A Hilbert Space of Probability Mass Functions and Applications on the Sum-Product Algorithm

Muhammet Fatih Bayramoğlu and Ali Özgür Yılmaz
METU Dept. of EEE Ankara, Turkey
Email: {fatih, aoyilmaz}@eee.metu.edu.tr

***Abstract*—In this paper a Hilbert space structure of probability mass functions (PMF) will be presented. The tools provided by the Hilbert space, specifically the norm and the inner product, may be useful while analyzing and improving the sum-product algorithm in many aspects. Our approach provides a metric distance between PMFs and a new point of a view of the log-likelihood ratio (LLR) such that the LLR representation is nothing but a Hilbert space representation.**

I. Introduction

The sum-product algorithm has been extremely popular in the communication community since the invention of turbo codes [3], [4], [5]. The sum-product algorithm elegantly solves the marginalized product density problem in a message passing fashion. The messages passed during the sum product algorithm (beliefs) are in fact probability density functions (PDF). Therefore, representation of the PDFs is crucial for the sum-product algorithm.

A good candidate for the representation of PDFs is a linear vector space structure. The linear vector spaces are very powerful mathematical structures such that they are employed in many and very different areas of science.

A linear vector space always has the algebraic operations, namely the addition and the scalar multiplication. Some of the vector spaces have some other tools which are used to define a geometric structure over the vector space. These tools are the norm and the inner product.

Roughly speaking, the norm of a vector is defined as the length of the vector. Together with the subtraction operation the norm defines a distance between two vectors. A proper distance function between PDFs is frequently needed durin the researches on the sum-product algorithm, such as analyzing the convergence of the sum product algorithm or evaluating the performance of lossy message computation algorithms. The Kullback-Leibler (KL) divergence is employed in the literature as the distance function between PDFs [6], [7]. However, KL divergence is not a true distance since it does not satisfy triangle inequality and is not symmetric.

Second geometric operation is the inner product. The inner product defines the angle between two vectors, consequently the concept of orthogonality and projections. An inner product always defines a proper norm and hence a proper distance. Furthermore, we claim that, the inner product of PDFs may be quite useful for developing lossy message computation algorithms. If a proper inner product can be defined over a vector space, the vector space becomes an inner product space. Moreover, if the vector space is complete with respect to the inner product induced norm it is called a Hilbert space.

The sum-product algorithm is usually employed in the communication receivers in order to detect the transmitted symbols. These symbols usually take values from a discrete alphabet. Hence, the PDFs describing their probabilities are of discrete type. Therefore, in this paper we will restrict ourselves to the discrete PDFs or probability mass functions (PMF) and derive the Hilbert space of PMFs. In other words, we will define the algebraic and geometric operations for the PMFs. While defining all of these operations it should be kept in mind that these operations should be meaningful in terms of detection and they should be useful in the sum-product algorithm.

Our definition of the inner product of PMFs suggests a trivial mapping from the set of PMFs to $\mathbb{R}^N$. Surprisingly, this mapping is nothing but the log-likelihood ratio (LLR) representation. In other words, our work on deriving a Hilbert space of PMFs ends up with the well known LLR. The LLRs first introduced in the classical detection theory in order to get rid of the exponential functions frequently faced due to the normal distribution [9]. Then the turbo decoding algorithm assigned more jobs to LLRs. Firstly, they are used for defining the extrinsic information. Secondly, LLRs are used for numerical stability of the sum-product algorithm. Finally, LLRs are employed in methods for analyzing the iterative algorithms. Now, we assign a different meaning to LLR representation such that the LLR representation is a Hilbert space representation of PMFs.

A Hilbert space of PDFs is first presented in literature in [8]. Their derivation is for a class of continuous PDFs. On the other hand our derivation is for discrete PDFs or PMFs. Although, both of our and their work can be derivable from each other easily, our derivation is independent of theirs. Moreover, we emphasize the connection between the Hilbert space of PMFs, the detection theory and the sum-product algorithm.

This paper is organized as follows. In the following section the algebraic structure of the vector space will be defined. In the next section the geometric structure of the Hilbert space of PMFs will be derived. Finally the representation of the sum-product algorithm by using the Hilbert space of PMFs will be presented. The details in the proofs in the paper are omitted due to the lack of space. However, all of the omitted details are usually direct consequences of the definitions.

978-1-4244-2863-2/08/$25.00 © 2008 IEEE

II. Algebraic Structure

The first step in constructing a Hilbert space over a set is constructing an algebraic structure over the set. The algebraic structure of a vector space consists of addition and scalar multiplication operations. While describing these operations, the emphasis will be on making them meaningful in the sense of detection.

First of all, the set should be defined formally. Let the set $\mathcal{A}$ be the support set of a discrete random variable. Then $\mathcal{V}_{\mathcal{A}}$ is defined as the set of all possible PMFs defined over $\mathcal{A}$:

$$\mathcal{V}_{\mathcal{A}} \triangleq \{q(x) : \mathcal{A} \to [0,1] | \sum_{\forall x_i \in \mathcal{A}} q(x_i) = 1\}. \tag{1}$$

This set contains the necessary PMFs encountered in a detection process of discrete random variables. The detection process can be summarized as estimating the value of an unknown random variable from an observed variable. For instance, assume that the random variable $\tilde{x}$, whose support set is $\mathcal{A}$, is to be estimated given the observed variable $\tilde{r}$. The known densities are usually $p_{\tilde{x}}(x)$ and $f_{\tilde{r}|\tilde{x}}(r|x)$. Obviously $p_{\tilde{x}}(x)$ is an element of $\mathcal{V}_{\mathcal{A}}$, whereas $f_{\tilde{r}|\tilde{x}}(r|x)$ is not. However, in a detection process we are usually more interested in $p_{\tilde{x}|\tilde{r}}(x|r)$ rather than $f_{\tilde{r}|\tilde{x}}(r|x)$. If we regard r in $p_{\tilde{x}|\tilde{r}}(x|r)$ as a parameter of the function rather than an argument, then $p_{\tilde{x}|\tilde{r}}(x|r)$ becomes an element of $\mathcal{V}_{\mathcal{A}}$.

Now the addition operation can be defined. In order to explain the meaning of the addition operation consider the following situation. A random variable $\tilde{x}$ is to be detected using two observed variables $\tilde{r_1}$ and $\tilde{r_2}$ which are conditionally independent given $\tilde{x}$. In other words;

$$f_{\tilde{r}_1,\tilde{r}_2|\tilde{x}}(r_1, r_2|x) = f_{\tilde{r_1}|\tilde{x}}(r_1|x) f_{\tilde{r_2}|\tilde{x}}(r_2|x). \tag{2}$$

Moreover, let all events of $\tilde{x}$ be equally likely. We can infer some information on $\tilde{x}$ given the value of $\tilde{r_1}$. This information is the a posteriori PMF of $\tilde{x}$ given $\tilde{r_1}$, i.e. $p_{\tilde{x}|\tilde{r_1}}(x|r_1)$. Let $q_1(x) = p_{\tilde{x}|\tilde{r_1}}(x|r_1)$. Similarly let the deduced information about $\tilde{x}$ given $\tilde{r_2}$ be $q_2(x) = p_{\tilde{x}|\tilde{r_2}}(x|r_2)$. Then it is meaningful to define the addition of $q_1(x)$ and $q_2(x)$ as the total information inferred about $\tilde{x}$ given $\tilde{r_1}$ and $\tilde{r_2}$ together. In other words;

$$q_1(x) \oplus q_2(x) = p_{\tilde{x}|\tilde{r}_1,\tilde{r}_2}(x|r_1, r_2) \tag{3}$$

where $\oplus$ denotes the addition operator. In order to derive $q_1(x) \oplus q_2(x)$ in terms of $q_1(x)$ and $q_2(x)$ Bayes' theorem and conditional independence can be used as follows:

$$\begin{aligned} q_1(x) \oplus q_2(x) &= \frac{f_{\tilde{r_1},\tilde{r_2}|\tilde{x}}(r_1, r_2|x) p_{\tilde{x}}(x)}{f_{\tilde{r_1},\tilde{r_2}}(r_1, r_2)} \\ &= \frac{f_{\tilde{r_1}|\tilde{x}}(r_1|x) f_{\tilde{r_2}|\tilde{x}}(r_2|x) p_{\tilde{x}}(x)}{f_{\tilde{r_1},\tilde{r_2}}(r_1, r_2)}. \end{aligned}$$

Using Bayes' theorem once more the following expression can be derived.

$$\begin{aligned} q_1(x) \oplus q_2(x) &= \frac{p_{\tilde{x}|\tilde{r_1}}(x|r_1) p_{\tilde{x}|\tilde{r_2}}(x|r_2) f_{\tilde{r_1}}(r_1) f_{\tilde{r_2}}(r_2)}{f_{\tilde{r_1},\tilde{r_2}}(r_1, r_2) p_{\tilde{x}}(x)} \\ &= q_1(x)\ q_2(x) \frac{f_{\tilde{r_1}}(r_1) f_{\tilde{r_2}}(r_2)}{f_{\tilde{r_1},\tilde{r_2}}(r_1, r_2) p_{\tilde{x}}(x)} \end{aligned} \tag{4}$$

Since it is assumed that all events of $\tilde{x}$ are equally likely, the last term in the multiplication in Equation 4 is independent of x and can be replaced with a constant, i.e.;

$$q_1(x) \oplus q_2(x) = C\ q_1(x)\ q_2(x). \tag{5}$$

We know from Equation 3 that $q_1(x) \oplus q_2(x)$ is a posteriori PMF. Hence, the area underneath should be equal to 1. Therefore,

$$C = \frac{1}{\sum_{\forall x_i \in \mathcal{A}} q_1(x_i) q_2(x_i)}. \tag{6}$$

Finally the addition of PMFs can be defined as follows;

$$q_1(x) \oplus q_2(x) \triangleq \frac{q_1(x)\ q_2(x)}{\sum_{\forall x_i \in \mathcal{A}} q_1(x_i) q_2(x_i)}. \tag{7}$$

After defining the addition operation the scalar multiplication operation, which will be denoted by $\odot$, can be defined for positive integer scalars easily as follows:

$$\begin{aligned} n \odot q(x) &= \underbrace{q(x) \oplus q(x) \oplus \ldots \oplus q(x)}_{\text{n times}} \qquad (8) \\ &= \frac{q^n(x)}{\sum_{\forall x_i \in \mathcal{A}} q^n(x)} \qquad (9) \end{aligned}$$

This result can be generalized to all real numbers and the definition of the scalar multiplication can be obtained as in Equation 10.

$$r \odot q(x) \triangleq \frac{q^r(x)}{\sum_{\forall x_i \in \mathcal{A}} q^r(x)} \tag{10}$$

Now it should be proven that these elements form a vector space over $\mathbb{R}$.

Theorem 1. *The set $\mathcal{V}_{\mathcal{A}}$ together with the $\oplus$ and $\odot$ operations forms a vector space over $\mathbb{R}$.*

Proof:. Firstly it should be proven that the $\mathcal{V}_A$ and the $\oplus$ operation forms an abelian group.

- Closure: For all $q_1(x), q_2(x) \in \mathcal{V}_{\mathcal{A}}$,
$$q_3(x) = q_1(x) \oplus q_2(x) \in \mathcal{V}_{\mathcal{A}}. \tag{11}$$
- Commutativity: For all $q_1(x), q_2(x)$ in $\mathcal{V}_{\mathcal{A}}$
$$q_1(x) \oplus q_2(x) = q_2(x) \oplus q_1(x). \tag{12}$$
- Associativity: For all $q_1(x), q_2(x), q_3(x)$ in $\mathcal{V}_{\mathcal{A}}$
$$\begin{aligned} (q_1(x) \oplus q_2(x)) \oplus q_3(x) &= q_1(x) \oplus (q_2(x) \oplus q_3(x)) \\ &= q_1(x) \oplus q_2(x) \oplus q_3(x) \end{aligned}$$
- Neutral element: Let $e(x) = \frac{1}{|\mathcal{A}|}$, where $|\mathcal{A}|$ denotes the number of elements in $\mathcal{A}$. Then for all $q(x)$ in $\mathcal{V}_{\mathcal{A}}$
$$q(x) \oplus e(x) = e(x) \oplus q(x) = q(x)$$
- Inverse element: For all $q(x) \in \mathcal{V}_{\mathcal{A}}$ there exist a
$$q^{-1}(x) = \frac{\frac{1}{q(x)}}{\sum_{\forall x_i \in \mathcal{A}} \frac{1}{q(x)}}$$
such that $q(x) \oplus q^{-1}(x) = q^{-1}(x) \oplus q(x) = e(x)$.

All of these properties can be proven as direct consequences of Equation 7.

Now the required properties of the scalar multiplication have to be proven.

- Closure: For all $\alpha \in \mathbb{R}$ and $q(x) \in \mathcal{V}_\mathcal{A}$,
$$\alpha \odot q(x) \in \mathcal{V}_\mathcal{A}. \tag{13}$$
- Compatibility with the multiplication in $\mathbb{R}$: For all α, β in $\mathbb{R}$ and all $q(x) \in \mathcal{V}_\mathcal{A}$:
$$\alpha \odot (\beta \odot q(x)) = (\alpha\beta) \odot q(x)$$
- Distributivity over vector addition: For all α in $\mathbb{R}$ and $q_1(x), q_2(x)$ in $\mathcal{V}_\mathcal{A}$:
$$\alpha \odot (q_1(x) \oplus q_2(x)) = \alpha \odot q_1(x) \oplus \alpha \odot q_2(x)$$
- Distributivity over scalar addition: For all α, β in $\mathbb{R}$ and $q(x)$ in $\mathcal{V}_\mathcal{A}$:
$$(\alpha + \beta) \odot q(x) = \alpha \odot q(x) \oplus \beta \odot q(x)$$

All these properties directly follows the definitions of the addition and the scalar multiplication. Finally, $(\mathcal{V}_\mathcal{A}, \oplus, \odot)$ is a vector space over $\mathbb{R}$.

$\diamond$

III. Geometric Structure

The next step in defining a Hilbert space is defining the geometric structure of the vector space and showing the completeness of the vector space. The geometric structure of a vector space can be described by an inner product. In order to define an inner product in the vector space of PMFs, firstly we will try to propose an answer to the question what the meaning of angle between two PMFs is. Secondly, we will propose an inner product function and prove that the proposed function satisfies the inner product axioms. After having defined the inner product we will investigate the inner product induced norm and distance function on the vector space of PMFs. Finally we will show that the vector space of PMFs is a finite dimensional vector space and hence, it is a Hilbert space.

A. The Angle Between Two PMFs

The angle between two vectors has an important property. This property is the fact that the angle between two vectors is kept constant when the vectors are scaled by nonzero scalars. In other words, for any $q_1(x), q_2(x)$ in $\mathcal{V}_\mathcal{A}$ and any nonzero scalars α, β in $\mathbb{R}$
$$\angle(q_1(x), q_2(x)) = \angle(\alpha \odot q_1(x), \beta \odot q_2(x)), \tag{14}$$
where $\angle(.,.)$ denotes the angle between two PMFs. This idea suggests that we should investigate the relations between families of PMFs which are constructed by scaling certain PMFs.

Representing PMFs in a coordinate space makes easier to analyze the relation between these families of PMFs. Consider any $q(x)$ in $\mathcal{V}_\mathcal{A}$. Let $q(x_i) = p_i$ for all i in $1, 2, \ldots, |\mathcal{A}|$. Each $q(x)$ can be represented by a point in the coordinate space of p_i's. However, the reverse is not true. Each point in the coordinate space of p_i's does not represent a valid PMF. Coordinates of a point should satisfy the following properties in order to be able to represent a PMF.

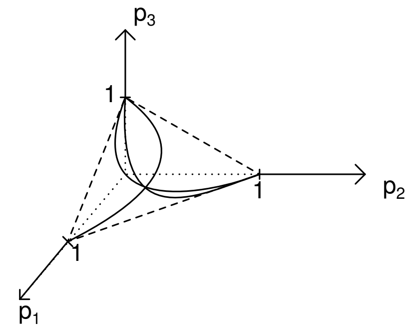

Fig. 1. Three different parametric curves which are constructed by scaling three different PMFs for $|\mathcal{A}| = 3$. Note that each curve intersects at the center of mass of the triangle.

$$\sum_{i=1}^{|\mathcal{A}|} p_i = 1 \tag{15}$$
$$p_i \in [0, 1] \quad \forall i \in 1, 2, \ldots, |\mathcal{A}| \tag{16}$$

The Equation 15 defines a hyperplane, and Equation 16 defines a region on the hyperplane.

Now consider a $q(x)$ and its scaled versions $\alpha \odot q(x)$ for all α in $\mathbb{R}$. The corresponding points to the $\alpha \odot q(x)$ in the coordinate space forms a parametric curve. For any $q(x)$ in $\mathcal{V}_\mathcal{A}$, $0 \odot q(x) = \frac{1}{|\mathcal{A}|}$. This means that the parametric curve $\alpha \odot q(x)$, passes through the point $p_i = \frac{1}{|\mathcal{A}|}$, which is the center of mass of the region defined by Equations 15 and 16. As a result any two parametric curves $\alpha \odot q_1(x)$ and $\beta \odot q_2(x)$ for any $q_1(x), q_2(x)$ in $\mathcal{V}_\mathcal{A}$ always intersect at the point $p_i = \frac{1}{|\mathcal{A}|}$. Figure 1 depicts an example of this idea for $|\mathcal{A}| = 3$.

At this stage we can reasonably define the angle between two PMFs as the angle between these two parametric curves at their intersection points. In order to derive the angle between them, a more formal definition of these curves is required. Let $\mathbf{c}_1(\alpha)$, which is a vector function of a scalar, be the curve composed of the points corresponding to $\alpha \odot q_1(x)$. Then $\mathbf{c}_1(\alpha)$ is given as
$$\mathbf{c}_1(\alpha) = \sum_{i=1}^{|\mathcal{A}|} \frac{p_i^\alpha}{\sum_{j=1}^{|\mathcal{A}|} p_j^\alpha} \mathbf{e}_i \tag{17}$$
where $\mathbf{e}_i$'s denote the canonical bases vectors of $\mathbb{R}^{|\mathcal{A}|}$. The tangential vector to this curve at $\alpha = 0$, which will be denoted by $\mathbf{t}_1$, is:
$$\mathbf{t}_1 = \left. \left(\sum_{i=1}^{|\mathcal{A}|} \frac{\mathrm{d}}{\mathrm{d}\alpha} \left(\frac{p_i^\alpha}{\sum_{j=1}^{|\mathcal{A}|} p_j^\alpha} \right) \mathbf{e_i} \right) \right|_{\alpha=0} \tag{18}$$
$$= \frac{1}{|\mathcal{A}|^2} \sum_{i=1}^{|\mathcal{A}|} \log \frac{q_1^{|\mathcal{A}|}(x_i)}{\prod_{j=1}^{|\mathcal{A}|} q_1(x_j)} \mathbf{e_i} \tag{19}$$

Similarly, let $\mathbf{c}_2(\beta)$ is the curve corresponding to $\beta \odot q_2(x)$. Then the tangential vector to this curve at $\beta = 0$, $\mathbf{t}_2$, is:

$$\mathbf{t}_2 = \frac{1}{|\mathcal{A}|^2} \sum_{i=1}^{|\mathcal{A}|} \log \frac{q_2^{|\mathcal{A}|}(x_i)}{\prod_{j=1}^{|\mathcal{A}|} q_2(x_j)} \mathbf{e_i} \tag{20}$$

Then the angle between $q_1(x)$ and $q_2(x)$ can be calculated by

$$\angle(q_1(x), q_2(x)) = \cos^{-1} \frac{< \mathbf{t}_1, \mathbf{t}_2 >}{\|\mathbf{t}_1\| \, \|\mathbf{t_2}\|}, \tag{21}$$

where $< .,. >$ and $\|.\|$ denotes the usual inner product and norm in $\mathbb{R}^{|\mathcal{A}|}$ respectively.

B. The Inner Product of PMFs

In the previous section, the angle between two PMFs is defined as the angle between the vectors $\mathbf{t}_1$ and $\mathbf{t}_2$ which are the tangential vectors to the parametric curves obtained by scaling the two PMFs. It is quite reasonable to define the inner product of two PMFs as the inner product of these tangential vectors. When the constants in the definitions of t hese tangential vectors are discarded, an inner product between two PMFs $q(x), r(x)$ in $\mathcal{V}_{\mathcal{A}}$ can be proposed as follows:

$$\sigma(q(x), r(x)) \triangleq \sum_{i=1}^{|\mathcal{A}|} \log \frac{q^{|\mathcal{A}|}(x_i)}{\prod_{j=1}^{|\mathcal{A}|} q(x_j)} \log \frac{r^{|\mathcal{A}|}(x_i)}{\prod_{j=1}^{|\mathcal{A}|} r(x_j)} \tag{22}$$

It should be shown that $\sigma(q(x), r(x))$ satisfies the inner product axioms.

Theorem 2. *The function $\sigma(q(x), r(x)) : \mathcal{V}_{\mathcal{A}} \times \mathcal{V}_{\mathcal{A}} \to \mathbb{R}$ is an inner product in $\mathcal{V}_{\mathcal{A}}$.*

Proof:. Three inner product axioms should be proven for $\sigma(q(x), r(x))$.

- Commutativity: For any $q(x), r(x)$ in $\mathcal{V}_{\mathcal{A}}$:

$$\sigma(q(x), r(x)) = \sigma(r(x), q(x))$$

- Linearity with respect to first argument: For any $q_1(x)$,$q_2(x)$, and $r(x)$ in $\mathcal{V}_{\mathcal{A}}$, and α, β in $\mathbb{R}$.

$$\begin{aligned} \sigma(\alpha \odot q_1(x) \oplus \beta \odot q_2(x), r(x)) = \\ \alpha\sigma(q_1(x), r(x)) + \beta\sigma(q_2(x), r(x)) \end{aligned}$$

- Nonnegativity: For any $q(x)$ in $\mathcal{V}_{\mathcal{A}}$:

$$\sigma(q(x), q(x)) \geq 0$$

Equality holds if and only if $q(x) = e(x)$.

All of these items can be proven directly from the Equation 22, and hence, $\sigma(q(x), r(x))$ is an inner product and the inner product notation can be used for it as follows:

$$< q(x), r(x) > \triangleq \sigma(q(x), r(x)) \tag{23}$$

Up to now we have proven that $(\mathcal{V}_{\mathcal{A}}, < .,. >)$ is an inner product space. The only missing step to show that $\mathcal{V}_{\mathcal{A}}$ is a Hilbert space, is to prove that $\mathcal{V}_{\mathcal{A}}$ is complete. Before proving that $\mathcal{V}_{\mathcal{A}}$ is complete, inner product induced norm will be investigated in the next section.

C. The Inner Product Induced Norm of PMFs

Every inner product defines a norm in the vector space it belongs to. So does $< .,. >$ in $\mathcal{V}_{\mathcal{A}}$. This norm is:

$$\|q(x)\| \triangleq \sqrt{< q(x), q(x) >} \tag{24}$$

This norm inherits all the properties of a norm. These are;

- Scalability: For any $\alpha \geq 0$ in $\mathbb{R}$ and $q(x)$ in $\mathcal{V}_{\mathcal{A}}$

$$\|\alpha \odot q(x)\| = \alpha \, \|q(x)\| \,. \tag{25}$$

- Nonnegativity: For any $q(x)$ in $\mathcal{V}_{\mathcal{A}}$

$$\|q(x)\| \geq 0 \tag{26}$$

and equality holds if and only if $q(x) = e(x)$.

- Triangle Inequality: For any $q(x), r(x)$ in $\mathcal{V}_{\mathcal{A}}$:

$$\|q(x) \oplus r(x)\| \leq \|q(x)\| + \|r(x)\| \,. \tag{27}$$

D. The Distance Between PMFs

An important tool that this norm provides is a distance function between PMFs which satisfies all requirements for being a metric distance. Unlike existing divergence functions, such as KL divergence and Jensen-Shannon divergence, the following distance function, defined in Equation 28, is symmetric, nonnegative, and zero if and only if the PMFs are equal. Moreover, this distance function satisfies triangle inequality.

$$D(q(x), r(x)) \triangleq \|q(x) \ominus r(x)\| \tag{28}$$

Our empirical studies show that, this distance function relates to the KL divergence as follows:

$$0 < \frac{D_{KL}(q(x), r(x))}{D(q(x), r(x))} < C, \tag{29}$$

where C is a constant depending on $|\mathcal{A}|$ and the ratio can be made arbitrarily close to zero by suitable selection of $p(x)$ and $q(x)$.

The question of whether this distance function is better than the KL divergence is very difficult to answer. The answer to this question depends on the application and the definition of being *better*. In the context of sum-product algorithm satisfaction of the triangle inequality is important as it will be explained in Section IV-C. Hence, we claim that in the context of sum-product algorithm the distance function defined in Equation 28 is more useful than the KL divergence.

Moreover, the Equation 29 states that, $CD(p(x), q(x))$ provides an upper bound for the KL divergence. Since the mutual information and channel capacity are defined using KL divergence, $D(p(x), q(x))$ can be used for calculating bounds for the capacity of some channels which have capacities difficult to calculate.

E. The LLR Representation of PMFs

The definition of inner product existing in Equation 22 suggests a transformation from $\mathcal{V}_{\mathcal{A}}$ to $\mathbb{R}^{|\mathcal{A}|}$. This transformation can be defined as;

$$\begin{aligned}\mathbf{q} &= L\{q(x)\} \\ &\triangleq \begin{pmatrix} \log \frac{q^{|\mathcal{A}|}(x_1)}{\prod_{j=1}^{|\mathcal{A}|} q(x_j)} \\ \log \frac{q^{|\mathcal{A}|}(x_2)}{\prod_{j=1}^{|\mathcal{A}|} q(x_j)} \\ \vdots \\ \log \frac{q^{|\mathcal{A}|}(x_{|\mathcal{A}|})}{\prod_{j=1}^{|\mathcal{A}|} q(x_j)} \end{pmatrix}\end{aligned} \tag{30}$$

Then the definition of the inner product can be rewritten more simply as;

$$\begin{aligned}< q(x), r(x) > &= L\{q(x)\} \cdot L\{r(x)\} \\ &= \mathbf{q} \cdot \mathbf{r}\end{aligned} \tag{31}$$

Besides simplifying the definition of inner product, this transformation has an important property stated by the following theorem.

Theorem 3. *$L\{.\} : \mathcal{V}_{\mathcal{A}} \to \mathbb{R}^{|\mathcal{A}|}$ is a linear map. In other words;*

$$L\{\alpha \odot q_1(x) \oplus \beta \odot q_2(x)\} = \alpha L\{q_1(x)\} + \beta L\{q_2(x)\}$$

for all $q_1(x), q_2(x)$ in $\mathcal{V}_{\mathcal{A}}$ and α, β in $\mathbb{R}$.

The proof of this theorem readily follows from the definitions of $L\{.\}$, the addition, and the multiplication of PMFs.

This theorem states that, PMFs can be transformed into $\mathbb{R}^{|\mathcal{A}|}$, and algebraic operations can be conducted in this domain. Moreover, as it is shown in the Equation 31 the inner products can be calculated in the $\mathbb{R}^{|\mathcal{A}|}$ as well. In other words, any PMF can be represented by a vector in $\mathbb{R}^{|\mathcal{A}|}$, and any algebraic or geometric operation conducted on the PMFs can be represented by the same operation in $\mathbb{R}^{|\mathcal{A}|}$.

Surprisingly, this representation is nothing but a redundant version of the LLR representation which is very familiar to the turbo coding community. For instance, for $|\mathcal{A}| = 2$ the transformation becomes

$$L\{q(x)\} = \begin{pmatrix} \log \frac{q(x_1)}{q(x_2)} \\ \log \frac{q(x_2)}{q(x_1)} \end{pmatrix}. \tag{32}$$

For alphabet sizes greater than 2 the representation is still redundant. The entries of the vector in the $\mathbb{R}^{|\mathcal{A}|}$ always sum up to zero as it is shown below. Let $\mathbf{q} = L\{q(x)\}$ for any alphabet size. Then

$$\begin{aligned}\sum_{i=1}^{|\mathcal{A}|} (\mathbf{q})_i &= \sum_{i=1}^{|\mathcal{A}|} \log \frac{q^{|\mathcal{A}|}(x_i)}{\prod_{j=1}^{|\mathcal{A}|} q(x_j)} \\ &= 0.\end{aligned} \tag{33}$$

Note that the Equation 33 also defines the range space of the mapping $L\{\}$. This range space is a hyperplane which is $|\mathcal{A}| - 1$ dimensional subspace of $\mathbb{R}^{|\mathcal{A}|}$.

The LLR representation always played an important role while analyzing the performance of sum-product algorithms. We believe that the underlying reason is that the LLR representation is a vector space representation of PMFs.

F. Completeness of the Vector Space

Completeness of an inner product space is important for working on convergence or approximations. It is known from functional analysis theory that a finite dimensional inner product space is always complete. Therefore, showing that the vector space of PMFs is finite dimensional is enough to show that the vector space of PMFs is complete.

Theorem 4. *The vector space $(\mathcal{V}_{\mathcal{A}}, \oplus, \odot)$ is a $|\mathcal{A}| - 1$ dimensional vector space.*

This theorem can be proven by finding a linear mapping from $\mathcal{V}_{\mathcal{A}}$ to $\mathbb{R}^{|\mathcal{A}|-1}$ which is one-to-one and onto. In the previous section we have defined the linear map $L\{.\}$ and observed that its range space is $|\mathcal{A}|-1$ dimensional. By slightly modifying the mapping $L\{.\}$ we can obtain a one-to-one and onto transformation from $\mathcal{V}_{\mathcal{A}}$ to $\mathbb{R}^{|\mathcal{A}|-1}$ and the vector space of PMFs becomes a Hilbert space.

IV. Representation of the Sum-Product Algorithm Using the Hilbert Space of PMFs

Since we have defined the addition operation between PMFs to be meaningful in the detection sense, the Hilbert space of PMFs results in a trivial representation for message calculation in the sum-product algorithm. Before presenting the representation we need to define two operators which will simplify the representation.

A. The Expansion Operator

The expansion operator is used for determining the joint PMF of a combined experiment. This combined experiment is assumed to be composed of independent individual experiments, one of which has a known PMF and others distributed equally likely. After this introduction the expansion operator can be defined formally as follows. Consider an ordered set of N independent experiments with support sets $\mathcal{A}_1$, $\mathcal{A}_2$, …, $\mathcal{A}_N$. The expansion operator of the i^{th} experiment, $E_i : \mathcal{V}_{\mathcal{A}_i} \to \mathcal{V}_{\mathcal{A}_1 \times \mathcal{A}_2 \times \ldots \times \mathcal{A}_N}$ is defined as

$$\begin{aligned}E_i\{q(x)\} &\triangleq \frac{1}{\prod_{j \in \{1..N\} \setminus \{i\}} |\mathcal{A}_j|} q(x_i) \\ &= q_{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_N}(x_1, x_2, \ldots, x_N)\end{aligned} \tag{34}$$

Then the joint PMF of a combined experiment which is composed of independent experiments all of which has a known PMF can be calculated from the individual PMFs as follows;

$$q_{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_N}(x_1, x_2, \ldots, x_N) = \bigoplus_{i=1}^{N} E_i\{q_i(x_i)\}. \tag{35}$$

Obviously the expansion operator is a linear operator.

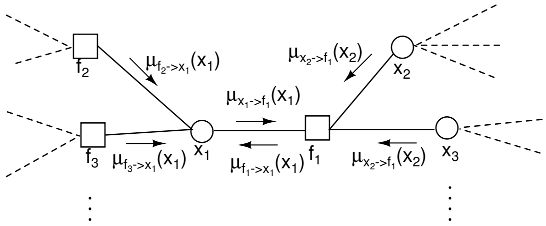

Fig. 2. A sample factor graph showing the message nomenclature

B. The Marginalization Operator

The marginalization operator, as its name implies, is used for obtaining the marginal PMF from a given joint PMF. The marginalization operator $M : \mathcal{V}_{\mathcal{A}_1\times\mathcal{A}_2\times\ldots\times\mathcal{A}_n} \rightarrow \mathcal{V}_{\mathcal{A}_i}$ is defined as

$$M_i\{q(x_1,x_2,\ldots,x_N)\} \triangleq \sum_{\forall(x_1,x_2,\ldots,x_{i-1},x_{i+1},\ldots,x_N)} q(x_1,x_2,\ldots,x_N). \quad (36)$$

Marginalization and expansion operators are not the inverse operators of each other in general. Although the equation

$$q_i(x_i) = M_i\{E_i\{q_i(x_i)\}\} \quad (37)$$

always holds, the reverse is not true in general.

On the contrary to the expansion operator the marginalization operator is not linear.

C. The Message Update Rules

After defining the expansion and marginalization operators we will show how the Hilbert space representation naturally fits to message update rules in the sum-product algorithm. We will adopt the same factor graph representation and message nomenclature as in [4], which is shown in Figure 2.

The representation of the message calculation in the variable nodes becomes extremely simple thanks to the definition of the addition of PMFs which is given below.

$$\mu_{x_i\rightarrow f_j}(x_i) = \bigoplus_{k\in n(x_i)\backslash\{f_j\}} \mu_{f_k\rightarrow x_i}(x_i), \quad (38)$$

where $n(x_i)$ denotes the set of neighbors of the node x_i.

Before passing to the message calculation rules at the factor nodes, it should be noted that factor functions can be regarded as PMFs after some proper scaling which does not affect the operation of the algorithm. Hence, they can be represented as vectors in the Hilbert space of PMFs. Then the message calculation rules at the factor nodes is given as follows:

$$\mu_{f_i\rightarrow x_j}(x_j) = M_j\left\{ f_i(n(f_i)) \oplus \bigoplus_{k\in n(f_i)\backslash\{x_j\}} E_k\{\mu_{x_k\rightarrow f_i}(x_k)\}\right\} \quad (39)$$

If the marginalization operator was a linear operator, the messages could be computed by multiplying the incoming message vectors with some matrices and adding the results up. This would result in a message computation complexity proportional to $|\mathcal{A}|^2$. Unfortunately, the marginalization operator is not a linear operator. However, the marginalization operator can be linearized over a specific region. This approach may lead to new lossy message computation methods.

Equations 38 and 39 show that the messages are added up during message computation. This fact emphasizes the importance of using a distance function satisfying the triangle inequality while evaluating the approximateness of the lossy messages. If the errors in the incoming lossy messages are bounded with respected to the distance defined in Equation 28 then it is guaranteed that the errors in the outgoing messages will also be bounded with respected to the same distance. For instance assume that two incoming messages to a node are $p(x)$ and $q(x)$. If these messages are approximated by $\hat{p}(x)$ and $\hat{q}(x)$ such that $D(p(x),\hat{p}(x)) < \epsilon$ and $D(q(x),\hat{q}(x)) < \epsilon$ then it is guaranteed due to the triangle inequality that

$$D(p(x)\oplus q(x), \hat{p}(x)\oplus\hat{q}(x)) < 2\epsilon. \quad (40)$$

Moreover, due to Equation 29 KL divergence in the output will also be bounded as follows:

$$D_{KL}(p(x)\oplus q(x), \hat{p}(x)\oplus\hat{q}(x)) < 2C\epsilon \quad (41)$$

However, no bound exist for the error in the output message in terms KL divergence between the input messages.

V. Conclusion

In this paper we have presented the Hilbert space of PMFs. We believe that the Hilbert space of PMFs can be a useful tool for analyzing and improving the sum-product algorithm. Actually, we have already known some of the applications of the Hilbert space of PMFs due to the fact that the LLR is actually a Hilbert space representation.

References

[1] A. Papoulis, "Probability, Random Variables and Stochastic Processes" McGraw Hill, 1991
[2] T. M. Cover and J. A. Thomas, "Elements of Information Theory" John Wiley&Sons, 1991
[3] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit error correcting coding and decoding: Turbo Codes" in Proc. 1993 IEEE Int. Conf. Comm., Gen., Switzerland, May. 1993, pp. 1064-1070
[4] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor Graphs and the Sum-Product Algorithm", IEEE Transactions on Information Theory, vol.47, No.2,pp.498-519 February 2001
[5] H. A. Loeliger, "An Introduction to Factor Graphs", IEEE Signal Processing Magazine, Vol. 21, Issue 1,pp.28-41 Jan. 2004,
[6] T. P. Minka, "A family of algorithms for approximate Bayesian inference", Ph. D. Thesis, Department of Electrical Eng. and Computer Science, Massachusetts Institute of Technology, 2001
[7] J. Hu, H.-A. Loeliger, J. Dauwels, and F. Kschischang, " A general computation rule for lossy summaries/messages with examples from equalization" Proc. 44th Allerton Conf. on Communication, Control, and Computing, Monticello, Illinois, Sept. 27-29, 2006.
[8] J. J. Egozcue, J. L. Diaz-Barrero, V. Pawlowsky-Glahn, "A Hilbert Space of Probability Density Functions Based on Aitchison Geometry", Acta Mathematica Sinica, Springer Berlin, July 2006
[9] H. L. van Trees, "Detection, Estimation, and Modulation Theory", John Wiley&Sons, 2001