Less Noisy Domination by Symmetric Channels

Anuran Makur and Yury Polyanskiy
EECS Department, Massachusetts Institute of Technology
Email: {a_makur, yp}@mit.edu

***Abstract*—Consider the family of all q-ary symmetric channels (q-SCs) with capacities decreasing from $\log(q)$ to 0. This paper addresses the following question: what is the member of this family with the smallest capacity that dominates a given channel V in the "less noisy" preorder sense. When the q-SCs are replaced by q-ary erasure channels, this question is known as the "strong data processing inequality." We provide several equivalent characterizations of the less noisy preorder in terms of χ^2-divergence, Löwner (PSD) partial order, and spectral radius. We then illustrate a simple criterion for domination by a q-SC based on degradation, and mention special improvements for the case where V is an additive noise channel over an Abelian group of order q. Finally, as an application, we discuss how logarithmic Sobolev inequalities for q-SCs, which are well-studied, can be transported to an arbitrary channel V.**

I. Introduction

The less noisy preorder over channels was developed in [1], and has since been primarily utilized in network information theory. For instance, it and its variants have been used to study the capacity regions of broadcast channels in [1]–[4] and the references therein. Formally, given two discrete channels $P_{Y|X} = W \in \mathbb{R}_{\mathsf{sto}}^{q\times r}$ and $P_{Z|X} = V \in \mathbb{R}_{\mathsf{sto}}^{q\times s}$ (where we represent channels as row stochastic matrices, and $X \in \mathcal{X}$, $Y \in \mathcal{Y}$, $Z \in \mathcal{Z}$ with $|\mathcal{X}| = q$, $|\mathcal{Y}| = r$, $|\mathcal{Z}| = s$), we say that W is *less noisy* than V, denoted $W \succeq_{\mathsf{ln}} V$, if and only if:

$$I(U;Y) \geq I(U;Z) \tag{1}$$

for every joint distribution $P_{U,X}$ such that $U \to X \to (Y,Z)$ forms a Markov chain and the random variable $U \in \mathcal{U}$ has some arbitrary range $\mathcal{U}$ [1, Proposition 2], where $\mathbb{R}_{\mathsf{sto}}^{q\times r}$ denotes the set of $q \times r$ row stochastic matrices. Indeed, this definition intuitively captures the notion that W is less noisy than V. In this paper, we examine the basic question of when a q-ary symmetric channel (q-SC) is less noisy than a given discrete channel $V \in \mathbb{R}_{\mathsf{sto}}^{q\times s}$ with common input alphabet.

Why would one be interested in knowing whether a q-SC dominates a given V? We present several reasons below. Firstly, $\succeq_{\mathsf{ln}}$ domination by a q-SC turns out to be a natural extension of the so called *strong data processing inequality* (SDPI) as noticed in [5, Proposition 15]. For every Markov chain $U \to X \to Z$, the data processing inequality states that $I(U;X) \geq I(U;Z)$. Fixing the channel $P_{Z|X} = V \in \mathbb{R}_{\mathsf{sto}}^{q\times s}$, this inequality can be tightened to [6]:

$$\eta_{\mathsf{KL}}(V) I(U;X) \geq I(U;Z) \tag{2}$$

using the *contraction coefficient* $\eta_{\mathsf{KL}}(V) \in [0,1]$, defined as:

$$\eta_{\mathsf{KL}}(V) \triangleq \sup_{P_{U,X}} \frac{I(U;Z)}{I(U;X)} \tag{3}$$

where the supremum is over all Markov chains $U \to X \to Z$ such that $P_{Z|X}$ is fixed and $0 < I(U;X) < +\infty$. Frequently, one gets $\eta_{\mathsf{KL}}(V) < 1$ and the resulting inequality is called an SDPI. Such inequalities have been recently simultaneously rediscovered and applied in several disciplines; see [5, Sections 1-2] for a short survey. In [5, Proposition 15], the authors demonstrate using an elementary calculation that a q-ary erasure channel (q-EC) $E_\epsilon \in \mathbb{R}_{\mathsf{sto}}^{q\times(q+1)}$ with erasure probability $\epsilon \in [0,1]$ is less noisy than the channel V, $E_\epsilon \succeq_{\mathsf{ln}} V$, if and only if $\eta_{\mathsf{KL}}(V) \leq 1 - \epsilon$. Hence, the entire study of SDPIs is equivalent to determining whether a given channel is dominated in the less noisy sense by a q-EC. There are several useful upper bounds on η_{KL} that are suitable for this purpose [7], [8, Remark III.2], but no such results exist for q-SC domination.

This paper initiates the inquiry of a natural extension of the concept of SDPI by replacing the distinguished role played by the q-EC with a q-SC. In analogy with the bounds on η_{KL}, one of our goals is to establish similar simple criteria for testing domination by a q-SC instead of a q-EC.

Secondly, the less noisy preorder tensorizes [9, Proposition 5], [5, Proposition 16]. In other words, if we know $W \succeq_{\mathsf{ln}} V$, we can also conclude that $W^{\otimes n} \succeq_{\mathsf{ln}} V^{\otimes n}$, which means that $I(U;Y^n) \geq I(U;Z^n)$ for every Markov chain $U \to X^n \to (Y^n, Z^n)$. Therefore, many impossibility results (in statistical decision theory for example) that are proven by exhibiting bounds on quantities such as $I(U;Y^n)$ transparently carry over to statistical experiments with observations on the basis of Z^n. Since it is much more common to study the q-ary symmetric observation model (especially with $q = 2$), we can leverage its sample complexity lower bounds for other V.

Thirdly, there is a self-contained information theoretic motivation that addresses our leading question. $W \succeq_{\mathsf{ln}} V$ if and only if the secrecy capacity $C_S = 0$, where C_S is the secrecy capacity of the *Wyner wiretap channel* with V as the main (legal receiver) channel and W as the eavesdropper channel [10, Corollary 17.11]. Thus, finding the maximally noisy q-SC that dominates V establishes the minimal noise required on the eavesdropper link so that secret communication is feasible.

Finally, $\succeq_{\mathsf{ln}}$ domination turns out to entail a comparison of Dirichlet forms (as we will illustrate), and consequently, allows us to prove *Poincaré (spectral gap) and logarithmic Sobolev inequalities* (LSIs) for V from well-known results on q-SCs.

This research was supported by the National Science Foundation CAREER award under grant agreement CCF-12-53205, and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370.

These inequalities are cornerstones of the modern approach to Markov chains and concentration of measure [11], [12].

We briefly delineate the ensuing discussion. In Section II, we derive equivalent characterizations of $\succeq_{\mathsf{ln}}$ which are useful for obtaining other results. We then illustrate conditions for domination by q-SCs in Section III. Finally, we elucidate the relationship between $\succeq_{\mathsf{ln}}$ domination and LSIs in Section IV.

II. Equivalent characterizations of less noisy

We commence this section with an alternative definition of $\succeq_{\mathsf{ln}}$. Given the channels $W \in \mathbb{R}^{q\times r}_{\mathsf{sto}}$ and $V \in \mathbb{R}^{q\times s}_{\mathsf{sto}}$, $W \succeq_{\mathsf{ln}} V$ if and only if [1, Proposition 2], [5, Proposition 14]:

$$\forall P_X, Q_X \in \mathcal{P}_q,\ \ D(P_X W||Q_X W) \geq D(P_X V||Q_X V) \quad (4)$$

where $\mathcal{P}_q$ denotes the probability simplex of row vectors in $\mathbb{R}^q$. Although $\succeq_{\mathsf{ln}}$ is characterized by KL divergence or mutual information (in a manner pertinent for channel coding) in [1], our most general result illustrates that this preorder can also be characterized using χ^2-divergence, the Löwner (PSD) partial order, or a spectral radius condition. Recall that for any two pmfs $P_X, Q_X \in \mathcal{P}_q$, their χ^2-divergence is given by:

$$\chi^2(P_X||Q_X) \triangleq \sum_{x\in\mathcal{X}} \frac{(P_X(x) - Q_X(x))^2}{Q_X(x)} \quad (5)$$

where $(0-0)^2/0 = 0$ and $(p-0)^2/0 = +\infty$ for every $p > 0$ based on continuity arguments. The next theorem presents these characterizations of the less noisy preorder.

Theorem 1 (Equivalent Characterizations of $\succeq_{\mathsf{ln}}$)**.** *For any pair of channels $W \in \mathbb{R}^{q\times r}_{\mathsf{sto}}$ and $V \in \mathbb{R}^{q\times s}_{\mathsf{sto}}$ on the same input alphabet, the following are equivalent:*

1) $W \succeq_{\mathsf{ln}} V$
2) *For every $P_X, Q_X \in \mathcal{P}_q$:*
$$\chi^2(P_X W||Q_X W) \geq \chi^2(P_X V||Q_X V)$$
3) *For every $P_X \in \mathcal{P}_q^{\circ}$:*
$$W \mathsf{diag}(P_X W)^{-1} W^T \succeq_{\mathsf{PSD}} V \mathsf{diag}(P_X V)^{-1} V^T$$
4) *For every $P_X \in \mathcal{P}_q^{\circ}$:*
$$\rho\Big(\Big(W \mathsf{diag}(P_X W)^{-1} W^T\Big)^{\dagger} V \mathsf{diag}(P_X V)^{-1} V^T\Big) = 1$$

where $\mathcal{P}_q^{\circ}$ denotes the relative interior of $\mathcal{P}_q$, $\mathsf{diag}(x)$ denotes a diagonal matrix with $x \in \mathbb{R}^q$ along its principal diagonal, $\succeq_{\mathsf{PSD}}$ denotes the Löwner partial order, $\dagger$ denotes the Moore-Penrose pseudoinverse, and $\rho(\cdot)$ denotes spectral radius.

Proof. (1 $\Leftrightarrow$ 2) To prove the forward direction, recall the local approximation of KL divergence [13, Proposition 4.2], which states that for any $P_X, Q_X \in \mathcal{P}_q$, if $\chi^2(P_X||Q_X) < +\infty$:

$$\lim_{\lambda\to 0^+} \frac{2}{\lambda^2} D(\lambda P_X + \bar{\lambda} Q_X||Q_X) = \chi^2(P_X||Q_X) \quad (6)$$

where $\bar{\lambda} = 1 - \lambda$ for $\lambda \in (0,1)$, and all logarithms are natural. For any $P_X, Q_X \in \mathcal{P}_q$, if $\chi^2(P_X W||Q_X W) = +\infty$, then there is nothing to prove. If $\chi^2(P_X W||Q_X W) < +\infty$, then $D(P_X V||Q_X V) \leq D(P_X W||Q_X W) < +\infty$ using (4) and [13, Equation 7.8]. Hence, $P_X V \ll Q_X V$, which implies that $\chi^2(P_X V||Q_X V) < +\infty$. Since $W \succeq_{\mathsf{ln}} V$, we have from (4):

$$D\big(\lambda P_X W + \bar{\lambda} Q_X W||Q_X W\big) \geq D\big(\lambda P_X V + \bar{\lambda} Q_X V||Q_X V\big)$$

for any $\lambda \in [0,1]$. Applying (6) to both sides of this produces:

$$\chi^2(P_X W||Q_X W) \geq \chi^2(P_X V||Q_X V)$$

which proves the forward direction. For the converse, we recall an integral representation of KL divergence [5, Appendix A.2]:

$$D(P_X||Q_X) = \int_0^{\infty} \chi^2(P_X||Q_X^t)\, dt \quad (7)$$

for any $P_X, Q_X \in \mathcal{P}_q$, where $Q_X^t = \frac{t}{1+t} P_X + \frac{1}{t+1} Q_X$ for $t \in [0,\infty)$. Since for every $P_X, Q_X \in \mathcal{P}_q$, $\chi^2(P_X W||Q_X^t W) \geq \chi^2(P_X V||Q_X^t V)$, we can integrate both sides and use (7) to get $W \succeq_{\mathsf{ln}} V$ as follows:

$$\int_0^{\infty} \chi^2(P_X W||Q_X^t W)\, dt \geq \int_0^{\infty} \chi^2(P_X V||Q_X^t V)\, dt$$
$$D(P_X W||Q_X W) \geq D(P_X V||Q_X V).$$

(2 $\Leftrightarrow$ 3) Observe that for every $P_X \in \mathcal{P}_q$ and $Q_X \in \mathcal{P}_q^{\circ}$:

$$\chi^2(P_X||Q_X) = J_X \mathsf{diag}(Q_X)^{-1} J_X^T \quad (8)$$

where $J_X = P_X - Q_X$. Hence, for every $P_X \in \mathcal{P}_q$ and every $Q_X \in \mathcal{P}_q^{\circ}$, the Löwner condition in part 3 implies that:

$$J_X W \mathsf{diag}(Q_X W)^{-1} W^T J_X^T \geq J_X V \mathsf{diag}(Q_X V)^{-1} V^T J_X^T$$

which, using (8), is equivalent to:

$$\chi^2(P_X W||Q_X W) \geq \chi^2(P_X V||Q_X V).$$

This inequality also holds for $Q_X \in \mathcal{P}_q \backslash \mathcal{P}_q^{\circ}$ (as $Q_X^k \to Q_X \Rightarrow \chi^2(P_X||Q_X^k) \to \chi^2(P_X||Q_X)$), which proves the converse. The forward direction should appear plausible due to (8), and we refer readers to [14, Proposition 8] for a complete proof.

(3 $\Leftrightarrow$ 4) $W \mathsf{diag}(P_X W)^{-1} W^T$ and $V \mathsf{diag}(P_X V)^{-1} V^T$ are both positive semidefinite matrices for every $P_X \in \mathcal{P}_q^{\circ}$. If $W \mathsf{diag}(P_X W)^{-1} W^T$ is invertible, then it is actually positive definite, and [15, Theorem 7.7.3 (a)] (which states that if $A \in \mathbb{R}^{q\times q}$ is positive definite and $B \in \mathbb{R}^{q\times q}$ is positive semidefinite, then $A \succeq_{\mathsf{PSD}} B$ if and only if $\rho\left(A^{-1}B\right) \leq 1$) implies that the Löwner condition in part 3 is equivalent to:

$$\rho\Big(\Big(W \mathsf{diag}(P_X W)^{-1} W^T\Big)^{-1} V \mathsf{diag}(P_X V)^{-1} V^T\Big) \leq 1 \quad (9)$$

for every $P_X \in \mathcal{P}_q^{\circ}$. Observe that $P_X W \mathsf{diag}(P_X W)^{-1} W^T = P_X V \mathsf{diag}(P_X V)^{-1} V^T = \mathbf{1}^T$, where $\mathbf{1} \triangleq [1 \cdots 1]^T \in \mathbb{R}^q$ is the column vector with all entries equal to unity. So, we have:

$$\mathbf{1}^T \Big(W \mathsf{diag}(P_X W)^{-1} W^T\Big)^{-1} V \mathsf{diag}(P_X V)^{-1} V^T = \mathbf{1}^T$$

which means that this matrix has an eigenvalue of 1, and the inequality in (9) is really an equality. This completes the proof for invertible $W \mathsf{diag}(P_X W)^{-1} W^T$. The more general case is derived in [14, Proposition 9] using an appropriate extension of [15, Theorem 7.7.3 (a)] proved in [14, Lemma 2]. ∎

Since (6) and (7) hold more generally, the first equivalence in Theorem 1 can be verified for more general Markov kernels. Furthermore, it is related to the following notable result [6]:

$$\eta_{\mathsf{KL}}(W) = \eta_{\chi^2}(W) \triangleq \sup_{P_X, Q_X \in \mathcal{P}_q} \frac{\chi^2(P_X W || Q_X W)}{\chi^2(P_X || Q_X)} \tag{10}$$

for any channel $W \in \mathbb{R}_{\mathsf{sto}}^{q \times r}$, where the supremum is over all pmfs $P_X, Q_X \in \mathcal{P}_q$ that satisfy $0 < \chi^2(P_X || Q_X) < +\infty$. Indeed, (10) portrays how less noisy domination by a q-EC (which is characterized by η_{KL}) can be characterized by χ^2-divergence, and our first equivalence generalizes this result to less noisy domination by an arbitrary channel. Finally, we remark that the Löwner characterization in Theorem 1 is useful for deriving our other results, and the spectral characterization can be useful for computationally deducing whether $W \succeq_{\mathsf{ln}} V$.

III. Conditions for less noisy domination

We next establish sufficient conditions for less noisy domination by q-SCs. Formally, a *q-ary symmetric channel* has input and output alphabet $\mathcal{X}$ with $|\mathcal{X}| = q$ and transition probability matrix $W_\delta \in \mathbb{R}_{\mathsf{sto}}^{q \times q}$:

$$W_\delta \triangleq \left(1 - \delta - \frac{\delta}{q-1}\right) I_q + \frac{\delta}{q-1} \mathbf{1}\mathbf{1}^T \tag{11}$$

where $I_q \in \mathbb{R}^{q \times q}$ denotes the identity matrix, and $\delta \in [0, 1]$ is the total crossover probability. Matrices of the form (11) satisfy several useful properties; they are symmetric, doubly stochastic, circulant, jointly diagonalizable by the DFT matrix, and $\left\{W_\delta \in \mathbb{R}^{q \times q} : \delta \in \mathbb{R} \backslash \left\{\frac{q-1}{q}\right\}\right\}$ with the multiplication operation is an Abelian group [14, Proposition 3]. Given another channel $V \in \mathbb{R}_{\mathsf{sto}}^{q \times q}$ on the same alphabet, our objective is to find the extremal $\delta^\star(V) \triangleq \sup\left\{\delta \in \left[0, \frac{q-1}{q}\right] : W_\delta \succeq_{\mathsf{ln}} V\right\}$ such that for every $0 \le \delta < \delta^\star(V)$, $W_\delta \succeq_{\mathsf{ln}} V$.

Computationally estimating $\delta^\star(V)$ corresponds to a problem of verifying collections of rational inequalities. Indeed, Theorem 1 suggests the following minimax formulation of $\delta^\star(V)$:

$$\delta^\star(V) = \inf_{P_X \in \mathcal{P}_q^\circ} \sup_{\delta \in \mathcal{S}(P_X)} \delta \tag{12}$$

where $\mathcal{S}(P_X) = \left\{\delta \in \left[0, \frac{q-1}{q}\right] : W_\delta \, \mathsf{diag}(P_X W_\delta)^{-1} W_\delta^T \succeq_{\mathsf{PSD}} V \mathsf{diag}(P_X V)^{-1} V^T\right\}$. The infimum can be naïvely approximated by sampling several $P_X \in \mathcal{P}_q^\circ$. Estimating the supremum entails testing collections of rational inequalities in δ, because positive semidefiniteness of a matrix can be established using the non-negativity of its principal minors by Sylvester's criterion [15, Theorem 7.2.5]. On the other hand, analytically determining $\delta^\star(V)$ appears to be intractable. So, we instead prove a sufficient condition for $W_\delta \succeq_{\mathsf{ln}} V$ in the next theorem. Our result can be construed as a lower bound on $\delta^\star(V)$:

$$\delta^\star(V) \ge \frac{\nu}{1 - (q-1)\nu + \frac{\nu}{q-1}} \tag{13}$$

where ν is the minimum conditional probability in V.

To present the ensuing theorem, we introduce the (output) degradation preorder over channels, which was also defined to study broadcast channels [16]. In particular, a channel $V \in \mathbb{R}_{\mathsf{sto}}^{q \times s}$ is said to be a *degraded* version of a channel $W \in \mathbb{R}_{\mathsf{sto}}^{q \times r}$ with the same input alphabet, denoted $W \succeq_{\mathsf{deg}} V$, if $V = WA$ for some channel $A \in \mathbb{R}_{\mathsf{sto}}^{r \times s}$. It is well-known that $W \succeq_{\mathsf{deg}} V$ implies that $W \succeq_{\mathsf{ln}} V$, which follows from the data processing inequality. Theorem 2 presents a simple sufficient condition for degradation by a q-SC.

Theorem 2 (Sufficient Condition for Degradation by Symmetric Channels)**.** *Given a channel $V \in \mathbb{R}_{\mathsf{sto}}^{q \times q}$ with $q \ge 2$ and minimum entry $\nu = \min\{[V]_{i,j} : 1 \le i, j \le q\}$, we have:*

$$0 \le \delta \le \frac{\nu}{1 - (q-1)\nu + \frac{\nu}{q-1}} \Rightarrow W_\delta \succeq_{\mathsf{deg}} V \Rightarrow W_\delta \succeq_{\mathsf{ln}} V.$$

Proof Sketch. We sketch the proof for the $\nu < \frac{1}{q}$ case here; the remaining details can be found in [14, Section VI]. Consider the q-SC $W_{(q-1)\nu} \in \mathbb{R}_{\mathsf{sto}}^{q \times q}$, and let $w_i \in \mathcal{P}_q$ and $v_i \in \mathcal{P}_q$ for $1 \le i \le q$ denote the ith rows of $W_{(q-1)\nu}$ and V, respectively. Using a majorization argument, we have:

$$\forall i \in \{1, \ldots, q\}, \;\; v_i = \sum_{j=1}^{q} p_{i,j} \, w_j$$

where $\{p_{i,j} \ge 0 : 1 \le i, j \le q\}$ are some convex weights such that $\sum_{j=1}^q p_{i,j} = 1$ for $1 \le i \le q$. Stacking the rows of V back into a matrix, we observe that:

$$V = \sum_{1 \le j_1, \ldots, j_q \le q} \left(\prod_{i=1}^{q} p_{i,j_i}\right) S_{j_1, \ldots, j_q}$$

where for each $1 \le j_1, \ldots, j_q \le q$, $S_{j_1, \ldots, j_q} = \left[w_{j_1}^T \cdots w_{j_q}^T\right]^T$ defines a matrix whose kth row is the j_kth row of $W_{(q-1)\nu}$, and $\{\prod_{i=1}^q p_{i,j_i} : 1 \le j_1, \ldots, j_q \le q\}$ forms a product pmf. Hence, if $\exists\, \delta \in \left[0, \frac{q-1}{q}\right)$ such that for every $1 \le j_1, \ldots, j_q \le q$, $W_\delta \succeq_{\mathsf{deg}} S_{j_1, \ldots, j_q}$, or equivalently $W_\delta^{-1} S_{j_1, \ldots, j_q} \in \mathbb{R}_{\mathsf{sto}}^{q \times q}$, then $W_\delta \succeq_{\mathsf{deg}} V$. The rows of $W_\delta^{-1} S_{j_1, \ldots, j_q}$ sum to unity, because W_δ^{-1} has the form (11) since matrices of the form (11) constitute a group. Moreover, we can verify that the minimum possible entry of $W_\delta^{-1} S_{j_1, \ldots, j_q}$ is non-negative if and only if:

$$0 \le \delta \le \frac{\nu}{1 - (q-1)\,\nu + \frac{\nu}{q-1}}.$$

This a sufficient condition for $W_\delta^{-1} S_{j_1, \ldots, j_q} \in \mathbb{R}_{\mathsf{sto}}^{q \times q}$ for every $1 \le j_1, \ldots, j_q \le q$. Therefore, $W_\delta \succeq_{\mathsf{deg}} V$, which in turn implies that $W_\delta \succeq_{\mathsf{ln}} V$. ■

The condition given in Theorem 2 is tight for degradation. For example, if $V = S_{2,1,\ldots,1} \in \mathbb{R}_{\mathsf{sto}}^{q \times q}$, then $W_\delta \succeq_{\mathsf{deg}} V$ if and only if $0 \le \delta \le \nu / \left(1 - (q-1)\nu + \frac{\nu}{q-1}\right)$ [14, Section VI].

A. Domination of additive noise channels

It is compelling to derive a sufficient condition for $W_\delta \succeq_{\mathsf{ln}} V$ that does not simply ensure $W_\delta \succeq_{\mathsf{deg}} V$. To this end, we study additive noise channels. Let $(\mathcal{X}, \oplus)$ be a finite Abelian group of order $q \ge 2$ equipped with a binary "addition" operation $\oplus$. An *additive noise channel* is defined by the relation:

$$Y = X \oplus N, \quad X \perp\!\!\!\perp N \tag{14}$$

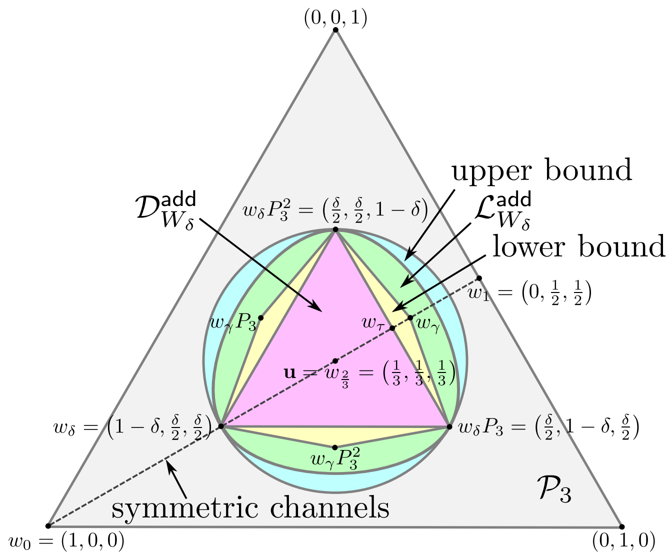

Fig. 1. Additive less noisy domination and degradation regions for a 3-ary symmetric channel with $\delta \in \left(0,\frac{2}{3}\right)$: The gray region denotes the probability simplex of noise pmfs $\mathcal{P}_3$. The dotted line denotes the symmetric channels $\left\{w_\delta = \left(1-\delta,\frac{\delta}{2},\frac{\delta}{2}\right) : \delta \in [0,1]\right\}$. As conveyed in Theorem 3, the yellow and (circular) cyan regions lower and upper bound the less noisy domination region, respectively. Note that $P_3 \in \mathbb{R}^{3\times 3}$ is a cyclic permutation matrix that cyclically shifts an input row vector to the right once, and $\tau = 1-(\delta/2)$ and $\gamma = (1-\delta)/(1-\delta+(\delta/4))$ correspond to the extremal degraded symmetric channel and the symmetric channel from Theorem 3, respectively.

where $X, Y, N \in \mathcal{X}$ are the input, output, and noise random variables respectively, and X is independent of N. The channel transition probability matrix corresponding to (14) is a doubly stochastic *$\mathcal{X}$-circulant matrix* $\mathsf{circ}_{\mathcal{X}}(P_N) \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ determined by the noise pmf $P_N \in \mathcal{P}_q$ [17, Chapter 3E, Section 4]:

$$\forall x, y \in \mathcal{X},\ \left[\mathsf{circ}_{\mathcal{X}}(P_N)\right]_{x,y} \triangleq P_N(-x \oplus y) = P_{Y|X}(y|x) \quad (15)$$

where $-x \in \mathcal{X}$ denotes the inverse of x. $\mathcal{X}$-circulant matrices form a commutative algebra, and are jointly unitarily diagonalizable by a "Fourier" matrix of characters. In the context of various channel symmetries in the literature, additive noise channels correspond to "group-noise" channels, and are input, output, Dobrushin, and Gallager symmetric [18, Section VI.B].

One can verify that a q-SC is an additive noise channel with noise pmf defined by $P_N(0) = 1-\delta$ and $P_N(x) = \delta/(q-1)$ for any $x \in \mathcal{X}\backslash\{0\}$, where $0 \in \mathcal{X}$ is the identity element and $\delta \in [0,1]$. To understand when a q-SC $W_\delta \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ dominates an additive noise channel, we define the *additive less noisy domination region*:

$$\mathcal{L}^{\mathsf{add}}_{W_\delta} \triangleq \{P_N \in \mathcal{P}_q : W_\delta \succeq_{\mathsf{ln}} \mathsf{circ}_{\mathcal{X}}(P_N)\} \quad (16)$$

and the *additive degradation region*:

$$\mathcal{D}^{\mathsf{add}}_{W_\delta} \triangleq \{P_N \in \mathcal{P}_q : W_\delta \succeq_{\mathsf{deg}} \mathsf{circ}_{\mathcal{X}}(P_N)\} \quad (17)$$

corresponding to W_δ. The next theorem exactly characterizes $\mathcal{D}^{\mathsf{add}}_{W_\delta}$, and "bounds" $\mathcal{L}^{\mathsf{add}}_{W_\delta}$ in a set theoretic sense.

Theorem 3 (Domination Regions of Symmetric Channels)**.** *Given $W_\delta \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ with $\delta \in \left[0,\frac{q-1}{q}\right]$ and $q \geq 2$, we have:*

$$\begin{aligned}\mathcal{D}^{\mathsf{add}}_{W_\delta} &= \mathsf{conv}\left(\left\{(W_\delta)_k : 1 \leq k \leq q\right\}\right) \\ &\subseteq \mathsf{conv}\left(\left\{(W_\delta)_k : 1 \leq k \leq q\right\} \cup \left\{(W_\gamma)_k : 1 \leq k \leq q\right\}\right) \\ &\subseteq \mathcal{L}^{\mathsf{add}}_{W_\delta} \subseteq \left\{v \in \mathcal{P}_q : \|v - \boldsymbol{u}\|_{\ell^2} \leq \|(W_\delta)_1 - \boldsymbol{u}\|_{\ell^2}\right\}\end{aligned}$$

where $\gamma = (1-\delta)/\left(1-\delta+\frac{\delta}{(q-1)^2}\right)$, $(W_\delta)_k$ is the kth row of W_δ, $\mathsf{conv}(\cdot)$ denotes convex hull, $\boldsymbol{u} \in \mathcal{P}_q$ is the uniform pmf, and the first set inclusion is strict for $\delta \in \left(0,\frac{q-1}{q}\right)$ and $q \geq 3$. Furthermore, $\mathcal{L}^{\mathsf{add}}_{W_\delta}$ is closed, convex, and invariant under permutations in the regular representation of $(\mathcal{X},\oplus)$.

For an additive noise channel $V \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$, the first set inclusion in Theorem 3 offers a sufficient condition for $W_\delta \succeq_{\mathsf{ln}} V$ without ensuring $W_\delta \succeq_{\mathsf{deg}} V$; it is derived using condition 3 of Theorem 1 [14, Proposition 13]. Theorem 3 is further explicated and proved in [14, Sections III, V], where similar properties of less noisy domination and degradation regions for more general channels, and some necessary conditions for less noisy domination are also established. In particular, degradation among additive noise channels is completely characterized using *group majorization* in [14, Proposition 5]. We remark that according to numerical evidence, the second and third set inclusions in Theorem 3 appear to be strict, and $\mathcal{L}^{\mathsf{add}}_{W_\delta}$ seems to be a strictly convex set. The results of Theorem 3 and these observations are depicted in Figure 1 when $q = 3$.

IV. Comparison of Dirichlet Forms

In this section, we show that proving $W_\delta \succeq_{\mathsf{ln}} V$ for a doubly stochastic channel $V \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ allows us to translate the LSI for W_δ to an LSI for V. Suppose V defines an *irreducible* discrete-time Markov chain with state space $\mathcal{X}$ (such that $|\mathcal{X}| = q$) and unique stationary distribution $\mathbf{u} \in \mathcal{P}_q$. Corresponding to this chain, we may define a continuous-time *Markov semigroup*:

$$\forall t \geq 0,\ \ H_t \triangleq \exp\left(-t\left(I_q - V\right)\right) \in \mathbb{R}^{q\times q}_{\mathsf{sto}} \quad (18)$$

with unique uniform stationary pmf, where $V - I_q$ is the *Laplacian operator* that forms the *generator* of the semigroup. Furthermore, we define the Hilbert space $\mathcal{L}^2(\mathcal{X},\mathbf{u})$ of all real functions with domain $\mathcal{X}$ endowed with the inner product:

$$\forall f, g \in \mathcal{L}^2(\mathcal{X},\mathbf{u}),\ \ \langle f, g\rangle_{\mathbf{u}} \triangleq \frac{1}{q}\sum_{x\in\mathcal{X}} f(x)g(x) = \frac{f^T g}{q} \quad (19)$$

and induced norm $\|\cdot\|_{\mathbf{u}}$, where we treat functions in $\mathcal{L}^2(\mathcal{X},\mathbf{u})$ as column vectors in $\mathbb{R}^q$. To present our result, we define the *Dirichlet form* $\mathcal{E}_V : \mathcal{L}^2(\mathcal{X},\mathbf{u}) \times \mathcal{L}^2(\mathcal{X},\mathbf{u}) \to \mathbb{R}^+$ [11]:

$$\mathcal{E}_V(f,f) \triangleq \langle (I_q - V)\, f, f\rangle_{\mathbf{u}} = \frac{1}{q} f^T\left(I_q - \frac{V+V^T}{2}\right) f \quad (20)$$

which is a quadratic form representing the energy of its input function. A salient specialization of (20) is the so called *standard Dirichlet form* associated to the channel $W_{(q-1)/q} = \mathbf{1u}$:

$$\mathcal{E}_{\mathsf{std}}(f,f) \triangleq \mathbb{VAR}_{\mathbf{u}}(f) = \sum_{x\in\mathcal{X}} \frac{f(x)^2}{q} - \left(\sum_{x\in\mathcal{X}} \frac{f(x)}{q}\right)^2, \quad (21)$$

and the Dirichlet forms corresponding to q-SCs satisfy the relation: $\mathcal{E}_{W_\delta}(f,f) = \frac{q\delta}{q-1}\mathcal{E}_{\mathsf{std}}(f,f)$ for every $f \in \mathcal{L}^2(\mathcal{X},\mathbf{u})$. The next theorem portrays that $W_\delta \succeq_{\mathsf{ln}} V$ implies a pointwise domination of Dirichlet forms: $\mathcal{E}_V \geq \mathcal{E}_{W_\delta} = \frac{q\delta}{q-1}\mathcal{E}_{\mathsf{std}}$.

Theorem 4 (Domination of Dirichlet Forms)**.** *Given the doubly stochastic channels $W_\delta \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ with $\delta \in \left[0, \frac{q-1}{q}\right]$ and $V \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$, if $W_\delta \succeq_{\mathsf{ln}} V$, then:*

$$\forall f \in \mathcal{L}^2(\mathcal{X},\boldsymbol{u}), \;\; \mathcal{E}_V(f,f) \geq \frac{q\delta}{q-1}\mathcal{E}_{\mathsf{std}}(f,f).$$

Proof Sketch. Since $W_\delta \succeq_{\mathsf{ln}} V$, we have $W_\delta^2 = W_\delta W_\delta^T \succeq_{\mathsf{PSD}} VV^T$ from part 3 of Theorem 1 after letting $P_X = \mathbf{u}$. Then, by the Löwner-Heinz theorem (cf. [15, Corollary 7.7.4 (b)]), we get $W_\delta^2 \succeq_{\mathsf{PSD}} VV^T \Rightarrow W_\delta \succeq_{\mathsf{PSD}} (VV^T)^{\frac{1}{2}}$, because W_δ and the Gramian matrix VV^T are positive semidefinite. Due to (20), it is sufficient to prove that: $W_\delta \succeq_{\mathsf{PSD}} (VV^T)^{\frac{1}{2}} \Rightarrow W_\delta \succeq_{\mathsf{PSD}} (V+V^T)/2$. This implication is proved in [14, Section VII] by carefully analyzing the spectral structure of W_δ, $(VV^T)^{\frac{1}{2}}$, and $(V+V^T)/2$, and then applying a corollary of the Courant-Fischer variational characterization of singular values. ■

We note that other variants of Theorem 4 are also presented in [14, Section VII]. A major consequence of the domination of Dirichlet forms in Theorem 4 is that we can immediately establish an LSI for V using the LSI for W_δ. An LSI for the Markov semigroup $\left\{H_t \in \mathbb{R}^{q\times q}_{\mathsf{sto}} : t \geq 0\right\}$ with constant $\alpha \in \mathbb{R}$ states that for all $f \in \mathcal{L}^2(\mathcal{X},\mathbf{u})$ with $\|f\|_{\mathbf{u}} = 1$, we have [11]:

$$D(f^2\mathbf{u}||\mathbf{u}) \leq \frac{1}{\alpha}\,\mathcal{E}_V(f,f) \tag{22}$$

where $\mu = f^2\mathbf{u} \in \mathcal{P}_q$ is a pmf such that $\forall x \in \mathcal{X}$, $\mu(x) = f(x)^2\,\mathbf{u}(x)$, and f^2 behaves like the density of μ with respect to $\mathbf{u}$. The largest constant α such that (22) holds is known as the *logarithmic Sobolev constant* of the Markov chain V:

$$\alpha(V) \triangleq \inf_{f\in\mathcal{F}} \frac{\mathcal{E}_V(f,f)}{D(f^2\mathbf{u}||\mathbf{u})} \tag{23}$$

where $\mathcal{F} = \left\{f \in \mathcal{L}^2(\mathcal{X},\mathbf{u}) : \|f\|_{\mathbf{u}} = 1 \text{ and } D(f^2\mathbf{u}||\mathbf{u}) > 0\right\}$. This constant is closely related to the ergodicity (convergence rate to invariant measure) and hypercontractivity properties of the corresponding Markov semigroup [11].

Typically, it is difficult to analytically compute logarithmic Sobolev constants. However, the logarithmic Sobolev constant corresponding to $\mathcal{E}_{\mathsf{std}}$ has been computed in [11, Appendix, Theorem A.1], and implies the following LSI for $q > 2$:

$$D(f^2\mathbf{u}||\mathbf{u}) \leq \frac{q\log(q-1)}{(q-2)}\,\mathcal{E}_{\mathsf{std}}(f,f) \tag{24}$$

for every $f \in \mathcal{L}^2(\mathcal{X},\mathbf{u})$ satisfying $\|f\|_{\mathbf{u}} = 1$ (where the $q = 2$ case uses the limiting value of 2 in front of $\mathcal{E}_{\mathsf{std}}$). Moreover, we can deduce $\alpha(W_\delta)$ for any q-SC $W_\delta \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ (which is irreducible if $\delta \in (0,1]$) from (24), cf. [14, Proposition 14]. So, we often compare the Dirichlet form of a given irreducible channel $V \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ with that of a q-SC $W_\delta \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ (or the standard Dirichlet form); generalizations of this idea are presented in [11, Lemmata 3.3 and 3.4]. If such a comparison yields a pointwise domination of the form:

$$\forall f \in \mathcal{L}^2(\mathcal{X},\mathbf{u}),\; \mathcal{E}_V(f,f) \geq \mathcal{E}_{W_\delta}(f,f) = \frac{q\delta}{q-1}\mathcal{E}_{\mathsf{std}}(f,f) \tag{25}$$

as shown in Theorem 4, we immediately establish the following LSI for V using (24):

$$D(f^2\mathbf{u}||\mathbf{u}) \leq \frac{(q-1)\log(q-1)}{\delta\,(q-2)}\,\mathcal{E}_V(f,f) \tag{26}$$

for every $f \in \mathcal{L}^2(\mathcal{X},\mathbf{u})$ satisfying $\|f\|_{\mathbf{u}} = 1$. Alternatively, we may perceive (26) as a lower bound on $\alpha(V)$:

$$\alpha(V) \geq \frac{\delta\,(q-2)}{(q-1)\log(q-1)}. \tag{27}$$

Therefore, Theorem 4 illustrates that less noisy domination of a given channel by a q-SC is a sufficient condition for deriving an LSI for the original channel.

References

[1] J. Körner and K. Marton, "Comparison of two noisy channels," in *Topics in Information Theory (Second Colloq., Keszthely, 1975)*, Amsterdam: North-Holland, 1977, pp. 411–423.

[2] A. El Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Inform. Theory*, vol. 25, no. 2, pp. 166–169, March 1979.

[3] C. Nair, "Capacity regions of two new classes of 2-receiver broadcast channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, South Korea, June 28 - July 3 2009, pp. 1839–1843.

[4] Y. Geng, C. Nair, S. Shamai, and Z. V. Wang, "On broadcast channels with binary inputs and symmetric outputs," *IEEE Trans. Inform. Theory*, vol. 59, no. 11, pp. 6980–6989, November 2013.

[5] Y. Polyanskiy and Y. Wu, "Strong data-processing inequalities for channels and Bayesian networks," July 2016, arXiv:1508.06025v4 [cs.IT].

[6] R. Ahlswede and P. Gács, "Spreading of sets in product spaces and hypercontraction of the Markov operator," *Ann. Probab.*, vol. 4, no. 6, pp. 925–939, December 1976.

[7] J. E. Cohen, Y. Iwasa, G. Rautu, M. B. Ruskai, E. Seneta, and G. Zbăganu, "Relative entropy under mappings by stochastic matrices," *Linear Algebra Appl.*, vol. 179, pp. 211–235, January 1993.

[8] M. Raginsky, "Strong data processing inequalities and ϕ-Sobolev inequalities for discrete channels," *IEEE Trans. Inform. Theory*, vol. 62, no. 6, pp. 3355–3389, June 2016.

[9] D. Sutter and J. M. Renes, "Universal polar codes for more capable and less noisy channels and sources," April 2014, arXiv:1312.5990v3 [cs.IT].

[10] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. New York: Cambridge University Press, 2011.

[11] P. Diaconis and L. Saloff-Coste, "Logarithmic Sobolev inequalities for finite Markov chains," *Ann. Appl. Probab.*, vol. 6, no. 3, pp. 695–750, August 1996.

[12] R. Montenegro and P. Tetali, "Mathematical aspects of mixing times in Markov chains," *Found. Trends Theor. Comput. Sci.*, vol. 1, no. 3, 2006.

[13] Y. Polyanskiy and Y. Wu, "Lecture notes on information theory," June 2016, Massachusetts Institute of Technology (Course 6.441) and University of Illinois Urbana-Champaign (ECE Course 563).

[14] A. Makur and Y. Polyanskiy, "Comparison of channels: criteria for domination by a symmetric channel," September 2016, arXiv:1609.06877v1 [cs.IT].

[15] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. New York: Cambridge University Press, 2013.

[16] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. 19, no. 2, pp. 197–207, March 1973.

[17] P. Diaconis, *Group Representations in Probability and Statistics*, S. S. Gupta, Ed. USA: Inst. Math. Stat. Monogr., 1988, vol. 11.

[18] Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," *IEEE Trans. Inform. Theory*, vol. 59, no. 5, pp. 2576–2595, May 2013.