3 The converse of the theorem 1.3.2 or HKLam Theory is also true.

Proof: Let $\beta_0 + \beta_1\underline{\mathbf{x}}_1 + \ldots + \beta_n\underline{\mathbf{x}}_n + \underline{\boldsymbol{\varepsilon}}$ be the regression approximation of the matrix linear transformation of $A\underline{\mathbf{x}} = \underline{\mathbf{b}}$. Then by the definition of inverse to a vector $\underline{\mathbf{x}}$, we may have [10]:

$$\underline{\mathbf{x}}^{-1} = \frac{1}{\|\underline{x}\|^2}\underline{\mathbf{x}} \qquad \text{(since the inner product of} < x\ x^{-1} > =1)$$

Hence, we may have $A' = (\beta_0 + \beta_1\underline{\mathbf{x}}_1 + \ldots + \beta_n\underline{\mathbf{x}}_n + \underline{\boldsymbol{\varepsilon}})\underline{\mathbf{x}}^{-1}$

$$= (\beta_0 + \beta_1\underline{\mathbf{x}}_1 + \ldots + \beta_n\underline{\mathbf{x}}_n + \underline{\boldsymbol{\varepsilon}})\left(\frac{1}{\|\underline{x}\|^2}\underline{\mathbf{x}}\right)$$

$$= (\underline{\mathbf{x}}_1 b'_1 + \underline{\mathbf{x}}_2 b'_2 + \ldots + \underline{\mathbf{x}}_n b'_n + \underline{\mathbf{c}}_1 + \underline{\mathbf{c}}_2 + \ldots + \underline{\mathbf{c}}_n)\ \left(\frac{1}{\|\underline{x}\|^2}\underline{\mathbf{x}}\right)$$

$$= (\underline{\mathbf{x}}_1\ b_1 + \ldots + \underline{\mathbf{x}}_n\ b_n)\ \left(\frac{1}{\|\underline{x}\|^2}\underline{\mathbf{x}}\right)$$

$$= A$$

(where the accuracy of A' or A depends completely on the precision of $\underline{\boldsymbol{\varepsilon}}$ that we have taken).

CONCLUSION

In a nutshell, this author has proved the HKLam Theory (both the "if" part and the "only if" part) by the elementary or basic linear algebra. It is no doubt that the theory has its extension for the tensor algebra and tensor analysis parts and has plenty of applications in physics and astronomphy etc. This author will left those discussions in the following series of the paper together with another one about its direct daily usage such as the real numerical data application of oral cavity teeth curvature in the field of 3D teeth modeling or even the controversial 3D cognitive brain signal input/output into human physical mind etc.

Reference

[1] Lang, S. (1987). Linear algebra (3rd ed.). Springer.
[2] Aggarwal, C. C. (2020). *Linear algebra and optimization for machine learning a textbook.* Springer.
[3] Halmos, P. R. (2016). *Finite dimensional vector spaces*. Princeton University Press.

[4] Lebedev, L. P., Lebedev, L. P., Cloud, M. J., & Eremeyev, V. A. (2010). *Tensor analysis with applications in mechanics*. World Scientific Publishing Co. Pte. Ltd.
[5] Woerdeman, H. J. (Hugo J. (2016). *Advanced linear algebra*. CRC Press, Taylor & Francis Group.
[6] Ben Klemens. (2008). Linear projections. In *Modeling with Data* (p. 264–). Princeton University Press.
[7] Leung, K. (1974). *Linear algebra and geometry*. Hong Kong University Press.
[8] Taboga, Marco (2021). "Matrix of a linear map", Lectures on matrix algebra. https://www.statlect.com/matrix-algebra/matrix-of-a-linear-map
[9] Lam, Kai Shun, Evaluation of the Weather-Influenza Pattern with a Regression Model Approximation to Causality (March 29, 2020). Available at SSRN: https://ssrn.com/abstract=3563601or http://dx.doi.org/10.2139/ssrn.3563601
[10] Gentle, J. E. (2017). *Matrix algebra : theory, computations and applications in statistics* (Second edition.). Springer. https://doi.org/10.1007/978-3-319-64867-5
[11] Favier, G. (2021). *Matrix and tensor decompositions in signal processing* (Vol. 2, p. 350–). Wiley-Blackwell.
[12] Wang, G., Wei, Y., & Qiao, S. (2018). *Generalized Inverses: Theory and Computations* (Vol. 53). Springer Singapore. https://doi.org/10.1007/978-981-13-0146-9