DTU Library

Extending models for two-dimensional constraints

Forchhammer, Søren

Published in:
Proceedings, ISIT

Link to article, DOI:
10.1109/ISIT.2009.5206052

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

ISIT 2009, Seoul, Korea, June 28 - July 3, 2009

Extending Models for Two-Dimensional Constraints

Søren Forchhammer
Department of Photonics Engineering
Technical University of Denmark
2800 Lyngby, Denmark
Email: sofo@fotonik.dtu.dk

***Abstract*— Random fields in two dimensions may be specified on 2×2 elements such that the probabilities of finite configurations and the entropy may be calculated explicitly. The Pickard random field is one example where probability of a new (non-boundary) element is conditioned on three previous elements. To extend the concept we consider extending such a field such that a vector or block of elements is conditioned on a larger set of previous elements. Given a stationary model defined on 2×2 elements, iterative scaling is used to define the extended model. The extended model may be used for models of two-dimensional constraints and as examples we apply it to the hard-square constraint and the no isolated bits (n.i.b) constraint. The iterative scaling can ensure that the entropy of the extension is optimized and that the entropy is increased compared to the initial model defined on 2×2 elements. Application to a simple stationary model with hidden states is also outlined. For the n.i.b constraint, the initial model is based on elements defined by blocks of (1×2) binary symbols.**

I. Introduction

We consider two-dimensional random fields, especially those fields, which must satisfy some shift invariant constraints of finite extent (M, N) over some finite alphabet $\mathcal{A}$. A constraint is defined by a list, $\mathcal{F}$, of forbidden configurations of symbols in $\mathcal{A}$. Each forbidden configuration is contained within a rectangle of maximum size $M \times N$. A configuration on a finite segment of the plane containing no forbidden configurations is called an admissible configuration. Let $E(m, n)$ be the set of admissible configurations on an $m \times n$ rectangle for a given field F. The combinatorial entropy (or capacity) of F is defined as

$$C(F) = \limsup_{m,n \to \infty} \frac{\log_2 |E(m,n)|}{mn}. \tag{1}$$

As discussed in [1] this limit is well defined, even though it may not be computable. The limit is identical if we consider rotated rectangles, parallelograms, or other segments where the length of the boundary is small compared to the area [2] [3] [4].

In this work, we shall approach the entropy by assigning probability measures to the configurations. Let F be the elements of an $m \times n$ rectangle. Let μ_F be a probability measure on $\mathcal{A}^{m \times n}$.

Let the elements of the $m \times n$ configuration be denoted $x_{ij}, 1 \leq i \leq m, 1 \leq j \leq n$. We shall also use $\mu_{m \times n}$ for the measure on a $m \times n$ rectangle. The (measure theoretic) entropy of μ_F is defined as

$$H(\mu_F) = -\frac{1}{mn} \sum_{\mathbf{x} \in \mathcal{A}^{m \times n}} \mu_F(\mathbf{x}) \log_2 \mu_F(\mathbf{x}). \tag{2}$$

For a two-dimensional constrained field, the measure is defined on the configurations, which agree with the constraint. That is, each of the $|\mathcal{A}|^{n \times m}$ possible configurations, $\mathbf{x}$, that contain forbidden words defined by $\mathcal{F}$ have probability zero according to μ_F, $0 \leq \mu_F \leq 1$.

A measure is said to be (locally) stationary if two configurations within the rectangle that differ only by a translation, i.e. a fixed offset (i', j') of indices of the elements, x_{ij}, have the same probability. Besides the entropy, a stationary measure may also provide useful information as correlation properties.

A basic example of a 2-D constraint is the 2-D hard-square constraint defined over binary elements by requiring that two neighbors can not be 1, i.e. $x_{ij}x_{i,j-1} \neq 1$ and $x_{i-1,j}x_{ij} \neq 1$. This constraint is also called the 2-D RLL(1,∞) as no row or column has a run of two or more 1s. Coding for the hard-square constraint was presented in [4] based on the probability distribution of four elements.

Initially we consider stationary probability measures of the form (throughout the paper we use short hand notation for probabilities $P()$ given by their arguments),

$$\begin{aligned} \mu_{m \times n}(\mathbf{x}) = & P(x_{11}) \cdot \Pi_{j=2}^{n} P(x_{1j}|x_{1(j-1)}) \\ & \cdot \Pi_{i=2}^{m} P(x_{i1}|x_{(i-1)1}) \\ & \cdot \Pi_{i=2}^{m} \Pi_{j=2}^{n} P(x_{ij}|x_{(i-1)(j-1)}, x_{(i-1)j} x_{i(j-1)}). \end{aligned} \tag{3}$$

An important example of (3) is the class of 2-D models presented in [5], which is often referred to as Pickard random fields (PRF). The class of binary Markov random fields defined as stationary extension of the distribution on 2×2 elements was completely characterized in [6].

We shall present an algorithm for extending a stationary field defined on 2×2 elements to fields where the conditional probabilities are defined on blocks or vectors and show that iterative scaling may be used to ensure that the entropy increases (or rather that it does not decrease) when extending the field. In Section II, 2-D random fields defined on 2×2 elements are briefly presented. Iterative scaling, as we shall use it, is introduced in Section III. The extension of a field defined on 2×2 elements is introduced in Section IV along with an algorithm using iterative scaling to extend the field.

978-1-4244-4313-0/09/$25.00 ©2009 IEEE

Examples of applying the algorithm to 2-D constraints and calculating the entropy is presented for the hard square and the n.i.b constraint in Section V. Application to a simple stationary model with hidden states is outlined in Section VI.

II. Finite Two-dimensional Random Fields

Let A, B, C, D be random variables over $\mathcal{A}$ in a 2×2 rectangle with relative positions given by, $x_{ij}, x_{i,j+1}, x_{i+1,j}$ and $x_{i+1,j+1}$:

$$\begin{matrix} A & B \\ C & D \end{matrix}$$

Consider $\mu_{2\times 2}$ defined by the joint distribution $(ABCD)$ of the variables. Likewise, let for a subset of variables, eg. (ABC) denote the joint distribution of the variables.

We consider stationary measures where the probability distribution on 2×2 elements is given by $\mu_{2\times 2}$. Furthermore we require that two rows of the field are described by a Markov chain with transition probabilities along the two rows given by the conditional probability $P(BD|AC)$.

It may be seen that the measure in (3) has this property. The Pickard random field is a prominent example of such a distribution and (3) is described by the probability distribution of four symbols [5] [6]:

The probabilities of $(ABCD)$ are expressed by

$$P(ABCD) = P(D|ABC)P(ABC) \tag{4}$$

and the independence condition of (3)

$$P(ABC) = P(B|A)P(C|A)P(A). \tag{5}$$

For a stationary field this implies that each row has a Markov chain description [4], [5].

To ensure stationarity there are further conditions on $P(D|ABC)$ which we assume satisfied, and refer to [4], [5], [6] for details.

The entropy of the elements which are not part of the boundary ($i = 1$ or $j = 1$) is given by the conditional entropy $H(D|ABC)$. $H(\mu_F) \geq H(D|ABC)$ and as the size of the $m \times n$ rectangle increases, the per symbol entropy $H(\mu_F)$ converges to $H(D|ABC)$ [4] [7].

III. Iterative Scaling

In [7], [8] iterative scaling was used for the construction of a PRF for a 2-D constraint, by determining a solution for the joint distribution of two variables. Later we shall apply iterative scaling to extend a PRF, or another stationary model with measure as (3), defined on 2×2 elements to a model defined on more elements. Here iterative scaling is briefly introduced.

The problem of determining a joint distribution, say (BC) on two random variables B and C, such that each variable attains a given marginal distribution and combined satisfies a constraint expressed by a set of forbidden combinations of b and c, may be addressed using an iterative process known as iterative scaling [9].

The iterative scaling is given by a matrix $\{m_{ij}\}$, which is modified to get row sums r_i and column sums c_j. In each step, the rows are updated

$$m'_{ij} = m_{ij} * r_i / \sum_j (m_{ij}) \tag{6}$$

and subsequently the columns are updated

$$m'_{ij} = m_{ij} * c_j / \sum_i (m_{ij}) \tag{7}$$

If the linear equations have a non-negative solution, the iteration will converge to it. Formally, let $\mathcal{L}$ be the union of linear families of distributions, e.g. as r_i and c_j. Iterative scaling [9], e.g. by (6-7), may take an initial probability distribution, $Q(x)$ and for a non-empty class of distributions as $\mathcal{L}$, find a distribution, P^* which minimizes the divergence, $D(P^*||Q)$ for $P^* \in \mathcal{L}$, where the divergence is given by $D(P||Q) \equiv \sum P(x) \log(P(x)/Q(x))$.

Given a joint constraint, the matrix is initialized with zeros in positions that are not admissible by the constraint. Starting with a uniform initial distribution over the admissible configurations as $Q(x)$, implies that $Q(x)$ is constant for $P(x) > 0$. Thus minimizing the divergence, $D(P||Q)$, by the iterative scaling (6-7) will minimize $\sum P(x) \log(P(x))$ and thereby maximize the entropy, $H(P)$, over the distributions $P \in \mathcal{L}$. We shall refer to this as *maximum entropy iterative scaling* and apply it to find a joint distribution on two variables given the marginal distribution on each variable. This was used in [7], [8] to determine the joint conditional distribution $P(AD|B = b, C = c)$ for each pair $\{b, c\} \in \mathcal{A}^2$ as part of constructing a PRF, given (ABC) and (BCD), i.e. $P(A|B = b, C = c)$ and $P(D|B = b, C = c)$. Here the iterative scaling shall be used to construct an extended model given a stationary model defined for 2×2 elements by a distribution $(ABCD)$.

IV. Extended Block Fields

Compared to a model defined on 2×2 elements a model defined on larger blocks (or vectors) may be of interest as a model for block based coding, to increase the entropy or to approximate a distribution over areas larger than 2×2. In this work we consider starting from a stationary 2×2 model (3) and extending this to a (quasi-)stationary model defined on $2\times(n+2)$ elements. The quasi-stationarity refers to the fact that the variables of column j mod $(n+1)$ have identical distributions, not necessarily so for different values of j mod $(n + 1)$ and the same characteristic applies when we consider the column of the say upper-left element of rectangles of a given size. The approach implies relaxing the assumption that two consecutive rows form a Markov chain, but maintaining the property that each row forms a Markov chain and it involves a modification of the conditional distribution of one row given the previous row.

First we return to a Pickard random field over a finite alphabet, $\mathcal{A}$. (For simplicity we shall refer to PRF, but the results are applicable to stationary models with a measure of the form (3).) Consider two rows and K columns of this

Pickard random field. The measure $\mu_{2\times K}$ is described by a Markov chain (MC) on the two rows [5] defined by the distribution on the four basic variables, $(ABCD)$, defining the transition probabilities $P(BD|AC)$. Let A and C be the first elements of the two rows (Section II) and let B_i and D_i denote the following elements where i denotes the column index. Further let $B_1^n = B_1, ..., B_n$ and $D_1^n = D_1, ..., D_n$ denote the first n of these elements. We shall first consider the $2 \times (n+2)$ elements

$$\begin{array}{ccccc} A & B_1 & \dots & B_n & B_{n+1} \\ C & D_1 & \dots & D_n & D_{n+1} \end{array}$$

Given a Pickard random field, a new block based construction shall be introduced such that the stationary Markov chain distribution of single rows is maintained, but the conditional probability of one row conditioned on the previous row is changed in order to increase the entropy if possible. The construction is based on changing the conditional probability $P(D_1^{n+1}|AB_1^{n+1}C)$. The corresponding conditional entropy for $n+1$ symbols, $H(D_1^{n+1}|AB_1^{n+1}C)$, may be rewritten as

$$\begin{aligned} &H(D_1^{n+1}|AB_1^{n+1}C) = \\ &H(B_1^{n+1}D_1^{n+1}|AC) - H(B_1^{n+1}|AC) = \\ &H(B_1^nD_1^n|ACB_{n+1}D_{n+1}) + H(B_{n+1}D_{n+1}|AC) - \\ &H(B_1^{n+1}|AC) \end{aligned} \tag{8}$$

Given a PRF, the conditional probabilities above may all be obtained as marginals from the PRF distribution, $\mu_{2\times(n+2)}$, on $(AB_1^{n+1}CD_1^{n+1})$. From this point on, let $P_P()$ denote marginals obtained from the (initial) PRF, summing out the other variables.

The new distribution on $(AB_1^{n+1}CD_1^{n+1})$ shall be composed by

$$\begin{aligned} &P(AB_1^{n+1}CD_1^{n+1}) = \\ &P_P(AC)P_P(B_{n+1}D_{n+1}|AC)P^*(B_1^nD_1^n|AB_{n+1}CD_{n+1}), \end{aligned} \tag{9}$$

i.e. $P(B_1^nD_1^n|AB_{n+1}CD_{n+1})$ shall be modified.

Given a PRF, a quasi-stationary extended field is defined by 1) the boundary distribution is given by the PRF boundary distribution ($i = 1$ or $j = 1$ in (3)) and 2) the conditional distribution of the elements in columns, $k(n+1) \le j \le (k+1)(n+1)$ of any row, i, is independent of i and k and has the causal form

$$\begin{aligned} &P(D_1^{n+1}|AB_1^{n+1}C) = \\ &P_P(D_{n+1}|AB_1^{n+1}C)P^*(D_1^n|AB_1^{n+1}CD_{n+1}). \end{aligned} \tag{10}$$

where $P^*(D_1^n|AB_1^{n+1}CD_{n+1})$ is consistent with $P_P(D_1^n|AB_{n+1}CD_{n+1})$ and $P_P(B_1^n|AB_{n+1}CD_{n+1})$ in the sense that these probability distributions are identical to the corresponding marginals of the new distribution $(AB_1^{n+1}CD_1^{n+1})$ (9).

Iterative scaling [9] may be applied to find a joint distribution $P^*(B_1^nD_1^n|AB_{n+1}CD_{n+1})$ and thereby $P^*(D_1^n|AB_1^{n+1}CD_{n+1})$ with marginals equal to $P_P(B_1^n|AB_{n+1}CD_{n+1})$ and $P_P(D_1^n|AB_{n+1}CD_{n+1})$ leading to (quasi-)stationarity as follows. Let $b_1^n(i)$ and $d_1^n(j)$ denote configurations of B_1^n and D_1^n indexed by i and j, respectively. For each configuration $ab_{n+1}cd_{n+1}$, a matrix is defined with elements $m_{ij} = P^*(b_1^n(i)d_1^n(j)|ab_{n+1}cd_{n+1})$. ($m_{ij} = 0$ for forbidden configurations of $ab_1^{n+1}cd_1^{n+1}$.) The row and column sums for the iterative scaling (11-12) are given by

$$\begin{aligned} r_i = &\sum_{d_1^n} P^*(b_1^n(i)d_1^n|ab_{n+1}cd_{n+1}) = \\ &P_P(b_1^n(i)|ab_{n+1}cd_{n+1}), \end{aligned} \tag{11}$$

$$\begin{aligned} c_j = &\sum_{b_1^n} P^*(b_1^nd_1^n(j)|ab_{n+1}cd_{n+1}) = \\ &P_P(d_1^n(j)|ab_{n+1}cd_{n+1}). \end{aligned} \tag{12}$$

Given a PRF, the distribution $P_P(B_1^nD_1^n|AB_{n+1}CD_{n+1})$ obviously already provides a solution to the iterative scaling (11-12). The maximum entropy iterative scaling (6-7) with row and column sums given by (11-12) will increase $H(B_1^nD_1^n|ACB_{n+1}D_{n+1})$ in (8) compared to the PRF distribution if this is possible. This in turn will increase $H(D_1^{n+1}|AB_1^{n+1}C)$ as the last two terms of (8) are not changed by the construction.

The iterative scaling solution $P^*(B_1^nD_1^n|AB_{n+1}CD_{n+1})$ gives a modified distribution (9) on the $2 \times (n+2)$ elements and $P^*(D_1^n|AB_1^{n+1}CD_{n+1})$, which leads to the modified conditional probabilities (10). Based on this, the conditional entropy (8) may be rewritten as

$$\begin{aligned} &H(D_1^{n+1}|AB_1^{n+1}C) = \\ &H^*(D_1^n|AB_1^{n+1}CD_{n+1}) + H_P(D_{n+1}|AB_1^{n+1}C), \end{aligned} \tag{13}$$

where $H^*()$ and $H_P()$ denotes the entropies based on distributions defined by iterative scaling and the initial PRF, respectively.

The construction above may first be extended to $k(n+1)+1$ columns for any $k > 0$ in the two rows: Let B_{n+1} and D_{n+1} take the place of A and C and define a new distribution on the elements of the next $n+1$ columns, B_{n+2}^{2n+2} and D_{n+2}^{2n+2}. Thereafter repeat the construction with $B_{k(n+1)}$ and $D_{k(n+1)}$ taking the place of A and C. In each step, $2\times(n+1)$ elements are appended in the two rows. As each of the (individual) rows have maintained the distribution of the original stationary model, the construction may be repeated row by row conditioning the new row on the previous one.

Theorem 4.1: Given a Pickard random field over a finite alphabet, $\mathcal{A}$, an extended quasi-stationary field with probability distribution (10) exists. The entropy of $n+1$ elements of the interior, $x_{ij}\dots x_{i,j+n}, 1 < i, 1 < j$, is $H(D_1^{n+1}|AB_1^{n+1}C)$, and it is given by (13). Defining $P^*(D_1^n|AB_1^{n+1}CD_{n+1})$ (10) by maximum entropy iterative scaling, the entropy of the extended field is not less than the entropy of the PRF as,

$$\begin{aligned} &H(D_1^{n+1}|AB_1^{n+1}C) = \\ &H^*(D_1^n|ACB_1^{n+1}D_{n+1}) + H_P(D_{n+1}|AB_1^{n+1}C) \\ &\ge (n+1)H_P(D|ABC) \end{aligned} \tag{14}$$

and the increase in entropy is given by

$$H^*(D_1^n|ACB_1^{n+1}D_{n+1}) - H_p(D_1^n|ACB_1^{n+1}D_{n+1}). \quad (15)$$

Proof: Consider the distribution of the elements in columns $0 \leq j \leq K = k(n+1)$ of the first two rows, $AB_1^K CD_1^K$. Given a PRF, the distribution of the first row is by definition given by the Markov chain of the PRF.

The independence condition of the PRF (5) gives $P_P(D_1^{n+1}|AB_1^{n+k}C) = P_P(D_1^{n+1}|AB_1^{n+1}C), k \geq 1$. The two-row MC property of the PRF gives $P_P(D_{n+1}|AB_1^{n+k}C) = P_P(D_{n+1}|AB_1^{n+1}C), k \geq 1$ and $P_P(D_1^n|AB_1^{n+k}CD_{n+1}) = P_P(D_1^n|AB_1^{n+1}CD_{n+1}), k \geq 1$ and that the distribution $(AB_1^nCD_1^n|B_{n+1}D_{n+1})$ is independent of values of B_k in columns, $k \geq n+1$.

The distribution on $AB_{n+1}CD_{n+1}$ is given by the 2-row MC derived from the PRF probabilities, $P_P(AB_{n+1}CD_{n+1})$. In the iterative scaling, $P_P(D_1^n|AB_{n+1}CD_{n+1})$ defines the marginals in terms of the column sums (12), therefore the distribution of the extended field on $(AB_{n+1}CD_1^{n+1})$ is maintained and the marginal distribution on CD_1^{n+1} is still given by $P_P(CD_1^{n+1})$. We note that the elements AC and $B_{n+1}D_{n+1}$ maintain their (stationary) distributions given by the two row MC, but this is not necessarily the case for the elements in between, $B_jD_j, 1 \leq j \leq n$.

For the PRF 2-row MC, each state $B_{k(n+1)}D_{k(n+1)}$ separates the distributions of the elements before and after the state. By construction and the arguments above, this also applies to the extended field and $P_P(B_{k(n+1)}D_{k(n+1)})$ gives the distribution. Thus, repeating the construction (10) the so defined distribution on $B_{k(n+1)}^{(k+1)(n+1)}D_{k(n+1)}^{(k+1)(n+1)}$ will, for any $k \geq 0$, be identical to the distribution $(AB_1^{n+1}CD_1^{n+1})$. The distribution of the new row $(CD_1^{(K'+1)(n+1)})$ is identical to the distribution given by the PRF MC distribution on rows, i.e. $P_P(CD_1^{(K'+1)(n+1)})$. Thus the second row has the same MC distribution as the first row.

By induction, the subsequent rows will have the same MC distribution and the distribution on the elements in columns $k(n+1)$ to $(k+1)(n+1)$ in any two consecutive rows will be identical to $(AB_1^{n+1}CD_1^{n+1})$ as given by (9) and the iterative scaling. The conditional probability is given by $P(D_1^{n+1}|AB_1^{n+1}C) = P_P(D_{n+1}|AB_1^{n+1}C)P^*(D_1^n|AB_1^{n+1}CD_{n+1})$ and the entropy is given by (13). When the maximum entropy iterative scaling is applied, the entropy per symbol will not be reduced as the PRF distribution obviously provides a valid solution to the row and column sum constraints (11-12) defined by the PRF itself. ■

It may also be noted, that the columns with index $k(n+1)$ for integer $k \geq 0$ have the same distribution as given by (AC).

The construction as a whole may also be generated by first defining the distribution on columns $k(n+1), k \geq 0$, based on the PRF given by the distribution $(AB_{n+1}CD_{n+1})$. The distribution of the first column is given by $P_P(AC)$ as is the case for all the columns $k(n+1), k \geq 0$. The distribution of the elements of the first row with column index $k(n+1)$ are given by the Markov chain $P_P(B_{n+1}|A)$. After these elements of the first row and the first column the remaining elements of columns $k(n+1), k \geq 0$ are determined by $P_P(D_{n+1}|AB_{n+1}C)$. The distribution of the intermediate columns is sequentially given by $P_P(B_1^n|AB_1^{n+1}CD_{n+1})$ for the first row and thereafter by $P^*(D_1^n|AB_1^{n+1}CD_{n+1})$.

V. Numerical Results

Models for the hard-square constraint [4] and the n.i.b constraint [7] were extended using the presented technique based on iterative scaling. Given a PRF, calculating the marginals, $P_P(B_{n+1}D_{n+1}|AC)$, $P_P(B_1^n|ACB_{n+1}D_{n+1})$ and $P_P(D_1^n|ACB_{n+1}D_{n+1})$, may efficiently be implemented using a trellis structure for the 'hidden' variables, i.e. the variables of $(AB_1^{n+1}CD_1^{n+1})$, which are to be summed out. To summarize the important steps:

1) The given initial PRF distribution $(ABCD)$ gives the distribution of (AC) and $P(BD|AC)$ defining the two-row MC and the distribution $(AB_1^{n+1}CD_1^{n+1})$.
2) Calculate $P_P(B_1^n|ACB_{n+1}D_{n+1})$ and $P_P(D_1^n|ACB_{n+1}D_{n+1})$ to define the iterative scaling (6-7).
3) Calculate $P^*(B_1^nD_1^n|AB_{n+1}CD_{n+1})$ by iterative scaling and
4) $P^*(D_1^n|AB_1^{n+1}CD_{n+1}) = P^*(B_1^nD_1^n|AB_{n+1}CD_{n+1})/P_P(B_1^n|ACB_{n+1}D_{n+1})$
5) $P_P(B_{n+1}D_{n+1}|AC)$ is also obtained from $(AB_1^{n+1}CD_1^{n+1})$
6) Calculate $H(D_1^{n+1}|AB_1^{n+1}C)$ by (13).

A. The hard-square constraint

A stationary model defined on 2×2 elements with probability measure of the form (3) was defined for the hard-square constraint in [4]. The basic elements of the second line (CD) are shifted one position to the left relative to A and B,

$$\begin{array}{ccc} - & A & B \\ C & D & - \end{array}$$

Thus the directions of the Markov chains are horizontal $(P(B|A))$ and diagonal (top-right to bottom-left) (given by $P(C|A)$). The model has two parameters [4], with optimal values w.r.t the entropy $(H(D|ABC) = 0.58727)$, $P(D = 0|A = 0, B = 0, C = 0) = 0.671833$ and $P(D = 0|A = 0, B = 1, C = 0) = 0.566932$. Starting with this model an extended model was determined, increasing entropy for increasing n. For $n = 12$ the extended model gave $H(D_1^{n+1}|AB_1^{n+1}C) = 0.5876$ per binary symbol. In [10] a lower bound of 0.58786 was presented based on combining PRF with tiling. Using the Calkin and Wilf bounds the capacity is $C \approx 0.58789$.

B. The n.i.b. constraint

The extended block construction was applied to the n.i.b. constraint with super symbols defined on 1×2 blocks along diagonals as in Fig. 1. The PRF solution presented in [7] was used as the starting point.

For $n = 4$ starting with the optimal PRF of [7], gave $H(D_1^{n+1}|AB_1^{n+1}C) = 0.9204$ bits per binary symbol. For

$$\begin{array}{c} c_1c_2 \\ a_1a_2d_1d_2 \\ b_1b_2 \end{array}$$

Fig. 1. The four 1×2 blocks of the PRF model for the n.i.b constraint.

fixed extended block size given by n, the four parameters of the PRF may be modified to increase $H(D_1^{n+1}|AB_1^{n+1}C)$. For $n = 4$ and using a gradient search over the PRF parameters, a maximum of $H = 0.9205$ bits per binary symbol was achieved. The PRF parameters were $P(01) = P(10) = 0.2260, P(00|00) = 0.2268, P(01|00) = 0.2485, P(00|11) = 0.3299$. For smaller values of n, the following entropies were obtained, $n = 1 : H = 0.9173$, $n = 2 : H = 0.9190$ and $n = 3 : H = 0.9199$ bits per binary symbol. In this example the iterative scaling was used twice, first to derive the intial PRF [7] and thereafter for the extension. In [11], a lower bound of 0.9226 was presented for the constraint.

In the two examples above, the absolute increase of entropy by the extended field based on maximum entropy iterative scaling is modest, but in fact the initial models are very good and a major part of the gap between the entropy for the initial models and the (bounded or estimated) capacity value is closed by the extension.

In the examples above the maximum entropy iterative scaling was used for the considered constraint, increasing the entropy. Some interesting alternative applications are also possible. The constraint in the iterative scaling could be different, so the initial model could be an approximation to the desired constraint, which is not enforced until the iterative scaling, which in this case will provide a solution if there is one. Another application would be to start with a desired distribution $\mu_{2\times(n+2)}$ and initialize the distribution Q of the iterative scaling with this, which would then return the best approximation in terms of the minimum divergence. In both of these alternative applications the method could address or incorporate structures defined over more elements than the 2×2 elements of the PRF.

VI. Models with Hidden States

In the stationary models considered so far, independence of B and C given A (5) has been assumed. Here we consider a broader class of stationary models where 2 rows are described by a MC and one row is described conditioned on the previous row [12]. An example is given by a two-row Markov chain, $P(BD|AC)$, which is symmetric: $P(B = b, D = d|A = a, C = c) = P(B = a, D = c|A = b, C = d)$. Thus the two rows will have identical (stationary) distributions. Only considering one row, the process is a function of a Markov chain and this process may be described using a forward and a backward pass similar to a hidden Markov process [12] if the independence (5) is not satisfied.

The iterative scaling (12) may be applied based on probabilities derived from the 2-row MC, which defines $\mu_{2\times(2n+2)}$ from which the relevant terms may be derived as before and outlined in Steps 1-5 in Section V. The reason is that considering the two-rows, (B_{n+1}, D_{n+1}) will again separate the values of $(AB_1^{n+1}CD_1^{n+1})$ from the rest of the 2-row MC, i.e. $B_k, D_k, k > n + 1$.

The difference comes when calculating $P(D_{n+1}|AB_1^KC)$ where the lack of independence (5) means that the conditional probability of D_{n+1} may depend on the rest of the row above, $\{B_k|k > n + 1\}$. This dependency may efficiently be handled by a backward pass using the two-row MC in the reverse direction and treating the variables D_k as hidden variables.

The entropy is

$$H(D_1^{n+1}|AB_1^{n+1}C) = \\ H^*(D_1^n|AB_1^{n+1}CD_{n+1}) + H_h(D_{n+1}|AB_1^KC), \quad (16)$$

where $H^*()$ and $H_h()$ denotes the entropies based on distributions defined by iterative scaling and the initial hidden Markov model derived from the 2-row MC, respectively. It will in general not be possible to calculate the exact value of $H_h()$, but the entropy of a function of a Markov chain may be bounded from above and below [12]. The increase of entropy by the iterative scaling is given by $H^*(D_1^n|AB_1^{n+1}CD_{n+1}) - H_h(D_1^n|AB_1^{n+1}CD_{n+1})$, which may be calculated though.

VII. Discussion

It may be noted that the procedure does not necessarily provide the maximum possible value of $H(D_1^{n+1}|AB_1^{n+1}C)$ for a given 2-D constraint, but searching over the full parameter space is not tractable. The advantage of using iterative scaling is that a good choice of values is automatically obtained for the large set of parameters.

References

[1] S. Friedland, "On the entropy of Z^d subshifts of finite type," *Linear algebra and its applications*, vol. 252 pp. 199–220, 1997.

[2] J. Justesen and Yu. M. Shtarkov, "The combinatorial entropy of images," *Probl. Pered. Inform.*, (in Russian) vol. 33, pp. 3-11, 1997.

[3] A. Kato and K. Zeger, "On the capacity of two-dimensional run-length constrained channels," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1527-1540, July 1999.

[4] R. M. Roth, P. H. Siegel, and J. K. Wolf, "Efficient coding scheme for the hard-square model," *IEEE Trans. Inform. Theory*, vol. 47, no. 3, pp. 1166-1176, March 2001.

[5] D. Pickard, "Unilateral Markov fields," *Adv. Appl. Probability*, vol. 12, pp. 655–671, 1980.

[6] F. Champagnat, J. Idier, and Y. Goussard, "Stationary Markov random fields on a finite rectangular lattice," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2901–2916, Nov. 1998.

[7] S. Forchhammer and T.V. Laursen, "A model for the two-dimensional no isolated bits constraint," in *Proc. ISIT*, Chicago, pp. 1189–1193, 2006.

[8] S. Forchhammer and J. Justesen, "Block Pickard models for two-dimensional constraints," submitted to *IEEE Trans. Inform. Theory*.

[9] I. Csiszár and P.C. Shields, "Information theory and statistics: A tutorial," in *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 4, pp. 422–527, 2004.

[10] A. Sharov and R.M. Roth, "Two-dimensional constrained coding based on tiling," in *Proc. ISIT*, Toronto, Canada, pp. 1468–1472, 2008.

[11] I. Tal and R.M. Roth, "Bounds on the rate of 2-D bit-stuffing encoders," in *Proc. ISIT*, Toronto, Canada, pp. 1463–1467, 2008.

[12] S. Forchhammer and J. Justesen, "Entropy bounds for constrained two-dimensional random fields," *IEEE Trans. Inform. Theory*, vol. 45, pp. 118–127, Jan. 1999.