Generalized notions of block symmetry for discrete memoryless channels

Flemming Topsøe
University of Copenhagen, Department of Mathematics

Introduction

Notions of *block symmetry* for discrete, memoryless channels were introduced in Pedersen and Topsøe [2] and some basic results concerning capacity and optimal distributions developed.

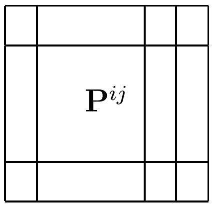

Figure 1

We generalize the results of [2] in two directions: Firstly we note that by adding a row to a transition matrix of a DMC, typical notions of symmetry are destroyed and yet, it is just as easy to find optimal (capacity achieving) distributions for the new DMC as for the original DMC. This problem is taken care of by extending the notions of symmetry to notions allowing what we shall call "connections".

Furthermore, we note that the results of [2] are concerned with the search for optimal distributions which are mixtures of uniform distributions. We generalize by searching for optimal distributions which have prescribed conditional distributions with conditioning understood in the natural sense in relation to the block decomposition given.

Main result of [2]

Let us first quote the main result of [2]. Let $\mathbf{P}$ be the transition matrix for a discrete memoryless channel (DMC) and consider a block decomposition $(\mathbf{P}^{ij})_{i,j}$ of $\mathbf{P}$ as indicated in Figure 1. Such a decomposition is induced by two decompositions, one of the input-, the other of the output alphabet. Assume that, within each block $\mathbf{P}^{ij}$, all row sums are equal and all column sums are equal. Assume further, that rows in the full matrix $\mathbf{P}$ which correspond to equivalent input letters have equal entropy. Then there exists an optimal input distribution which is consistent with the decomposition of the input alphabet in the sense that equivalent input letters are sent with equal probability. Furthermore, the optimal output distribution is consistent in a similar way, hence equivalent output letters are received with equal probability. This is the main result of [2]. In [2] examples are given to illustrate that the result goes beyond known results as contained in standard textbooks.

Coverings, decompositions

For any set Z, we denote by $\mathrm{DEC}(Z)$ the set of decompositions of Z and by $\mathrm{COV}(Z)$ the set of coverings of Z. For $\eta_1,\, \eta_2 \in \mathrm{DEC}(Z)$ we write $\eta_1 \leq \eta_2$ if every η_1-set is a union of η_2-sets. $\mathrm{DEC}(Z)$ is a lattice in this ordering. The smallest and largest elements of $\mathrm{DEC}(Z)$ we denote by $o(Z)$, respectively $e(Z)$. Thus $o(Z) = \{Z\}$ and $e(Z) = \{\{z\} \mid z \in Z\}$.

Every covering ξ of Z induces a decomposition of Z, defined via a corresponding equivalence relation, called *ξ-connection.* Here, $z \in Z$ and $z' \in Z$ are *ξ-connected* if there exists $n \in \mathbf{N}$ and sets $A_1, \cdots, A_n$ in ξ such that $z \in A_1$, $z' \in A_n$ and $A_i \cap A_{i+1} \neq \emptyset$ for $i = 1, \cdots, n-1$. The decomposition thus defined is denoted $\mathrm{fr}(\xi)$ and called the *frame of* ξ.

Consider finite sets X and Y (later on taken as input- and output alphabets, respectively) and put $Z = X \times Y$. A *block decomposition* of Z is a decomposition of Z of the form $\eta_X \times \eta_Y = \{A \times B \mid A \in \eta_X\,,\, B \in \eta_Y\}$ with $\eta_X \in \mathrm{DEC}(X)$, and $\eta_Y \in \mathrm{DEC}(Y)$. The set of block decompositions of Z is denoted $\mathrm{BDE}(Z)$. This set is a sublattice of $\mathrm{DEC}(Z)$.

Block symmetric decompositions

The natural projections on $Z = X \times Y$ are denoted $\pi_X : Z \to X$ and $\pi_Y : Z \to Y$. By a *horizontal section*, respectively a *vertical section*, we understand a subset $A \subseteq Z$ such that $\pi_X(A)$, respectively $\pi_Y(A)$ is a singleton. Let τ_{-} ($\tau_{|}$) be an equivalence relation on the set of horizontal (vertical) sections and asume that if two such sections are equivalent, they have the same projection on Y (on X). We use τ_{+} as a common symbol for τ_{-} and $\tau_{|}$ and consider the subset $\mathrm{BSD}(Z;\tau_{+})$ of $\mathrm{BDE}(Z)$ consisting of those $\eta \in \mathrm{BDE}(Z)$ for which every block $A \times B \in \eta$ satisfies:

$$x_1 \in A \,\wedge\, x_2 \in A \Rightarrow \{x_1\} \times B \equiv_{\tau_{-}} \{x_2\} \times B \tag{1}$$

$$y_1 \in B \,\wedge\, y_2 \in B \Rightarrow A \times \{y_1\} \equiv_{\tau_{|}} A \times \{y_2\}\,. \tag{2}$$

In other words, the requirement to η is that within any η-block, all horizontal sections are quivalent and also, all vertical sections are equivalent.

We call $\mathrm{BSD}(Z;\tau_{+})$ the *symmetry type* specified by τ_{+} and if $\eta \in \mathrm{BSD}(Z;\tau_{+})$, we say that η is a *block symmetric decomposition* of Z w.r.t. τ_{+}. The qualifying term “symmetric” will be justified by the intended applications where the equivalences τ_{-} and $\tau_{|}$ will signal some kind of regular pattern or “symmetry”, typically defined with reference to a given fixed DMC.

τ_- is *composition stable* if, for $x_1 \in X$, $x_2 \in X$ and for any family $(B_j)_{j\in J}$ of pairwise disjoint subsets of Y,

$$\forall\, j \in J : \{x_1\} \times B_j \equiv_{\tau_-} \{x_2\} \times B_j \Rightarrow \{x_1\} \times \bigcup_{j\in J} B_j \equiv_{\tau_-} \{x_2\} \times \bigcup_{j\in J} B_j \,.$$

If the analogous relation (with vertical, rather than horizontal sections) holds for $\tau_{|}$, then $\tau_{|}$ is *composition stable.*

Lemma 1. *Let τ_- and $\tau_{|}$ be composition stable equivalence relations as described above. Then $e(Z) \in \mathrm{BSD}(Z;\tau_+)$, and $\mathrm{BSD}(Z;\tau_+)$ is closed under the lattice operation $\wedge$. In particular, $\mathrm{BSD}(Z;\tau_+)$ contains a coarsest block decomposition.*

We shall not give the proof here. We add that under suitable conditions one can device an efficient algorithm which allows the construction of the coarsest block symmetric decomposition.

We consider $\mathrm{BSD}(Z;\tau_+)$ to be a *notion of block symmetry.* In general, such a notion, say κ, is nothing but a subset of $\mathrm{BDE}(Z)$. If κ contains a coarsest element, we call this element the *profile* of κ and denote it by $[\kappa]$. In this terminology, Lemma 1 asserts that $[\mathrm{BSD}(Z;\tau_+)]$ exists under the conditions stated regarding τ_- and $\tau_{|}$.

Allowing for connections

Let, again, τ_- and $\tau_{|}$ be equivalence relations on the horizontal, respectively the vertical sections of Z. See p.6 for an important choice of τ_- and $\tau_{|}$.

Occasionally we find it appropriate to extend the notion $\mathrm{BSD}(Z;\tau_+)$ in a way which is "asymmetric" in that horizontal notions and vertical notions are treated differently. We denote the new type $\mathrm{BDE}^*(Z;\tau_+)$ and call it the *notion of block symmetry specified by τ_+, allowing for connections* (more suggestive: allowing for *vertical* connections). Let $\eta = \eta_X \times \eta_Y \in \mathrm{BDE}(Z)$. In order to decide if $\eta \in \mathrm{BSD}^*(Z;\tau_+)$ we first consider $\xi \subseteq \mathcal{P}(X)$ which consists of those subsets $A \subseteq X$ which are contained in an η_X-equivalence class and, in addition satisfy the requirement that $\{A\} \times \eta_Y \in \mathrm{BSD}(A \times Y;\tau_+)$. By definition, $\eta \in \mathrm{BSD}^*(Z;\tau_+)$ if $\xi \in \mathrm{COV}(X)$ and $\mathrm{fr}(\xi) = \eta_X$.

Clearly, $\mathrm{BDE}^*(Z;\tau_+) \supseteq \mathrm{BSD}(Z;\tau_+)$ hence, if the profile $[\mathrm{BDE}^*(Z;\tau_+)]$ exists, it is coarser than $[\mathrm{BSD}(Z;\tau_+)]$. However, we know of no result similar to that of Lemma 1 which tells us that, under suitable conditions on τ_- and $\tau_{|}$, the profile defined exists when we allow for connections.

Capacity under benefit (or cost)

For the remainder of the manuscript there is also given a DMC, characterized by the *transition matrix* $\mathbf{P}$ over $X \times Y$.

The *capacity* $C(\mathbf{P})$ is the supremum of $I(\vec{p})$, the *information transmission rate*, over all input distributions. With our finiteness assumption, the supremum is attained. Any input distribution with $I(\vec{p}) = C$ is said to be *optimal.* The *optimal output distribution* is the distribution – known to be unique – induced by an optimal input distribution.

We need a refined notion of capacity, taking into regard that the sending of an input symbol may be associated with a certain benefit. This idea, and the basic result connected with it, has been considered before, cf. Blahut [1] Theorem 9 (where it was found more natural to associate an "expense" or "cost" , rather than a "benefit", with the input symbols sent).

To a given transition matrix $\mathbf{P}$ and a given *benefit function* $\mathbf{a} : x \curvearrowright a_x$ which maps X into the reals, we consider the *modified capacity with benefit* $\mathbf{a}$, $C(\mathbf{P}, \mathbf{a})$, defined via the *information transmission rate*, $I(\vec{p})$, by

$$C(\mathbf{P}; \mathbf{a}) = \sup_{\vec{p}} (I(\vec{p}) + \langle \mathbf{a}, \vec{p} \rangle) \,. \tag{3}$$

Here, $\langle \mathbf{a}, \vec{p} \rangle = \sum_x p_x a_x$. Clearly, the supremum is attained, and we are led to consider optimal input- and output distributions for the modified problem, thereby generalizing the usual concepts (which correspond to the case with zero benefit).

Lemma 2. *Let $\vec{p}^*$ be an input distribution and denote by $\vec{q}^*$ the induced output distribution. A necessary and sufficient condition that $\vec{p}^*$ be an optimal input distribution for the modified problem with benefit $\mathbf{a}$ is that, for some constant C, the following two conditions hold:*

$$D(\vec{q}_x \| \vec{q}^*) + a_x \leq C \quad \text{for all } x \tag{4}$$

$$D(\vec{q}_x \| \vec{q}^*) + a_x = C \quad \text{for all } x \quad \text{with } p_x^* > 0 \,. \tag{5}$$

If these conditions are satisfied, C is the modified capacity: $C = C(\mathbf{P}; \mathbf{a})$.

For a proof, see [1] or [2].

Explicit formulas for calculation of optimal distributions and modified (or usual) capacity does not exist in general, and even when they do, they get complicated. Formulas for a 2×2 transition matrix can be found in [2].

The derived DMC

Let $\mathbf{P} = (p_{xy})_{x \in X, y \in Y}$ and $\eta = \eta_X \times \eta_Y \in \mathrm{BDE}(X \times Y)$. The number of classes in η_X and η_Y are denoted M, respectively N. We put $\eta_X = \{X_i \mid i \leq M\}$ and $\eta_Y = \{Y_j \mid j \leq N\}$. Let $\mathbf{P}^{ij}$ be the ij'th block in $\mathbf{P}$, i.e. $\mathbf{P}^{ij} = (p_{xy})_{x \in X_i, j \in Y_j}$. The row vectors in $\mathbf{P}^{ij}$ are denoted $\vec{q}_x^{\,ij}; x \in X_i$ and the sums of the elements of these row vectors are denoted $\sigma_x^{ij}; x \in X_i$. If $\sigma_x^{ij} > 0$, we denote by ${}^*\vec{q}_x^{\,ij}$, the probability vector obtained by normalization of $\vec{q}_x^{\,ij}$.

In case the σ_x^{ij} are all equal when x ranges over X_i, the common value is denoted σ_-^{ij}.

As the block decomposition η is seen in relation to $\mathbf{P}$, we write $\eta \in \mathrm{BDE}(\mathbf{P})$. We write $\eta \in BDE(\mathbf{P};\sigma_-)$ if, within each block $\mathbf{P}^{ij}$, the row sums are equal, i.e. if all the σ_-^{ij} are well defined. If $\eta \in \mathrm{BDE}(\mathbf{P};\sigma_-)$, we define the *derived DMC* as the *DMC* with transition matrix $\partial_\eta \mathbf{P} = (\sigma_-^{ij})_{i\le M, j\le N}$. The i'th row vector in $\partial_\eta \mathbf{P}$ is denoted $\vec{\sigma}^i$.

If $\mathbf{a} = (a_i)_{i\le M}$ are certain benefits, we denote by $C(\partial_\eta \mathbf{P};\mathbf{a})$ the capacity of $\partial_\eta \mathbf{P}$ with benefits given by the a_i's. By choosing the benefits suitably, we can in some cases relate this capacity to the capacity of the original DMC.

Prescribing conditionals, the matching condition

We assume that we are given two systems of probability vectors $(\vec{u}^i)_{i\le M}$ and $(\vec{v}^j)_{j\le N}$. For $i \le M, \vec{u}^i = (u_x^i)_{x\in X_i}$ is assummed to define a probability distribution over X_i with positive point probabilities: $u_x^i > 0$ for $x \in X_i$, and for $j \le N$, $\vec{v}^j = (v_y^j)_{y\in Y_j}$ is, similarly, supposed to satisfy $v_y^j > 0$ for $y \in Y_j$. We refer to $(\vec{u}^i)_{i\le M}$ and $(\vec{v}^j)$ as the *prescribed* or *given conditionals*.

Consider a $\mathbf{P}$-input distribution $\vec{p} = (p_x)_{x\in X}$. Note that $\vec{p}$ has the $\vec{u}^i$ as conditionals – by which we mean that for each $i \le M$ with $\sum_{x\in X_i} p_x > 0$, the conditional distribution given X_i equals $\vec{u}^i$ – if and only if $\vec{p}$ is a mixture of the $\vec{u}^i$, i.e. $\vec{p} = \sum_{i\le M} \alpha_i \vec{u}^i$ for some probability vector $\vec{\alpha} = (\alpha_i)_{i\le M}$. Similarly, a $\mathbf{P}$-output distribution $\vec{q} = (q_y)_{y\in Y}$ has the $\vec{v}$ as conditionals if and only if it is a mixture of the form $\vec{q} = \sum_{j\in N} \beta_j \vec{v}^j$ for some probability vector $\vec{\beta} = (\beta_j)_{j\le N}$. Also note that when $\vec{p}$ and $\vec{q}$ are of the form considered, we find that $\partial_{\eta_X} \vec{p} = \vec{\alpha}$ and $\partial_{\eta_Y} \vec{q} = \vec{\beta}$, applying a natural notation for datareduction of $\vec{p}$, respectively $\vec{q}$.

We say that the *matching condition* holds, if every $\mathbf{P}$-input distribution with the $\vec{u}^i$ as conditionals induces a $\mathbf{P}$-output distribution with the $\vec{v}^j$ as conditionals. Easy inspection shows what this requirement amounts to:

Lemma 3. *The matching condition holds if and only if, for every $i \le M$ and $j \le N$, the sum*

$$\sum_{x\in X_i} \frac{u_x^i}{v_y^j} p_{xy}$$

is independent of y for $y \in Y_j$. If so, if $\eta \in \mathrm{BDE}(\mathbf{P};\sigma_-)$ then $\vec{\beta}$ is the $\partial_\eta \mathbf{P}$-output distribution induced by the $\partial_\eta \mathbf{P}$-input distribution $\vec{\alpha}$.

A basic result

In the result below, $D(\cdot\|\cdot)$ denotes Kullback-Leibler divergence.

Theorem 1. *Let η, $\mathbf{P}$ and conditionals $(\vec{u}^i)_{i\le M}$ and $(\vec{v}^j)_{j\le N}$ be given and assume that $\eta \in \mathrm{BDE}(\mathbf{P};\sigma_-)$ and that the matching condition holds.*

If, for every $i \le M$, the expression

$$a_i = \sum_{j \le N} \sigma_-^{ij} D\left({}^*\vec{q}_x^{\,ij} \| \vec{v}^j\right) \tag{6}$$

is independent of x for $x \in X_i$ then there exists an optimal $\mathbf{P}$-input distribution with the $\vec{u}$ as conditionals, and the optimal $\mathbf{P}$-output distribution has the $\vec{v}^j$ as conditionals. Furthermore, $C(\mathbf{P}) = C(\partial_\eta \mathbf{P}; \mathbf{a})$ where the benefits $\mathbf{a} = (a_i)_{i \le M}$ are given by (6)*, and for any $\vec{\alpha} = (\alpha_i)_{i \le M}$, $\vec{p} = \sum_{i \le M} \alpha_i \vec{u}^i$ is an optimal $\mathbf{P}$-input distribution if and only if $\vec{\alpha}$ is an optimal $\partial_\eta \mathbf{P}$-input distribution for the modified problem with benefits given by* (6)*.*

The proof is based on lemmas 2 and 3.

A RESULT ON BLOCK SYMMETRY UNDER CONNECTIONS

Theorem 1 can be used to derive various concrete results. We consider the case $\mathrm{BSD}^*(Z; \tau_+)$ where $\mathrm{BSD}(Z; \tau_+)$ corresponds to the situation in the main result of [2] cited on p.1-2. In other words, $\eta_X \times \eta_Y \in \mathrm{BSD}((Z; \tau_+)$ means that within blocks row-sums are equal and column sums are equal and, furthermore, full rows with equivalent input letters have identical entropies.

Theorem 2. *Assume that $\eta \in \mathrm{BSD}^*(Z; \tau_+)$ in the situation indicated above. Then the condition of Theorem 1 holds. Let $\vec{\alpha}$ and $\vec{\beta}$ be, respectively an optimal input-, and the optimal output distribution for the capacity optimization problem pertaining to $\partial_\eta \mathbf{P}$ with benefits as prescribed by (6). Then, the distribution $\Sigma_{j \le N} \beta_j \vec{v}_j$, is the optimal output distribution for $\mathbf{P}$.*

References

[1] R. E. Blahut. Computation of channel capacity and rate-distortion functions. *IEEE Trans. Inform. Theory*, 18:460–473, 1972.

[2] Jakob Bøje Pedersen and Flemming Topsøe. Block symmetry in discrete memoryless channels. In *Proceedings of 2002 IEEE Information Theory Workshop, Bangalore*, pages 131–134, 2002.