

Article

Non-linear Information Inequalities

Terence Chan * and Alex Grant

Institute for Telecommunications Research, University of South Australia, Australia. E-mails: {terence.chan, alex.grant}@unisa.edu.au

* Author to whom correspondence should be addressed.

Received: 24 May 2008 / Accepted: 9 December 2008 / Published: 22 December 2008

Abstract: We construct non-linear information inequalities from Matúš' infinite series of linear information inequalities. Each single non-linear inequality is sufficiently strong to prove that the closure of the set of all entropy functions is not polyhedral for four or more random variables, a fact that was already established using the series of linear inequalities. To the best of our knowledge, they are the first non-trivial examples of non-linear information inequalities.

Keywords: Entropy, entropy function, nonlinear information inequality, nonshannon type information inequality

1. Introduction

Information inequalities play a crucial role in the proofs for almost all source and channel coding converse theorems. Roughly speaking, these inequalities govern the impossibility in information theory. Among information inequalities discovered to date, the most well-known are the Shannon-type inequalities, including the non-negativity of (conditional) entropies and (conditional) mutual information. In [2], a non-Shannon information inequality (that cannot be deduced from any set of Shannon-type inequalities) involving more than three random variables was discovered. Since then, many additional information inequalities have been discovered [4].

Apart from their application in proving converse coding theorems, information inequalities (either linear or non-linear) were shown to have a very close relation with inequalities involving the cardinality of a group and its subgroups [3]. Specifically, an information inequality is valid if and only if its group-theoretic counterpart (obtained by mechanical substitution of symbols) is also valid. For example, the

non-negativity of mutual information is equivalent to the group inequality $|G||G_1 \cap G_2| \geq |G_1||G_2|$, where G_1 and G_2 are subgroups of the group G .

Information inequalities are also the most common tool (perhaps even unique), for the characterization of entropy functions (see Definition 1 below). In fact, entropy functions and information inequalities are two sides of the same coin. A complete characterization for entropy functions requires complete knowledge of the set of all information inequalities.

The set of entropy functions involving n random variables, Γ_n^* , and its closure $\bar{\Gamma}_n^*$ are of extreme importance not only because of their relation to information inequalities [6], but also for determination of the set of feasible multicast rates in communication networks employing network coding [5, 7]. Furthermore, determination of Γ^* would resolve the implication problem of conditional independence (determination of every other conditional independence relation implied by a given set of conditional independence relationships). A simple and explicit characterization of Γ^* , and $\bar{\Gamma}^*$ will indeed be very useful. Unfortunately, except in the case when $n < 4$, such a characterization is still missing [1, 2, 4].

Recently, it was shown by Matúš that there are countably infinite many information inequalities [1]. This result, summarized below in Section 2, implies that $\bar{\Gamma}_n^*$ is not polyhedral. The main result of this paper is *non-linear* inequalities, which we derive from Matúš' series in Section 3. To the best of our knowledge this is the first example of a non-trivial non-linear information inequality. We use the non-linear inequality to deduce that the closure of the set of all entropy functions is not polyhedral – a fact previously proved in [1] using the infinite sequence of linear inequalities. Finally, in Section 4, we compare the series of linear inequalities and the proposed nonlinear inequality on a projection of $\bar{\Gamma}_n^*$.

2. Background

Let the index set $N = \{1, 2, \dots, n\}$ induce a real 2^n dimensional Euclidean space \mathcal{F}_n with coordinates indexed by the set of all subsets of N . Specifically, if $\mathbf{g} \in \mathcal{F}_n$, then its coordinates are denoted $(\mathbf{g}(\alpha) : \alpha \subseteq N)$. Consequently, points $\mathbf{g} \in \mathcal{F}_n$ can be regarded as functions $\mathbf{g} : 2^N \mapsto \mathbb{R}$. The focus of this paper is the subset of \mathcal{F}_n corresponding to (almost) entropic functions.

Definition 1 (Entropic function) A function $\mathbf{g} \in \mathcal{F}_n$ is entropic if $\mathbf{g}(\emptyset) = 0$ and there exists discrete random variables X_1, \dots, X_n such that the joint entropy of $\{X_i : i \in \alpha\}$ is $\mathbf{g}(\alpha)$ for all $\emptyset \neq \alpha \subseteq N$. Furthermore, \mathbf{g} is almost entropic if it is the limit of a sequence of entropic functions.

Let Γ_n^* be the set of all entropic functions. Its closure $\bar{\Gamma}_n^*$ (i.e., the set of all almost entropic functions) is well-known to be a closed, convex cone [6]. An important recent result with significant implications for $\bar{\Gamma}_n^*$ is the series of linear information inequalities obtained by Matúš [1] (restated below in Theorem 1). Using this series, $\bar{\Gamma}_n^*$ was proved to be non-polyhedral for $n \geq 4$. This means $\bar{\Gamma}_n^*$ cannot be defined by an intersection of any finite set of linear information inequalities.

Following [1], we will use the following notational conventions. Specific subsets of N will be denoted by concatenation of elements, e.g. 123 will be written for $\{1, 2, 3\}$. For any $\mathbf{g} \in \mathcal{F}_n$ and sets $I, J \subseteq N$,

define

$$\begin{aligned} \Delta_{I,J} \mathbf{g} &\triangleq \mathbf{g}(I) + \mathbf{g}(J) - \mathbf{g}(I \cup J) + \mathbf{g}(I \cap J) \\ \square_{12,34} \mathbf{g} &\triangleq \mathbf{g}(13) + \mathbf{g}(23) + \mathbf{g}(14) + \mathbf{g}(24) + \mathbf{g}(34) \\ &\quad - \mathbf{g}(12) - \mathbf{g}(3) - \mathbf{g}(4) - \mathbf{g}(134) - \mathbf{g}(234). \end{aligned}$$

Furthermore, for singletons $i, j, k \in N$, write $\Delta_{ij|k}$ as shorthand for $\Delta_{ik,jk}$.

Theorem 1 (Matúš) *Let $s \in \mathbb{Z}^+$, the set of positive integers, and $\mathbf{g} \in \Gamma_n^*$ be the entropy function of discrete random variables $\{X_1, \dots, X_n\}$. Then*

$$s (\square_{12,34} \mathbf{g} + \Delta_{34|5} \mathbf{g} + \Delta_{45|3} \mathbf{g}) + \Delta_{35|4} \mathbf{g} + \frac{s(s-1)}{2} (\Delta_{24|3} \mathbf{g} + \Delta_{34|2} \mathbf{g}) \geq 0. \tag{1}$$

Furthermore, assuming that $X_5 = X_2$, the inequality reduces to

$$s (\square_{12,34} \mathbf{g} + \Delta_{34|2} \mathbf{g} + \Delta_{24|3} \mathbf{g}) + \Delta_{23|4} \mathbf{g} + \frac{s(s-1)}{2} (\Delta_{24|3} \mathbf{g} + \Delta_{34|2} \mathbf{g}) \geq 0. \tag{2}$$

To the best of our knowledge, this is the only result indicating the existence of infinitely many linear information inequalities. Reductions to $\bar{\Gamma}_4^*$ with $s = 1$ recovers the Zhang-Yeung inequality [2] and $s = 2$ obtains an inequality of [4].

3. Main Results

3.1. Non-linear information inequalities

The series of information inequalities given in Theorem 1 are all “quadratic” in the parameter $s \in \mathbb{Z}^+$,

$$Q(s; a(\mathbf{g}), b(\mathbf{g}), c(\mathbf{g})) \triangleq sb(\mathbf{g}) + c(\mathbf{g}) + s(s-1)a(\mathbf{g}) \geq 0$$

or equivalently

$$s^2 a(\mathbf{g}) + s(b(\mathbf{g}) - a(\mathbf{g})) + c(\mathbf{g}) \geq 0, \tag{3}$$

where in the first series of inequalities (1)

$$\begin{aligned} a(\mathbf{g}) &\triangleq \frac{1}{2} (\Delta_{24|3} \mathbf{g} + \Delta_{34|2} \mathbf{g}) \\ b(\mathbf{g}) &\triangleq \square_{12,34} \mathbf{g} + \Delta_{34|5} \mathbf{g} + \Delta_{45|3} \mathbf{g} \\ c(\mathbf{g}) &\triangleq \Delta_{35|4} \mathbf{g} \end{aligned} \tag{4}$$

and in the second series of inequalities (2)

$$\begin{aligned} a(\mathbf{g}) &\triangleq \frac{1}{2} (\Delta_{24|3} \mathbf{g} + \Delta_{34|2} \mathbf{g}) \\ b(\mathbf{g}) &\triangleq \square_{12,34} \mathbf{g} + \Delta_{34|2} \mathbf{g} + \Delta_{24|3} \mathbf{g} \\ c(\mathbf{g}) &\triangleq \Delta_{23|4} \mathbf{g}. \end{aligned} \tag{5}$$

Proposition 1 Suppose $\mathbf{g} \in \mathcal{F}_n$ satisfies (3) for all positive integers s and $c(\mathbf{g}) \geq 0$ (or equivalently, $Q(0; a(\mathbf{g}), b(\mathbf{g}), c(\mathbf{g})) \geq 0$). Then

- $a(\mathbf{g}) \geq 0$,
- $a(\mathbf{g}) = 0 \Rightarrow b(\mathbf{g}) \geq 0$, and
- $a(\mathbf{g}) + b(\mathbf{g}) + 2c(\mathbf{g}) \geq 0$. Furthermore, equality holds if and only if $a(\mathbf{g}) = b(\mathbf{g}) = c(\mathbf{g}) = 0$.

Proof: Direct verification. \square

In the following, we will derive non-linear information inequalities from the sequence of linear inequalities (3).

Theorem 2 Suppose that $\mathbf{g} \in \mathcal{F}_n$ and $b(\mathbf{g}) \leq 2a(\mathbf{g})$. Let

$$w(\mathbf{g}) \triangleq \begin{cases} -\frac{b(\mathbf{g})-a(\mathbf{g})}{2a(\mathbf{g})} & \text{if } a(\mathbf{g}) > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then \mathbf{g} satisfies (3) for all nonnegative integers s if and only if $a(\mathbf{g}), c(\mathbf{g}) \geq 0$ and

$$(b(\mathbf{g}) - a(\mathbf{g}))^2 - 4a(\mathbf{g})c(\mathbf{g}) \leq \min(4a(\mathbf{g})^2(w(\mathbf{g}) - \lfloor w(\mathbf{g}) \rfloor)^2, 4a(\mathbf{g})^2(\lceil w(\mathbf{g}) \rceil - w(\mathbf{g}))^2). \quad (6)$$

Proof: To simplify notation, $a(\mathbf{g}), b(\mathbf{g})$ and $c(\mathbf{g})$ will simply be denoted as a, b and c . We will first prove the only-if part. Assume that \mathbf{g} satisfies (3) for all nonnegative integers s . When $s = 0, c \geq 0$. By Proposition 1, $a \geq 0$. It remains to prove that (6) holds.

Suppose first that $a > 0$. If the quadratic $Q(s; a, b, c)$ has no distinct real roots in s , then clearly $(b - a)^2 - 4ac \leq 0$ and the theorem holds. On the other hand, if $Q(s; a, b, c)$ has distinct real roots, implying $(b - a)^2 - 4ac > 0$, then $Q(s; a, b, c)$ is negative and is at its minimum when $s = -(b - a)/2a$ which is greater than $-1/2$ by assumption.

Since $Q(s; a, b, c) \geq 0$ for all non-negative integer s , the “distance” between the two roots can be at most $2 \min(w - \lfloor w \rfloor, \lceil w \rceil - w)$. In other words,

$$\frac{\sqrt{(b - a)^2 - 4ac}}{a} \leq 2 \min(w - \lfloor w \rfloor, \lceil w \rceil - w)$$

or equivalently, $(b - a)^2 - 4ac \leq \min(4a^2(w - \lfloor w \rfloor)^2, 4a^2(\lceil w \rceil - w)^2)$.

If on the other hand that $a = 0$, then the assumption $b \leq 2a$ and Proposition 1 implies that $b = 0$. As such, the quadratic inequality $(b - a)^2 - 4ac \leq 0$ and (6) clearly holds. Hence, the only-if part of the theorem is proved.

Now, we will prove the if-part. If $a = 0$, then (6) and the assumption $b \leq 2a$ implies that $b = 0$. The theorem then holds as $c \geq 0$ by assumption. Now suppose $a > 0$ and $b \leq 2a$. Using a similar argument as before, (6) implies that either $Q(s; a, b, c) \geq 0$ has no real roots or the two real roots are within the closed interval $[\lfloor w \rfloor, \lceil w \rceil]$. Since $a > 0$, for all nonnegative integer s , we have $Q(s; a, b, c) \geq 0$, or equivalently, $s^2a + s(b - a) + c \geq 0$ and hence the theorem is proved. \square

Theorem 2 showed that Matúš series of linear inequalities is equivalent to the single non-linear inequality (6) under the condition that that $b(\mathbf{g}) \leq 2a(\mathbf{g})$ and $a(\mathbf{g}), c(\mathbf{g}) \geq 0$.

Clearly, $a(\mathbf{g}), c(\mathbf{g}) \geq 0$ holds for all entropic \mathbf{g} because of the nonnegativity of conditional mutual information. Therefore, imposing these two conditions does not very much weaken (6). If on the other hand that $b(\mathbf{g}) \leq 2a(\mathbf{g})$ does not hold, then Matúš series of inequalities are implied by that $a(\mathbf{g}), c(\mathbf{g}) \geq 0$. In that case, Matúš’ inequalities will not be of interest. Therefore, our proposed nonlinear inequality essentially is not much weaker than Matúš’ ones.

While (6) is interesting in its own right, it is not so easy to work with. In the following, we shall consider a weaker form.

Corollary 1 (Quadratic information inequality) *Suppose that \mathbf{g} satisfies (3) for all nonnegative integers s . If $b(\mathbf{g}) \leq 2a(\mathbf{g})$, then*

$$(b(\mathbf{g}) - a(\mathbf{g}))^2 - 4a(\mathbf{g})c(\mathbf{g}) \leq a(\mathbf{g})^2. \tag{7}$$

Consequently, if \mathbf{g} is almost entropic and $\square_{12,34} \mathbf{g} \leq 0$ then

$$\left(\square_{12,34} \mathbf{g} + \frac{\Delta_{24|3} \mathbf{g} + \Delta_{34|2} \mathbf{g}}{2} \right)^2 - 2(\Delta_{24|3} \mathbf{g} + \Delta_{34|2} \mathbf{g})\Delta_{32|4} \mathbf{g} \leq \frac{(\Delta_{24|3} \mathbf{g} + \Delta_{34|2} \mathbf{g})^2}{4}.$$

Proof: Since $\min(w(\mathbf{g}) - \lfloor w(\mathbf{g}) \rfloor, \lceil w(\mathbf{g}) \rceil - w(\mathbf{g})) \leq 1/2$, the corollary then follows directly from Theorem 2. \square

Despite the fact that the above “quadratic” information inequality is a consequence of a series of linear inequalities, to the best of our knowledge, it is indeed the first non-trivial *non-linear* information inequality.

3.2. Implications of Corollary 1

In Proposition 1, we showed that Matúš’ inequalities imply that if $a(\mathbf{g}) = 0$, then $b(\mathbf{g}) \geq 0$. The same result can also be proved by using the quadratic information inequality in (7).

Implication 1 *For any $\mathbf{g} \in \mathcal{F}_n$ such that*

$$b(\mathbf{g}) \leq 2a(\mathbf{g}) \Rightarrow (b(\mathbf{g}) - a(\mathbf{g}))^2 - 4a(\mathbf{g})c(\mathbf{g}) \leq a(\mathbf{g})^2, \tag{8}$$

then $a(\mathbf{g}) = 0$ implies $b(\mathbf{g}) \geq 0$.

Proof: If $a(\mathbf{g}) = 0$, then $(b(\mathbf{g}) - a(\mathbf{g}))^2 - 4a(\mathbf{g})c(\mathbf{g}) - a(\mathbf{g})^2 = b(\mathbf{g})^2$. Hence, if $b(\mathbf{g}) < 0$, then (8) will be violated leading to a contradiction. \square

In [1], it was proved that the cone $\bar{\Gamma}_n^*$ is not polyhedral for $n \geq 4$. Ignoring the technical details, the idea of the proof is very simple. First, a sequence of entropic functions \mathbf{g}_t was constructed such that (1) the sequence converges to \mathbf{g}_0 , and (2) it has a one-side tangent $\dot{\mathbf{g}}_{0+}$ which is defined as $\lim_{t \rightarrow 0^+} (\mathbf{g}_t - \mathbf{g}_0)/t$. Clearly, if $\bar{\Gamma}_n^*$ is polyhedral, then there exists $\epsilon > 0$ such that $\mathbf{g}_0 + \epsilon \dot{\mathbf{g}}_{0+}$ is contained in $\bar{\Gamma}_n^*$. It was then shown that for any $\epsilon > 0$, the function $\mathbf{g}_0 + \epsilon \dot{\mathbf{g}}_{0+}$ is not in $\bar{\Gamma}_n^*$ because it violates (3) for sufficiently large s . Therefore, $\bar{\Gamma}_n^*$ is not polyhedral, or equivalently, there are infinitely many information inequalities.

In fact, we can also show that $\mathbf{g}_0 + \epsilon \dot{\mathbf{g}}_{0+}$ also violates the quadratic information inequality obtained in Corollary 1 for any positive ϵ . As such, (7) is sufficient to prove that $\bar{\Gamma}_n^*$ is not polyhedral for $n \geq 4$ and hence the following implication.

Implication 2 *The quadratic inequality (7) is strong enough to imply that $\bar{\Gamma}_n^*$ is not polyhedral.*

Some nonlinear information inequalities are direct consequences of basic linear information inequalities (e.g., $H(X)^2 I(X; Y) \geq 0$). Such inequalities are trivial in that they are obtained directly as nonlinear transformations of known linear inequalities. Our proposed quadratic inequality (7) is non-trivial, as proved in the following.

Implication 3 *The quadratic inequality (7) is a non-linear inequality that cannot be implied by any finite number of linear information inequalities. Specifically, for any given finite set of valid linear information inequalities, there exists $\mathbf{g} \notin \bar{\Gamma}_n^*$ such that \mathbf{g} does not satisfy (7) but satisfies all the given linear inequalities.*

Proof: Suppose we are given a finite set of valid linear information inequalities. Then the set of $\mathbf{g} \in \mathcal{F}_n$ satisfying all these linear inequalities is polyhedral. In other words, the set is obtained by taking intersection of a finite number of half-spaces. For simplicity, such a polyhedron will be denoted by Ψ .

We will once again use the sequence of entropic functions $\{\mathbf{g}_t\}_{t=1}^\infty$ constructed in [1]. Clearly, $\mathbf{g}_t \in \Psi$ for all t since $\mathbf{g} \in \bar{\Gamma}_n^*$. Again, as Ψ is polyhedral, $\mathbf{g}_\epsilon \triangleq \mathbf{g}_0 + \epsilon \dot{\mathbf{g}}_{0+} \in \Psi$ for sufficiently small $\epsilon > 0$. In other words, $\mathbf{g}_0 + \epsilon \dot{\mathbf{g}}_{0+}$ satisfies all the given linear inequalities. However, as explained earlier, \mathbf{g}_ϵ violates the quadratic inequality (7) and hence the theorem follows. \square

4. Characterizing $\bar{\Gamma}_n^*$ by projection

Although the set of almost entropic functions $\bar{\Gamma}_n^*$ is a closed and convex cone, finding a complete characterization is an extremely difficult task. Therefore, instead of tackling the hard problem directly, it is sensible to consider a relatively simpler problem – the characterization of a “projection” of $\bar{\Gamma}_n^*$. This projection problem is easier because the dimension of a projection can be much smaller, making it easier to be visualized and to be described. Furthermore, its low dimensionality may also facilitate the use of numerical techniques to find an approximation for the projection.

In this section, we consider a particular projection and will show how inequalities obtained in the previous section be expressed by equivalent ones on the proposed projection. As such, we can have a better idea how the projection looks like. First, we will define our proposed projection Υ .

Define $\Upsilon = \{(a(\mathbf{g}), b(\mathbf{g}) - a(\mathbf{g})) : \mathbf{g} \in \bar{\Gamma}_n^* \text{ and } a(\mathbf{g}) + b(\mathbf{g}) + 2c(\mathbf{g}) = 1\}$, or equivalently,

$$\Upsilon = \left\{ \left(\frac{a(\mathbf{g})}{a(\mathbf{g}) + b(\mathbf{g}) + 2c(\mathbf{g})}, \frac{b(\mathbf{g}) - a(\mathbf{g})}{a(\mathbf{g}) + b(\mathbf{g}) + 2c(\mathbf{g})} \right) : \mathbf{g} \in \bar{\Gamma}_n^* \text{ and } \mathbf{g} \neq \mathbf{0} \right\}. \tag{9}$$

Lemma 1 *Υ is a closed and convex set.*

Proof: Since the set $\{(a(\mathbf{g}), b(\mathbf{g}) - a(\mathbf{g})) : \mathbf{g} \in \bar{\Gamma}_n^*\}$ is a closed and convex one, its cross-section (and its affine transform) Υ is also closed and convex. \square

Since Υ is obtained by projecting $\bar{\Gamma}_n^*$ onto a two-dimensional Euclidean space, any inequality satisfied by all points in Υ induces a corresponding information inequality. Specifically, we have the following proposition.

Proposition 2 Suppose that there exists $k \geq 0$ such that

$$(a + b + 2c)^k \psi \left(\frac{a}{a + b + 2c}, \frac{b - a}{a + b + 2c} \right) \geq 0 \text{ if } a = b = c = 0. \tag{10}$$

Then

$$\psi(u, v) \geq 0, \forall (u, v) \in \Upsilon \tag{11}$$

if and only if

$$(a(\mathbf{g}) + b(\mathbf{g}) + 2c(\mathbf{g}))^k \psi \left(\frac{a(\mathbf{g})}{a(\mathbf{g}) + b(\mathbf{g}) + 2c(\mathbf{g})}, \frac{b(\mathbf{g}) - a(\mathbf{g})}{a(\mathbf{g}) + b(\mathbf{g}) + 2c(\mathbf{g})} \right) \geq 0, \forall \mathbf{g} \in \bar{\Gamma}_n^*. \tag{12}$$

Similarly, (11) holds for all $(u, v) \in \Upsilon$ and $v \leq u$ if and only if (12) holds for all $\mathbf{g} \in \bar{\Gamma}_n^*$ and $b(\mathbf{g}) \leq 2a(\mathbf{g})$.

Proof: First, we will prove that (11) implies (12). For any $\mathbf{g} \in \bar{\Gamma}_n^*$. If $a(\mathbf{g}) + b(\mathbf{g}) + 2c(\mathbf{g}) = 0$, then by Proposition 1, $a(\mathbf{g}) = b(\mathbf{g}) = c(\mathbf{g}) = 0$ and (12) follows from (10). Otherwise, $a(\mathbf{g}) + b(\mathbf{g}) + 2c(\mathbf{g}) > 0$ and (12) follows from (9).

Conversely, for any $(u, v) \in \Upsilon$, by definition, there exists $\mathbf{g} \in \bar{\Gamma}_n^*$ such that (1) $\mathbf{g} \neq 0$ and (2) $u = a(\mathbf{g}) / (a(\mathbf{g}) + b(\mathbf{g}) + 2c(\mathbf{g}))$ and $v = (b(\mathbf{g}) - a(\mathbf{g})) / (a(\mathbf{g}) + b(\mathbf{g}) + 2c(\mathbf{g}))$. The inequality (11) then follows from (12) and that $\mathbf{g} \neq 0$ (hence, $a(\mathbf{g}) + b(\mathbf{g}) + 2c(\mathbf{g}) > 0$).

Finally, the constrained counterpart follows from that $(b(\mathbf{g}) - a(\mathbf{g})) \leq a(\mathbf{g})$ if and only if $b(\mathbf{g}) \leq 2a(\mathbf{g})$. □

By Proposition 2, there is a mechanical way to rewrite inequalities for $\bar{\Gamma}_n^*$ as ones for Υ , and vice versa. Therefore, we will abuse notations by calling that (11) and (12) equivalent. In the following, we will rewrite inequalities obtained in previous sections by using Proposition 2.

Proposition 3 (Matúš’ inequalities) When s is a positive integer, the inequality (3) is equivalent to

$$v \geq \frac{2u - 2s^2u - 1}{2s - 1}. \tag{13}$$

Proof: A direct consequence of Proposition 2 and that

$$\frac{c(\mathbf{g})}{a(\mathbf{g}) + b(\mathbf{g}) + 2c(\mathbf{g})} = \frac{1}{2} \left(1 - \frac{b(\mathbf{g}) - a(\mathbf{g})}{a(\mathbf{g}) + b(\mathbf{g}) + 2c(\mathbf{g})} - 2 \frac{a(\mathbf{g})}{a(\mathbf{g}) + b(\mathbf{g}) + 2c(\mathbf{g})} \right). \tag{14}$$

□

By optimizing the choice of s , we can obtain a stronger piecewise linear inequality which can be rewritten as follows.

Theorem 3 (Piecewise linear inequality) The piecewise linear inequality

$$\min_{s \in \mathbb{Z}^+} s^2 a(\mathbf{g}) + s(b(\mathbf{g}) - a(\mathbf{g})) + c(\mathbf{g}) \geq 0 \tag{15}$$

is equivalent to that

$$v \geq L^{li}(u), \tag{16}$$

where $L^{li}(u) \triangleq \sup_{s \in \mathbb{Z}^+} (2u - 2s^2u - 1) / (2s - 1)$.

Proof: A direct consequence of Propositions 2 and 3. \square

As we shall see in the following lemma, $L^{li}(u)$ can be explicitly characterized.

Lemma 2 $L^{li}(0) = 0$ and

$$L^{li}(u) = (2u - 2s_o^2u - 1)/(2s_o - 1)$$

for any $0 < u \leq 1$, where s_o is the smallest positive integer such that $1/(1 + 2s_o^2) \leq u$.

Proof: Let $f(s, u) \triangleq (2u - 2s^2u - 1)/(2s - 1)$. First, $f(s, 0) = -1/(2s - 1)$. Therefore, $L^{li}(0) = \sup_{s \in \mathbb{Z}^+} f(s, 0) = 0$. Also, it is straightforward to prove that

- for any fixed $u \geq 1/2$, $f(s, u)$ is a decreasing function of s and hence $\sup_{s \in \mathbb{Z}^+} f(s, u) = f(1, u) = -1$.
- for $0 < u \leq 1/2$, $f(s, u)$ is a strictly concave function of s for $s \geq 1$ and is at its maximum when $s = \frac{1}{2} + \sqrt{\frac{1}{2u} - \frac{3}{4}} \geq 1$. As a result, $L^{li}(u) = \max(f(s_l, u), f(s_h, u))$ where $s_l = \lfloor \frac{1}{2} + \sqrt{\frac{1}{2u} - \frac{3}{4}} \rfloor$ and $s_h = \lceil \frac{1}{2} + \sqrt{\frac{1}{2u} - \frac{3}{4}} \rceil$.

Clearly, for any positive integer s_o ,

$$L^{li}(u) = \begin{cases} f(s_o, u) & \text{if } u = 1/2(1 - s_o + s_o^2) \\ f(s_o + 1, u) & \text{if } u = 1/2(1 + s_o + s_o^2) \end{cases}.$$

Furthermore, if $1/2(1 + s_o + s_o^2) < u < 1/2(1 - s_o + s_o^2)$, we have $s_l = s_o$ and $s_h = s_o + 1$ and hence,

$$L^{li}(u) = \max \left(\frac{2u - 2s_o^2u - 1}{2s_o - 1}, \frac{-4s_o u - 2s_o^2u - 1}{2s_o + 1} \right). \tag{17}$$

By solving a system of linear equations, we can show that $f(s_o, u) = f(s_o + 1, u)$ if and only if $u = 1/(1 + 2s_o^2)$. Therefore,

$$L^{li}(u) = \begin{cases} f(s_o + 1, u) & \text{if } 1/2(1 + s_o + s_o^2) < u \leq 1/(1 + 2s_o^2) \\ f(s_o, u) & \text{if } 1/(1 + 2s_o^2) \leq u \leq 1/2(1 - s_o + s_o^2). \end{cases}$$

Together with the fact that $L^{li}(u) = -1 = f(1, u)$ for $1/2 \leq u \leq 1$, the lemma follows. \square

Proposition 4 (Quadratic inequality) *The quadratic inequality (7) (subject to that $b(\mathbf{g}) \leq 2a(\mathbf{g})$) is equivalent to*

$$(v + u)^2 + 2u^2 \leq 2u \tag{18}$$

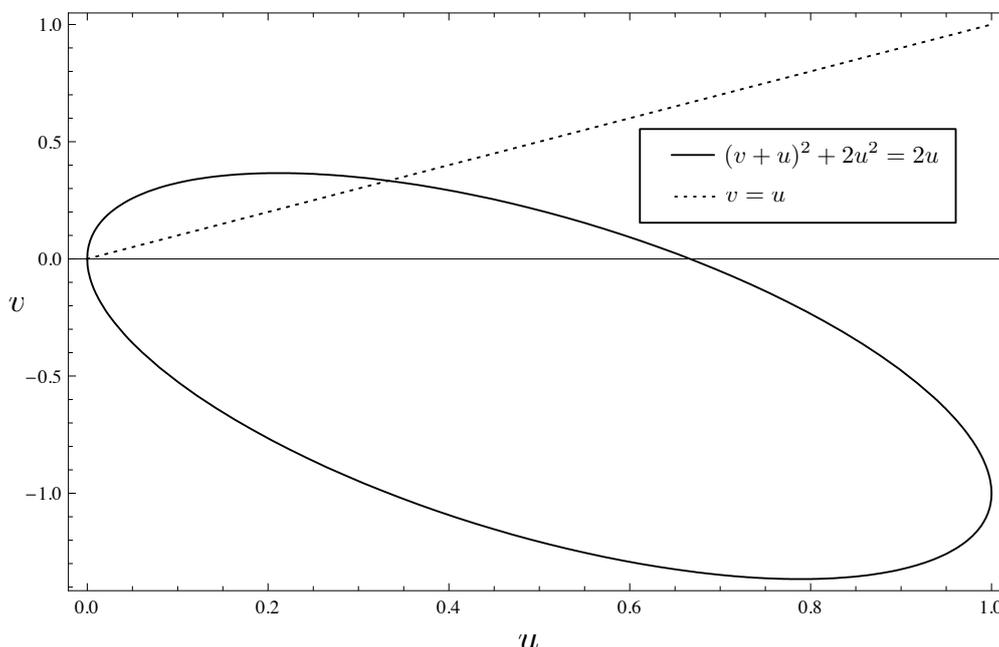
subject to that $v \leq u$.

Proof: By using Proposition 2 and (14), it is straightforward to rewrite (7) as (18). \square

To illustrate (18), we plot the curves $(v + u)^2 + 2u^2 = 2u$ and $v = u$ in Figure 1. From the proposition, if $v \leq u$ (i.e., the point (u, v) is below the dotted line), then $(u, v) \in \Upsilon$ implies that (u, v) is inside the ellipse.

Proposition 4 gives a nonlinear information inequality on Υ subject to a condition that $v \leq u$. In the following theorem, we relax the inequality so as to remove the condition.

Figure 1. Quadratic inequality (18).



Theorem 4 (Non-linear inequality) Let

$$L^{nl}(u) = -u - \sqrt{2u - 2u^2}. \tag{19}$$

For any $(u, v) \in \Upsilon$, $v \geq L^{nl}(u)$. Consequently, by Proposition 2,

$$b(\mathbf{g}) \geq -\sqrt{2a(\mathbf{g})(a(\mathbf{g}) + b(\mathbf{g}) + 2c(\mathbf{g})) - 2a(\mathbf{g})^2} \tag{20}$$

$$= -\sqrt{2a(\mathbf{g})(b(\mathbf{g}) + 2c(\mathbf{g}))} \tag{21}$$

Proof: By Proposition 4, if $(u, v) \in \Upsilon$ such that $v \leq u$, then

$$(v + u)^2 + 2u^2 \leq 2u.$$

As a result, $v + u \geq -\sqrt{2u - 2u^2}$ or equivalently, $v \geq -u - \sqrt{2u - 2u^2}$. On the other hand, if $v \geq u$, then $v \geq 0$ and hence $v \geq -u - \sqrt{2u - 2u^2}$. The theorem then follows from Proposition 2. \square

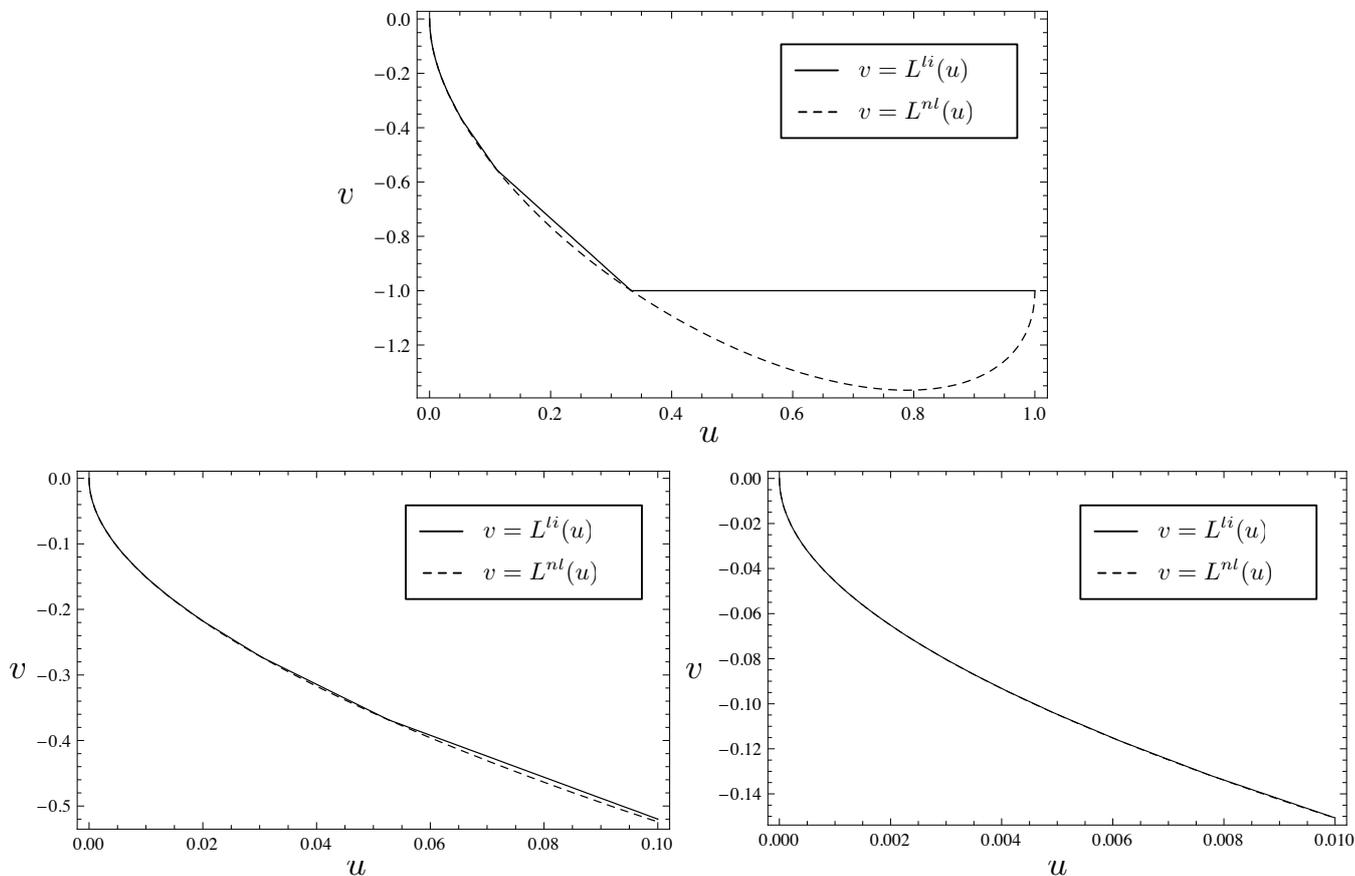
In the next proposition, we will show that the piecewise linear inequality $v \geq L^{li}(u)$ and the proposed nonlinear inequality $v \geq L^{nl}(u)$ coincides for countably infinite number of u .

Proposition 5 For any $0 \leq u \leq 1$, we have $L^{nl}(u) \leq L^{li}(u)$. Furthermore, equality holds if $u = 1/(1 + 2s^2)$ for some nonnegative integer s .

Proof: By definition, $L^{nl}(0) = L^{li}(0) = 0$ and the proposition holds in this case. Assume now that $0 < u \leq 1$. We first show that $L^{nl}(u) = L^{li}(u)$ when $u = 1/(1 + 2s^2)$ for some nonnegative integer s . Suppose that $s = 0$, then $L^{nl}(u) = L^{li}(u) = -1$. On the other hand, if $u = 1/(1 + 2s^2)$ where s is a positive integer, then it is straightforward to prove that

$$L^{li}(u) = f(s, u) = f(s + 1, u) = \frac{-1 - 2s}{1 + 2s^2} = L^{nl}(u). \tag{22}$$

Figure 2. Piecewise linear inequality and nonlinear inequality.



By differentiating $L^{nl}(u)$ with respect to u , we can prove that $L^{nl}(u)$ is convex over $[0, 1]$. For each nonnegative integer s , $L^{li}(u)$ is linear over the interval $[1/(1+2(s+1)^2), 1/(1+2s^2)]$ and $L^{nl}(u) = L^{li}(u)$ when $u = 1/(1 + 2s^2)$ or $1/(1 + 2(s + 1)^2)$. Hence, $L^{li}(u) \geq L^{nl}(u)$ over the interval by the convexity of $L^{nl}(u)$. As s can be arbitrarily large, $L^{li}(u) \geq L^{nl}(u)$ for $u \in (0, 1]$ and the theorem then follows. \square

5. Conclusion

In this paper, we constructed several piecewise linear and quadratic information inequalities from a series of information inequalities proved in [1]. Our proposed nonlinear inequality (6) was shown to be equivalent to the whole set of Matúš’ linear inequalities. Hence, we can replace all Matúš’ inequalities with our proposed ones.

However, the inequality is not smooth and may not be easy to work with. Therefore, we relax these nonlinear inequalities to quadratic ones. These quadratic inequalities are strong enough to show that the set of almost entropic functions is not polyhedral.

It is certain that the proposed quadratic inequalities we obtained in (16) and (19) are a consequence of Matúš’ linear inequalities. Yet, the non-linear inequality has a much simpler form. By comparing the inequalities on projections of $\bar{\Gamma}_n^*$, our figures suggested that these nonlinear inequalities are indeed

fairly good approximations to the corresponding piecewise linear inequalities. Furthermore, they are of particular interest for several reasons.

First, all these inequalities are non-trivial and cannot be deduced from any finite number of linear information inequalities. To the best of our knowledge, they are the first non-trivial nonlinear information inequalities. Second, in some cases, it will be relatively easier to work with a single nonlinear inequality, rather than an infinite number of linear inequalities. For example, in order to compute some bounds on a capacity region (say, in a network coding problem), a characterization of $\bar{\Gamma}_n^*$ may be needed as input to a computing system. Surely, $\bar{\Gamma}_n^*$ is unknown and hence an outer bound of $\bar{\Gamma}_n^*$ will be used instead. If one replace the countably infinite number of linear inequalities with a single nonlinear inequality, it may greatly simplify the computing problem. Third, these nonlinear inequalities prompt us to ask new fundamental questions - are nonlinear information inequalities more fundamental than linear information inequalities? Would it be possible that the set $\bar{\Gamma}_n^*$ be completely characterized by a finite number of nonlinear inequalities? If so, what will they look like?

As a final remark, Matúš' inequalities, and also all the non-linear inequalities we obtained, are "tighter" than the Shannon inequalities only in the region where $b(\mathbf{g}) \leq 2a(\mathbf{g})$. When $b(\mathbf{g}) \geq 2a(\mathbf{g})$, the two inequalities are direct consequences of non-negativity of conditional mutual information. This phenomenon seems to suggest that entropic functions are much more difficult to characterize in the region $b(\mathbf{g}) < 2a(\mathbf{g})$. An explanation for this phenomenon is still lacking.

Acknowledgements

This work was supported by the Australian Government under ARC grant DP0557310.

References and Notes

1. Matúš, F. Infinitely many information inequalities. In *IEEE Int. Symp. Inform. Theory*, Nice, France, July 2007, pp. 41–44.
2. Zhang, Z.; Yeung, R.W. On the characterization of entropy function via information inequalities. *IEEE Trans. Inform. Theory* **1998**, *44*, 1440–1452.
3. Chan, T. H.; Yeung, R.W. On a relation between information inequalities and group theory. *IEEE Trans. Inform. Theory* **2002**, *48*, 1992–1995.
4. Dougherty, R.; Freiling, C.; Zeger, K. Six new non-Shannon information inequalities. In *IEEE Int. Symp. Inform. Theory*, Seattle, USA, July 2006, pp. 233–236.
5. Chan, T.H.; Grant, A. Dualities between entropy functions and network codes. *IEEE Trans. Inform. Theory* **2008**, *54*, 4470–4487.
6. Yeung, R. A framework for linear information inequalities. *IEEE Trans. Inform. Theory* **1997**, *43*, 1924–1934.
7. Song, L.; Yeung, R. W.; Cai, N. Zero-error network coding for acyclic networks. *IEEE Trans. Inform. Theory* **2003**, *49*, 3129–3139.