Comparison of channels: criteria for domination by a symmetric channel

Anuran Makur and Yury Polyanskiy*

Abstract

This paper studies the basic question of whether a given channel V can be dominated (in the precise sense of being more noisy) by a q-ary symmetric channel. The concept of "less noisy" relation between channels originated in network information theory (broadcast channels) and is defined in terms of mutual information or Kullback-Leibler divergence. We provide an equivalent characterization in terms of χ^2-divergence. Furthermore, we develop a simple criterion for domination by a q-ary symmetric channel in terms of the minimum entry of the stochastic matrix defining the channel V. The criterion is strengthened for the special case of additive noise channels over finite Abelian groups. Finally, it is shown that domination by a symmetric channel implies (via comparison of Dirichlet forms) a logarithmic Sobolev inequality for the original channel.

Index Terms

Less noisy, degradation, q-ary symmetric channel, additive noise channel, Dirichlet form, logarithmic Sobolev inequalities.

CONTENTS

I Introduction 2
- I-A Preliminaries 2
- I-B Channel preorders in information theory 2
- I-C Symmetric channels and their properties 4
- I-D Main question and motivation 6

II Main results 6
- II-A χ^2-divergence characterization of the less noisy preorder 7
- II-B Less noisy domination by symmetric channels 7
- II-C Structure of additive noise channels 8
- II-D Comparison of Dirichlet forms 8
- II-E Outline 9

III Less noisy domination and degradation regions 10
- III-A Less noisy domination and degradation regions for additive noise channels 11
- III-B Less noisy domination and degradation regions for symmetric channels 12

IV Equivalent characterizations of less noisy preorder 13
- IV-A Characterization using χ^2-divergence 13
- IV-B Characterizations via the Löwner partial order and spectral radius 14

V Conditions for less noisy domination over additive noise channels 16
- V-A Necessary conditions 16
- V-B Sufficient conditions 17

VI Sufficient conditions for degradation over general channels 19

VII Less noisy domination and logarithmic Sobolev inequalities 22

VIII Conclusion 26

*A. Makur and Y. Polyanskiy are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA (e-mail: a_makur@mit.edu; yp@mit.edu).

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

This work was presented at the 2017 IEEE International Symposium on Information Theory (ISIT) [1].

Appendix A: Basics of majorization theory 26

Appendix B: Proofs of propositions 4 and 12 27

Appendix C: Auxiliary results 28

References 30

I. INTRODUCTION

For any Markov chain $U \to X \to Y$, it is well-known that the data processing inequality, $I(U;Y) \leq I(U;X)$, holds. This result can be strengthened to [2]:

$$I(U;Y) \leq \eta I(U;X) \tag{1}$$

where the contraction coefficient $\eta \in [0,1]$ only depends on the channel $P_{Y|X}$. Frequently, one gets $\eta < 1$ and the resulting inequality is called a *strong data processing inequality* (SDPI). Such inequalities have been recently simultaneously rediscovered and applied in several disciplines; see [3, Section 2] for a short survey. In [3, Section 6], it was noticed that the validity of (1) for all $P_{U,X}$ is equivalent to the statement that an erasure channel with erasure probability $1-\eta$ is *less noisy* than the given channel $P_{Y|X}$. In this way, the entire field of SDPIs is equivalent to determining whether a given channel is dominated by an erasure channel.

This paper initiates the study of a natural extension of the concept of SDPI by replacing the distinguished role played by erasure channels with q-ary symmetric channels. We give simple criteria for testing this type of domination and explain how the latter can be used to prove logarithmic Sobolev inequalities. In the next three subsections, we introduce some basic definitions and notation. We state and motivate our main question in subsection I-D, and present our main results in section II.

A. Preliminaries

The following notation will be used in our ensuing discussion. Consider any $q, r \in \mathbb{N} \triangleq \{1,2,3,\dots\}$. We let $\mathbb{R}^{q\times r}$ (respectively $\mathbb{C}^{q\times r}$) denote the set of all real (respectively complex) $q \times r$ matrices. Furthermore, for any matrix $A \in \mathbb{R}^{q\times r}$, we let $A^T \in \mathbb{R}^{r\times q}$ denote the transpose of A, $A^\dagger \in \mathbb{R}^{r\times q}$ denote the *Moore-Penrose pseudoinverse* of A, $\mathcal{R}(A)$ denote the range (or column space) of A, and $\rho(A)$ denote the *spectral radius* of A (which is the maximum of the absolute values of all complex eigenvalues of A) when $q = r$. We let $\mathbb{R}^{q\times q}_{\succeq 0} \subsetneq \mathbb{R}^{q\times q}_{\mathsf{sym}}$ denote the sets of positive semidefinite and symmetric matrices, respectively. In fact, $\mathbb{R}^{q\times q}_{\succeq 0}$ is a closed convex cone (with respect to the Frobenius norm). We also let $\succeq_{\mathsf{PSD}}$ denote the *Löwner partial order* over $\mathbb{R}^{q\times q}_{\mathsf{sym}}$: for any two matrices $A, B \in \mathbb{R}^{q\times q}_{\mathsf{sym}}$, we write $A \succeq_{\mathsf{PSD}} B$ (or equivalently, $A - B \succeq_{\mathsf{PSD}} 0$, where 0 is the zero matrix) if and only if $A - B \in \mathbb{R}^{q\times q}_{\succeq 0}$. To work with probabilities, we let $\mathcal{P}_q \triangleq \{p = (p_1,\dots,p_q) \in \mathbb{R}^q : p_1,\dots,p_q \geq 0 \text{ and } p_1 + \cdots + p_q = 1\}$ be the probability simplex of row vectors in $\mathbb{R}^q$, $\mathcal{P}_q^\circ \triangleq \{p = (p_1,\dots,p_q) \in \mathbb{R}^q : p_1,\dots,p_q > 0 \text{ and } p_1 + \cdots + p_q = 1\}$ be the relative interior of $\mathcal{P}_q$, and $\mathbb{R}^{q\times r}_{\mathsf{sto}}$ be the convex set of row stochastic matrices (which have rows in $\mathcal{P}_r$). Finally, for any (row or column) vector $x = (x_1,\dots,x_q) \in \mathbb{R}^q$, we let $\mathsf{diag}(x) \in \mathbb{R}^{q\times q}$ denote the diagonal matrix with entries $[\mathsf{diag}(x)]_{i,i} = x_i$ for each $i \in \{1,\dots,q\}$, and for any set of vectors $\mathcal{S} \subseteq \mathbb{R}^q$, we let $\mathsf{conv}(\mathcal{S})$ be the convex hull of the vectors in $\mathcal{S}$.

B. Channel preorders in information theory

Since we will study preorders over discrete channels that capture various notions of relative "noisiness" between channels, we provide an overview of some well-known channel preorders in the literature. Consider an input random variable $X \in \mathcal{X}$ and an output random variable $Y \in \mathcal{Y}$, where the alphabets are $\mathcal{X} = [q] \triangleq \{0,1,\dots,q-1\}$ and $\mathcal{Y} = [r]$ for $q, r \in \mathbb{N}$ without loss of generality. We let $\mathcal{P}_q$ be the set of all probability mass functions (pmfs) of X, where every pmf $P_X = (P_X(0),\dots,P_X(q-1)) \in \mathcal{P}_q$ and is perceived as a row vector. Likewise, we let $\mathcal{P}_r$ be the set of all pmfs of Y. A *channel* is the set of conditional distributions $W_{Y|X}$ that associates each $x \in \mathcal{X}$ with a conditional pmf $W_{Y|X}(\cdot|x) \in \mathcal{P}_r$. So, we represent each channel with a stochastic matrix $W \in \mathbb{R}^{q\times r}_{\mathsf{sto}}$ that is defined entry-wise as:

$$\forall x \in \mathcal{X}, \forall y \in \mathcal{Y}, \;\; [W]_{x+1,y+1} \triangleq W_{Y|X}(y|x) \tag{2}$$

where the $(x+1)$th row of W corresponds to the conditional pmf $W_{Y|X}(\cdot|x) \in \mathcal{P}_r$, and each column of W has at least one non-zero entry so that no output alphabet letters are redundant. Moreover, we think of such a channel as a (linear) map $W : \mathcal{P}_q \to \mathcal{P}_r$ that takes any row probability vector $P_X \in \mathcal{P}_q$ to the row probability vector $P_Y = P_X W \in \mathcal{P}_r$.

One of the earliest preorders over channels was the notion of *channel inclusion* proposed by Shannon in [4]. Given two channels $W \in \mathbb{R}^{q\times r}_{\mathsf{sto}}$ and $V \in \mathbb{R}^{s\times t}_{\mathsf{sto}}$ for some $q, r, s, t \in \mathbb{N}$, he stated that W includes V, denoted $W \succeq_{\mathsf{inc}} V$, if there exist a pmf $g \in \mathcal{P}_m$ for some $m \in \mathbb{N}$, and two sets of channels $\{A_k \in \mathbb{R}^{r\times t}_{\textsf{sto}} : k = 1, \dots, m\}$ and $\{B_k \in \mathbb{R}^{s\times q}_{\textsf{sto}} : k = 1, \dots, m\}$, such that:

$$V = \sum_{k=1}^{m} g_k B_k W A_k. \tag{3}$$

Channel inclusion is preserved under channel addition and multiplication (which are defined in [5]), and the existence of a code for V implies the existence of as good a code for W in a probability of error sense [4]. The channel inclusion preorder includes the *input-output degradation* preorder, which can be found in [6], as a special case. Indeed, V is an input-output degraded version of W, denoted $W \succeq_{\textsf{iod}} V$, if there exist channels $A \in \mathbb{R}^{r\times t}_{\textsf{sto}}$ and $B \in \mathbb{R}^{s\times q}_{\textsf{sto}}$ such that $V = BWA$. We will study an even more specialized case of Shannon's channel inclusion known as *degradation* [7], [8].

Definition 1 (Degradation Preorder)**.** *A channel $V \in \mathbb{R}^{q\times s}_{\textsf{sto}}$ is said to be a degraded version of a channel $W \in \mathbb{R}^{q\times r}_{\textsf{sto}}$ with the same input alphabet, denoted $W \succeq_{\textsf{deg}} V$, if $V = WA$ for some channel $A \in \mathbb{R}^{r\times s}_{\textsf{sto}}$.*

We note that when Definition 1 of degradation is applied to general matrices (rather than stochastic matrices), it is equivalent to Definition C.8 of *matrix majorization* in [9, Chapter 15]. Many other generalizations of the majorization preorder over vectors (briefly introduced in Appendix A) that apply to matrices are also presented in [9, Chapter 15].

Körner and Marton defined two other preorders over channels in [10] known as the *more capable* and *less noisy* preorders. While the original definitions of these preorders explicitly reflect their significance in channel coding, we will define them using equivalent mutual information characterizations proved in [10]. (See [11, Problems 6.16-6.18] for more on the relationship between channel coding and some of the aforementioned preorders.) We say a channel $W \in \mathbb{R}^{q\times r}_{\textsf{sto}}$ is *more capable* than a channel $V \in \mathbb{R}^{q\times s}_{\textsf{sto}}$ with the same input alphabet, denoted $W \succeq_{\textsf{mc}} V$, if $I(P_X, W_{Y|X}) \geq I(P_X, V_{Y|X})$ for every input pmf $P_X \in \mathcal{P}_q$, where $I(P_X, W_{Y|X})$ denotes the mutual information of the joint pmf defined by P_X and $W_{Y|X}$. The next definition presents the less noisy preorder, which will be a key player in our study.

Definition 2 (Less Noisy Preorder)**.** *Given two channels $W \in \mathbb{R}^{q\times r}_{\textsf{sto}}$ and $V \in \mathbb{R}^{q\times s}_{\textsf{sto}}$ with the same input alphabet, let Y_W and Y_V denote the output random variables of W and V, respectively. Then, W is less noisy than V, denoted $W \succeq_{\textsf{ln}} V$, if $I(U; Y_W) \geq I(U; Y_V)$ for every joint distribution $P_{U,X}$, where the random variable $U \in \mathcal{U}$ has some arbitrary range $\mathcal{U}$, and $U \to X \to (Y_W, Y_V)$ forms a Markov chain.*

An analogous characterization of the less noisy preorder using Kullback-Leibler (KL) divergence or relative entropy is given in the next proposition.

Proposition 1 (KL Divergence Characterization of Less Noisy [10])**.** *Given two channels $W \in \mathbb{R}^{q\times r}_{\textsf{sto}}$ and $V \in \mathbb{R}^{q\times s}_{\textsf{sto}}$ with the same input alphabet, $W \succeq_{\textsf{ln}} V$ if and only if $D(P_X W||Q_X W) \geq D(P_X V||Q_X V)$ for every pair of input pmfs $P_X, Q_X \in \mathcal{P}_q$, where $D(\cdot||\cdot)$ denotes the KL divergence.*[1]

We will primarily use this KL divergence characterization of $\succeq_{\textsf{ln}}$ in our discourse because of its simplicity. Another well-known equivalent characterization of $\succeq_{\textsf{ln}}$ due to van Dijk is presented below, cf. [12, Theorem 2]. We will derive some useful corollaries from it later in subsection IV-B.

Proposition 2 (van Dijk Characterization of Less Noisy [12])**.** *Given two channels $W \in \mathbb{R}^{q\times r}_{\textsf{sto}}$ and $V \in \mathbb{R}^{q\times s}_{\textsf{sto}}$ with the same input alphabet, consider the functional $F : \mathcal{P}_q \to \mathbb{R}$:*

$$\forall P_X \in \mathcal{P}_q, \;\; F(P_X) \triangleq I(P_X, W_{Y|X}) - I(P_X, V_{Y|X}).$$

Then, $W \succeq_{\textsf{ln}} V$ if and only if F is concave.

The more capable and less noisy preorders have both been used to study the capacity regions of broadcast channels. We refer readers to [13]–[15], and the references therein for further details. We also remark that the more capable and less noisy preorders tensorize, as shown in [11, Problem 6.18] and [3, Proposition 16], [16, Proposition 5], respectively.

On the other hand, these preorders exhibit rather counter-intuitive behavior in the context of Bayesian networks (or directed graphical models). Consider a Bayesian network with "source" nodes (with no inbound edges) X and "sink" nodes (with no outbound edges) Y. If we select a node Z in this network and replace the channel from the parents of Z to Z with a less noisy channel, then we may reasonably conjecture that the channel from X to Y also becomes less noisy (motivated by the results in [3]). However, this conjecture is false. To see this, consider the Bayesian network in

[1]Throughout this paper, we will adhere to the convention that $\infty \geq \infty$ is true. So, $D(P_X W||Q_X W) \geq D(P_X V||Q_X V)$ is not violated when both KL divergences are infinity.

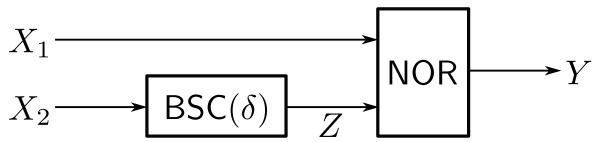

Fig. 1. Illustration of a Bayesian network where $X_1, X_2, Z, Y \in \{0,1\}$ are binary random variables, $P_{Z|X_2}$ is a $\mathsf{BSC}(\delta)$ with $\delta \in (0,1)$, and $P_{Y|X_1,Z}$ is defined by a deterministic NOR gate.

Figure 1 (inspired by the results in [17]), where the source nodes are $X_1 \sim \text{Ber}\left(\frac{1}{2}\right)$ and $X_2 = 1$ (almost surely), the node Z is the output of a binary symmetric channel (BSC) with crossover probability $\delta \in (0,1)$, denoted $\mathsf{BSC}(\delta)$, and the sink node Y is the output of a NOR gate. Let $I(\delta) = I(X_1, X_2; Y)$ be the end-to-end mutual information. Then, although $\mathsf{BSC}(0) \succeq_{\mathsf{ln}} \mathsf{BSC}(\delta)$ for $\delta \in (0,1)$, it is easy to verify that $I(\delta) > I(0) = 0$. So, when we replace the $\mathsf{BSC}(\delta)$ with a less noisy $\mathsf{BSC}(0)$, the end-to-end channel does *not* become less noisy (or more capable).

The next proposition illustrates certain well-known relationships between the various preorders discussed in this subsection.

Proposition 3 (Relations between Channel Preorders)**.** *Given two channels $W \in \mathbb{R}_{\mathsf{sto}}^{q \times r}$ and $V \in \mathbb{R}_{\mathsf{sto}}^{q \times s}$ with the same input alphabet, we have:*

1) $W \succeq_{\mathsf{deg}} V \;\Rightarrow\; W \succeq_{\mathsf{iod}} V \;\Rightarrow\; W \succeq_{\mathsf{inc}} V$,
2) $W \succeq_{\mathsf{deg}} V \;\Rightarrow\; W \succeq_{\mathsf{ln}} V \;\Rightarrow\; W \succeq_{\mathsf{mc}} V$.

These observations follow in a straightforward manner from the definitions of the various preorders. Perhaps the only nontrivial implication is $W \succeq_{\mathsf{deg}} V \Rightarrow W \succeq_{\mathsf{ln}} V$, which can be proven using Proposition 1 and the data processing inequality.

C. Symmetric channels and their properties

We next formally define q-ary symmetric channels and convey some of their properties. To this end, we first introduce some properties of Abelian groups and define additive noise channels. Let us fix some $q \in \mathbb{N}$ with $q \geq 2$ and consider an *Abelian group* $(\mathcal{X}, \oplus)$ of order q equipped with a binary "addition" operation denoted by $\oplus$. Without loss of generality, we let $\mathcal{X} = [q]$, and let 0 denote the identity element. This endows an ordering to the elements of $\mathcal{X}$. Each element $x \in \mathcal{X}$ permutes the entries of the row vector $(0, \dots, q-1)$ to $(\sigma_x(0), \dots, \sigma_x(q-1))$ by (left) addition in the Cayley table of the group, where $\sigma_x : [q] \rightarrow [q]$ denotes a permutation of $[q]$, and $\sigma_x(y) = x \oplus y$ for every $y \in \mathcal{X}$. So, corresponding to each $x \in \mathcal{X}$, we can define a permutation matrix $P_x \triangleq \left[e_{\sigma_x(0)} \cdots e_{\sigma_x(q-1)}\right] \in \mathbb{R}^{q \times q}$ such that:

$$\left[v_0 \;\cdots\; v_{q-1}\right] P_x = \left[v_{\sigma_x(0)} \;\cdots\; v_{\sigma_x(q-1)}\right] \tag{4}$$

for any $v_0, \dots, v_{q-1} \in \mathbb{R}$, where for each $i \in [q]$, $e_i \in \mathbb{R}^q$ is the ith standard basis column vector with unity in the $(i+1)$th position and zero elsewhere. The permutation matrices $\{P_x \in \mathbb{R}^{q \times q} : x \in \mathcal{X}\}$ (with the matrix multiplication operation) form a group that is isomorphic to $(\mathcal{X}, \oplus)$ (see *Cayley's theorem*, and permutation and regular representations of groups in [18, Sections 6.11, 7.1, 10.6]). In particular, these matrices commute as $(\mathcal{X}, \oplus)$ is Abelian, and are jointly unitarily diagonalizable by a *Fourier matrix of characters* (using [19, Theorem 2.5.5]). We now recall that given a row vector $x = (x_0, \dots, x_{q-1}) \in \mathbb{R}^q$, we may define a corresponding *$\mathcal{X}$-circulant matrix*, $\mathsf{circ}_{\mathcal{X}}(x) \in \mathbb{R}^{q \times q}$, that is defined entry-wise as [20, Chapter 3E, Section 4]:

$$\forall a, b \in [q], \;\; \left[\mathsf{circ}_{\mathcal{X}}(x)\right]_{a+1,b+1} \triangleq x_{-a \oplus b}. \tag{5}$$

where $-a \in \mathcal{X}$ denotes the inverse of $a \in \mathcal{X}$. Moreover, we can decompose this $\mathcal{X}$-circulant matrix as:

$$\mathsf{circ}_{\mathcal{X}}(x) = \sum_{i=0}^{q-1} x_i P_i^T \tag{6}$$

since $\sum_{i=0}^{q-1} x_i \left[P_i^T\right]_{a+1,b+1} = \sum_{i=0}^{q-1} x_i \left[e_{\sigma_i(a)}\right]_{b+1} = x_{-a \oplus b}$ for every $a, b \in [q]$. Using similar reasoning, we can write:

$$\mathsf{circ}_{\mathcal{X}}(x) = \left[P_0 y \;\cdots\; P_{q-1} y\right] = \left[P_0 x^T \;\cdots\; P_{q-1} x^T\right]^T \tag{7}$$

where $y = \left[x_0 \;\; x_{-1} \cdots x_{-(q-1)}\right]^T \in \mathbb{R}^q$, and $P_0 = I_q \in \mathbb{R}^{q \times q}$ is the $q \times q$ identity matrix. Using (6), we see that $\mathcal{X}$-circulant matrices are normal, form a commutative algebra, and are jointly unitarily diagonalizable by a Fourier matrix.

Furthermore, given two row vectors $x, y \in \mathbb{R}^q$, we can define $x \,\mathsf{circ}_{\mathcal{X}}(y) = y \,\mathsf{circ}_{\mathcal{X}}(x)$ as the *$\mathcal{X}$-circular convolution* of x and y, where the commutativity of $\mathcal{X}$-circular convolution follows from the commutativity of $\mathcal{X}$-circulant matrices.

A salient specialization of this discussion is the case where $\oplus$ is addition modulo q, and $(\mathcal{X} = [q], \oplus)$ is the cyclic Abelian group $\mathbb{Z}/q\mathbb{Z}$. In this scenario, $\mathcal{X}$-circulant matrices correspond to the standard circulant matrices which are jointly unitarily diagonalized by the *discrete Fourier transform* (DFT) matrix. Furthermore, for each $x \in [q]$, the permutation matrix $P_x^T = P_q^x$, where $P_q \in \mathbb{R}^{q\times q}$ is the *generator cyclic permutation matrix* as presented in [19, Section 0.9.6]:

$$\forall a, b \in [q], \;\; [P_q]_{a+1,b+1} \triangleq \Delta_{1,(b-a \,(\mathsf{mod}\, q))} \tag{8}$$

where $\Delta_{i,j}$ is the Kronecker delta function, which is unity if $i = j$ and zero otherwise. The matrix P_q cyclically shifts any input row vector to the right once, i.e. $(x_1, x_2, \ldots, x_q)\, P_q = (x_q, x_1, \ldots, x_{q-1})$.

Let us now consider a channel with common input and output alphabet $\mathcal{X} = \mathcal{Y} = [q]$, where $(\mathcal{X}, \oplus)$ is an Abelian group. Such a channel operating on an Abelian group is called an *additive noise channel* when it is defined as:

$$Y = X \oplus Z \tag{9}$$

where $X \in \mathcal{X}$ is the input random variable, $Y \in \mathcal{X}$ is the output random variable, and $Z \in \mathcal{X}$ is the additive noise random variable that is independent of X with pmf $P_Z = (P_Z(0), \ldots, P_Z(q-1)) \in \mathcal{P}_q$. The channel transition probability matrix corresponding to (9) is the $\mathcal{X}$-circulant stochastic matrix $\mathsf{circ}_{\mathcal{X}}(P_Z) \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$, which is also doubly stochastic (i.e. both $\mathsf{circ}_{\mathcal{X}}(P_Z)\,, \mathsf{circ}_{\mathcal{X}}(P_Z)^T \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$). Indeed, for an additive noise channel, it is well-known that the pmf of Y, $P_Y \in \mathcal{P}_q$, can be obtained from the pmf of X, $P_X \in \mathcal{P}_q$, by $\mathcal{X}$-circular convolution: $P_Y = P_X \,\mathsf{circ}_{\mathcal{X}}(P_Z)$. We remark that in the context of various channel symmetries in the literature (see [21, Section VI.B] for a discussion), additive noise channels correspond to "group-noise" channels, and are input symmetric, output symmetric, Dobrushin symmetric, and Gallager symmetric.

The *q-ary symmetric channel* is an additive noise channel on the Abelian group $(\mathcal{X}, \oplus)$ with noise pmf $P_Z = w_\delta \triangleq (1-\delta, \delta/(q-1), \ldots, \delta/(q-1)) \in \mathcal{P}_q$, where $\delta \in [0,1]$. Its channel transition probability matrix is denoted $W_\delta \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$:

$$W_\delta \triangleq \mathsf{circ}_{\mathcal{X}}(w_\delta) = \left[{w_\delta}^T \;\; P_q^T {w_\delta}^T \cdots \left(P_q^T\right)^{q-1} {w_\delta}^T \right]^T \tag{10}$$

which has $1-\delta$ in the principal diagonal entries and $\delta/(q-1)$ in all other entries regardless of the choice of group $(\mathcal{X}, \oplus)$. We may interpret δ as the total crossover probability of the symmetric channel. Indeed, when $q = 2$, W_δ represents a BSC with crossover probability $\delta \in [0,1]$. Although W_δ is only stochastic when $\delta \in [0,1]$, we will refer to the parametrized convex set of matrices $\left\{W_\delta \in \mathbb{R}^{q\times q}_{\mathsf{sym}} : \delta \in \mathbb{R}\right\}$ with parameter δ as the "symmetric channel matrices," where each W_δ has the form (10) such that every row and column sums to unity. We conclude this subsection with a list of properties of symmetric channel matrices.

Proposition 4 (Properties of Symmetric Channel Matrices)**.** *The symmetric channel matrices, $\left\{W_\delta \in \mathbb{R}^{q\times q}_{\mathsf{sym}} : \delta \in \mathbb{R}\right\}$, satisfy the following properties:*

1) *For every $\delta \in \mathbb{R}$, W_δ is a symmetric circulant matrix.*
2) *The DFT matrix $F_q \in \mathbb{C}^{q\times q}$, which is defined entry-wise as $[F_q]_{j,k} = q^{-1/2}\omega^{(j-1)(k-1)}$ for $1 \leq j,k \leq q$ where $\omega = \exp(2\pi i/q)$ and $i = \sqrt{-1}$, jointly diagonalizes W_δ for every $\delta \in \mathbb{R}$. Moreover, the corresponding eigenvalues or Fourier coefficients, $\{\lambda_j(W_\delta) = \left[F_q^H W_\delta F_q\right]_{j,j} : j = 1, \ldots, q\}$ are real:*
$$\lambda_j(W_\delta) = \begin{cases} 1, & j = 1 \\ 1 - \delta - \frac{\delta}{q-1}, & j = 2, \ldots, q \end{cases}$$
where F_q^H denotes the Hermitian transpose of F_q.
3) *For all $\delta \in [0,1]$, W_δ is a doubly stochastic matrix that has the uniform pmf $\boldsymbol{u} \triangleq (1/q, \ldots, 1/q)$ as its stationary distribution: $\boldsymbol{u}W_\delta = \boldsymbol{u}$.*
4) *For every $\delta \in \mathbb{R}\backslash\left\{\frac{q-1}{q}\right\}$, $W_\delta^{-1} = W_\tau$ with $\tau = -\delta/\left(1-\delta-\frac{\delta}{q-1}\right)$, and for $\delta = \frac{q-1}{q}$, $W_\delta = \frac{1}{q}\boldsymbol{1}\boldsymbol{1}^T$ is unit rank and singular, where $\boldsymbol{1} = [1 \cdots 1]^T$.*
5) *The set $\left\{W_\delta \in \mathbb{R}^{q\times q}_{\mathsf{sym}} : \delta \in \mathbb{R}\backslash\left\{\frac{q-1}{q}\right\}\right\}$ with the operation of matrix multiplication is an Abelian group.*

Proof. See Appendix B. ∎

D. Main question and motivation

As we mentioned at the outset, our work is partly motivated by [3, Section 6], where the authors demonstrate an intriguing relation between less noisy domination by an erasure channel and the contraction coefficient of the SDPI (1). For a common input alphabet $\mathcal{X} = [q]$, consider a channel $V \in \mathbb{R}_{\mathsf{sto}}^{q\times s}$ and a q*-ary erasure channel* $E_\epsilon \in \mathbb{R}_{\mathsf{sto}}^{q\times (q+1)}$ with erasure probability $\epsilon \in [0, 1]$. Recall that given an input $x \in \mathcal{X}$, a q-ary erasure channel erases x and outputs e (erasure symbol) with probability ϵ, and outputs x itself with probability $1 - \epsilon$; the output alphabet of the erasure channel is $\{\mathsf{e}\} \cup \mathcal{X}$. It is proved in [3, Proposition 15] that $E_\epsilon \succeq_{\mathsf{ln}} V$ if and only if $\eta_{\mathsf{KL}}(V) \leq 1 - \epsilon$, where $\eta_{\mathsf{KL}}(V) \in [0, 1]$ is the *contraction coefficient* for KL divergence:

$$\eta_{\mathsf{KL}}(V) \triangleq \sup_{\substack{P_X, Q_X \in \mathcal{P}_q \\ 0 < D(P_X||Q_X) < +\infty}} \frac{D\left(P_X V || Q_X V\right)}{D\left(P_X || Q_X\right)} \tag{11}$$

which equals the best possible constant η in the SDPI (1) when $V = P_{Y|X}$ (see [3, Theorem 4] and the references therein). This result illustrates that the q-ary erasure channel E_ϵ with the largest erasure probability $\epsilon \in [0, 1]$ (or the smallest channel capacity) that is less noisy than V has $\epsilon = 1 - \eta_{\mathsf{KL}}(V)$.[2] Furthermore, there are several simple upper bounds on η_{KL} that provide sufficient conditions for such less noisy domination. For example, if the ℓ^1-distances between the rows of V are bounded by $2(1 - \alpha)$ for some $\alpha \in [0, 1]$, then $\eta_{\mathsf{KL}} \leq 1 - \alpha$, cf. [22]. Another criterion follows from *Doeblin minorization* [23, Remark III.2]: if for some pmf $p \in \mathcal{P}_s$ and some $\alpha \in (0, 1)$, $V \geq \alpha\, \mathbf{1} p$ entry-wise, then $E_\alpha \succeq_{\mathsf{deg}} V$ and $\eta_{\mathsf{KL}}(V) \leq 1 - \alpha$.

To extend these ideas, we consider the following question: *What is the q-ary symmetric channel W_δ with the largest value of $\delta \in \left[0, \frac{q-1}{q}\right]$ (or the smallest channel capacity) such that $W_\delta \succeq_{\mathsf{ln}} V$?*[3] Much like the bounds on η_{KL} in the erasure channel context, the goal of this paper is to address this question by establishing simple criteria for testing $\succeq_{\mathsf{ln}}$ domination by a q-ary symmetric channel. We next provide several other reasons why determining whether a q-ary symmetric channel dominates a given channel V is interesting.

Firstly, if $W \succeq_{\mathsf{ln}} V$, then $W^{\otimes n} \succeq_{\mathsf{ln}} V^{\otimes n}$ (where $W^{\otimes n}$ is the n-fold tensor product of W) since $\succeq_{\mathsf{ln}}$ tensorizes, and $I(U; Y_W^n) \geq I(U; Y_V^n)$ for every Markov chain $U \to X^n \to (Y_W^n, Y_V^n)$ (see Definition 2). Thus, many impossibility results (in statistical decision theory for example) that are proven by exhibiting bounds on quantities such as $I(U; Y_W^n)$ transparently carry over to statistical experiments with observations on the basis of Y_V^n. Since it is common to study the q-ary symmetric observation model (especially with $q = 2$), we can leverage its sample complexity lower bounds for other V.

Secondly, we present a self-contained information theoretic motivation. $W \succeq_{\mathsf{ln}} V$ if and only if $C_S = 0$, 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 [24, Corollary 3], [11, Corollary 17.11]. Therefore, finding the maximally noisy q-ary symmetric channel that dominates V establishes the minimal noise required on the eavesdropper link so that secret communication is feasible.

Thirdly, $\succeq_{\mathsf{ln}}$ domination turns out to entail a comparison of Dirichlet forms (see subsection II-D), and consequently, allows us to prove *Poincaré and logarithmic Sobolev inequalities* for V from well-known results on q-ary symmetric channels. These inequalities are cornerstones of the modern approach to Markov chains and concentration of measure [25], [26].

II. Main Results

In this section, we first delineate some guiding sub-questions of our study, indicate the main results that address them, and then present these results in the ensuing subsections. We will delve into the following four leading questions:

1) *Can we test the less noisy preorder $\succeq_{\mathsf{ln}}$ without using KL divergence?*
   Yes, we can use χ^2-divergence as shown in Theorem 1.
2) *Given a channel $V \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$, is there a simple sufficient condition for less noisy domination by a q-ary symmetric channel $W_\delta \succeq_{\mathsf{ln}} V$?*
   Yes, a condition using degradation (which implies less noisy domination) is presented in Theorem 2.
3) *Can we say anything stronger about less noisy domination by a q-ary symmetric channel when $V \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$ is an additive noise channel?*
   Yes, Theorem 3 outlines the structure of additive noise channels in this context (and Figure 2 depicts it).

[2] A q-ary erasure channel E_ϵ with erasure probability $\epsilon \in [0, 1]$ has channel capacity $C(\epsilon) = \log(q)(1 - \epsilon)$, which is linear and decreasing.

[3] A q-ary symmetric channel W_δ with total crossover probability $\delta \in \left[0, \frac{q-1}{q}\right]$ has channel capacity $C(\delta) = \log(q) - H(w_\delta)$, which is convex and decreasing. Here, $H(w_\delta)$ denotes the Shannon entropy of the pmf w_δ.

4) *Why are we interested in less noisy domination by q-ary symmetric channels?*
Because this permits us to compare Dirichlet forms as portrayed in Theorem 4.

We next elaborate on these aforementioned theorems.

A. χ^2-divergence characterization of the less noisy preorder

Our most general result illustrates that although less noisy domination is a preorder defined using KL divergence, one can equivalently define it using χ^2-divergence. Since we will prove this result for general measurable spaces, we introduce some notation pertinent only to this result. Let $(\mathcal{X}, \mathcal{F})$, $(\mathcal{Y}_1, \mathcal{H}_1)$, and $(\mathcal{Y}_2, \mathcal{H}_2)$ be three measurable spaces, and let $W : \mathcal{H}_1 \times \mathcal{X} \to [0, 1]$ and $V : \mathcal{H}_2 \times \mathcal{X} \to [0, 1]$ be two *Markov kernels* (or channels) acting on the same source space $(\mathcal{X}, \mathcal{F})$. Given any probability measure P_X on $(\mathcal{X}, \mathcal{F})$, we denote by $P_X W$ the probability measure on $(\mathcal{Y}_1, \mathcal{H}_1)$ induced by the push-forward of P_X.[4] Recall that for any two probability measures P_X and Q_X on $(\mathcal{X}, \mathcal{F})$, their KL divergence is given by:

$$D(P_X||Q_X) \triangleq \begin{cases} \displaystyle\int_{\mathcal{X}} \log\left(\frac{dP_X}{dQ_X}\right) dP_X, & \text{if } P_X \ll Q_X \\ +\infty, & \text{otherwise} \end{cases} \tag{12}$$

and their χ^2-*divergence* is given by:

$$\chi^2(P_X||Q_X) \triangleq \begin{cases} \displaystyle\int_{\mathcal{X}} \left(\frac{dP_X}{dQ_X}\right)^2 dQ_X - 1, & \text{if } P_X \ll Q_X \\ +\infty, & \text{otherwise} \end{cases} \tag{13}$$

where $P_X \ll Q_X$ denotes that P_X is absolutely continuous with respect to Q_X, $\frac{dP_X}{dQ_X}$ denotes the Radon-Nikodym derivative of P_X with respect to Q_X, and $\log(\cdot)$ is the natural logarithm with base e (throughout this paper). Furthermore, the characterization of $\succeq_{\mathsf{ln}}$ in Proposition 1 extends naturally to general Markov kernels; indeed, $W \succeq_{\mathsf{ln}} V$ if and only if $D(P_X W||Q_X W) \geq D(P_X V||Q_X V)$ for every pair of probability measures P_X and Q_X on $(\mathcal{X}, \mathcal{F})$. The next theorem presents the χ^2-divergence characterization of $\succeq_{\mathsf{ln}}$.

Theorem 1 (χ^2-Divergence Characterization of $\succeq_{\mathsf{ln}}$)**.** *For any Markov kernels $W : \mathcal{H}_1 \times \mathcal{X} \to [0, 1]$ and $V : \mathcal{H}_2 \times \mathcal{X} \to [0, 1]$ acting on the same source space, $W \succeq_{\mathsf{ln}} V$ if and only if:*

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

for every pair of probability measures P_X and Q_X on $(\mathcal{X}, \mathcal{F})$.

Theorem 1 is proved in subsection IV-A.

B. Less noisy domination by symmetric channels

Our remaining results are all concerned with less noisy (and degraded) domination by q-ary symmetric channels. Suppose we are given a q-ary symmetric channel $W_\delta \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ with $\delta \in [0, 1]$, and another channel $V \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ with common input and output alphabets. Then, the next result provides a sufficient condition for when $W_\delta \succeq_{\mathsf{deg}} V$.

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

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

Theorem 2 is proved in section VI. We note that the sufficient condition in Theorem 2 is tight as there exist channels V that violate $W_\delta \succeq_{\mathsf{deg}} V$ when $\delta > \nu/(1 - (q-1)\nu + \frac{\nu}{q-1})$. Furthermore, Theorem 2 also provides a sufficient condition for $W_\delta \succeq_{\mathsf{ln}} V$ due to Proposition 3.

[4]Here, we can think of X and Y as random variables with codomains $\mathcal{X}$ and $\mathcal{Y}$, respectively. The Markov kernel W behaves like the conditional distribution of Y given X (under regularity conditions). Moreover, when the distribution of X is P_X, the corresponding distribution of Y is $P_Y = P_X W$.

C. Structure of additive noise channels

Our next major result is concerned with understanding when q-ary symmetric channels operating on an Abelian group $(\mathcal{X}, \oplus)$ dominate other additive noise channels on $(\mathcal{X}, \oplus)$, which are defined in (9), in the less noisy and degraded senses. Given a symmetric channel $W_\delta \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$ with $\delta \in [0,1]$, we define the *additive less noisy domination region* of W_δ as:

$$\mathcal{L}_{W_\delta}^{\mathsf{add}} \triangleq \left\{v \in \mathcal{P}_q : W_\delta = \mathsf{circ}_{\mathcal{X}}(w_\delta) \succeq_{\mathsf{ln}} \mathsf{circ}_{\mathcal{X}}(v)\right\} \tag{14}$$

which is the set of all noise pmfs whose corresponding channel transition probability matrices are dominated by W_δ in the less noisy sense. Likewise, we define the *additive degradation region* of W_δ as:

$$\mathcal{D}_{W_\delta}^{\mathsf{add}} \triangleq \left\{v \in \mathcal{P}_q : W_\delta = \mathsf{circ}_{\mathcal{X}}(w_\delta) \succeq_{\mathsf{deg}} \mathsf{circ}_{\mathcal{X}}(v)\right\} \tag{15}$$

which is the set of all noise pmfs whose corresponding channel transition probability matrices are degraded versions of W_δ. The next theorem exactly characterizes $\mathcal{D}_{W_\delta}^{\mathsf{add}}$, and "bounds" $\mathcal{L}_{W_\delta}^{\mathsf{add}}$ in a set theoretic sense.

Theorem 3 (Additive Less Noisy Domination and Degradation Regions for Symmetric Channels)**.** *Given a symmetric channel $W_\delta = \mathsf{circ}_{\mathcal{X}}(w_\delta) \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$ with $\delta \in \left[0, \frac{q-1}{q}\right]$ and $q \geq 2$, we have:*

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

where the first set inclusion is strict for $\delta \in \left(0, \frac{q-1}{q}\right)$ and $q \geq 3$, P_q denotes the generator cyclic permutation matrix as defined in (8), $\mathbf{u}$ denotes the uniform pmf, $\|\cdot\|_{\ell^2}$ is the Euclidean ℓ^2-norm, and:

$$\gamma = \frac{1-\delta}{1-\delta+\frac{\delta}{(q-1)^2}}.$$

Furthermore, $\mathcal{L}_{W_\delta}^{\mathsf{add}}$ is a closed and convex set that is invariant under the permutations $\{P_x \in \mathbb{R}^{q\times q} : x \in \mathcal{X}\}$ defined in (4) corresponding to the underlying Abelian group $(\mathcal{X}, \oplus)$ (i.e. $v \in \mathcal{L}_{W_\delta}^{\mathsf{add}} \Rightarrow vP_x \in \mathcal{L}_{W_\delta}^{\mathsf{add}}$ for every $x \in \mathcal{X}$).

Theorem 3 is a compilation of several results. As explained at the very end of subsection V-B, Proposition 6 (in subsection III-A), Corollary 1 (in subsection III-B), part 1 of Proposition 9 (in subsection V-A), and Proposition 11 (in subsection V-B) make up Theorem 3. We remark that according to numerical evidence, the second and third set inclusions in Theorem 3 appear to be strict, and $\mathcal{L}_{W_\delta}^{\mathsf{add}}$ seems to be a strictly convex set. The content of Theorem 3 and these observations are illustrated in Figure 2, which portrays the probability simplex of noise pmfs for $q = 3$ and the pertinent regions which capture less noisy domination and degradation by a q-ary symmetric channel.

D. Comparison of Dirichlet forms

As mentioned in subsection I-D, one of the reasons we study q-ary symmetric channels and prove Theorems 2 and 3 is because less noisy domination implies useful bounds between Dirichlet forms. Recall that the q-ary symmetric channel $W_\delta \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$ with $\delta \in [0,1]$ has uniform stationary distribution $\mathbf{u} \in \mathcal{P}_q$ (see part 3 of Proposition 4). For any channel $V \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$ that is doubly stochastic and has uniform stationary distribution, we may define a corresponding *Dirichlet form*:

$$\forall f \in \mathbb{R}^q,\ \ \mathcal{E}_V(f,f) = \frac{1}{q} f^T (I_q - V) f \tag{16}$$

where $f = [f_1 \cdots f_q]^T \in \mathbb{R}^q$ are column vectors, and $I_q \in \mathbb{R}^{q\times q}$ denotes the $q \times q$ identity matrix (as shown in [25] or [26]). Our final theorem portrays that $W_\delta \succeq_{\mathsf{ln}} V$ implies that the Dirichlet form corresponding to V dominates the Dirichlet form corresponding to W_δ pointwise. The Dirichlet form corresponding to W_δ is in fact a scaled version of the so called *standard Dirichlet form*:

$$\forall f \in \mathbb{R}^q,\ \ \mathcal{E}_{\mathsf{std}}(f,f) \triangleq \mathbb{VAR}_{\mathbf{u}}(f) = \frac{1}{q}\sum_{k=1}^{q} f_k^2 - \left(\frac{1}{q}\sum_{k=1}^{q} f_k\right)^2 \tag{17}$$

which is the Dirichlet form corresponding to the q-ary symmetric channel $W_{(q-1)/q} = \mathbf{1u}$ with all uniform conditional pmfs. Indeed, using $I_q - W_\delta = \frac{q\delta}{q-1}(I_q - \mathbf{1u})$, we have:

$$\forall f \in \mathbb{R}^q,\ \ \mathcal{E}_{W_\delta}(f,f) = \frac{q\delta}{q-1}\mathcal{E}_{\mathsf{std}}(f,f). \tag{18}$$

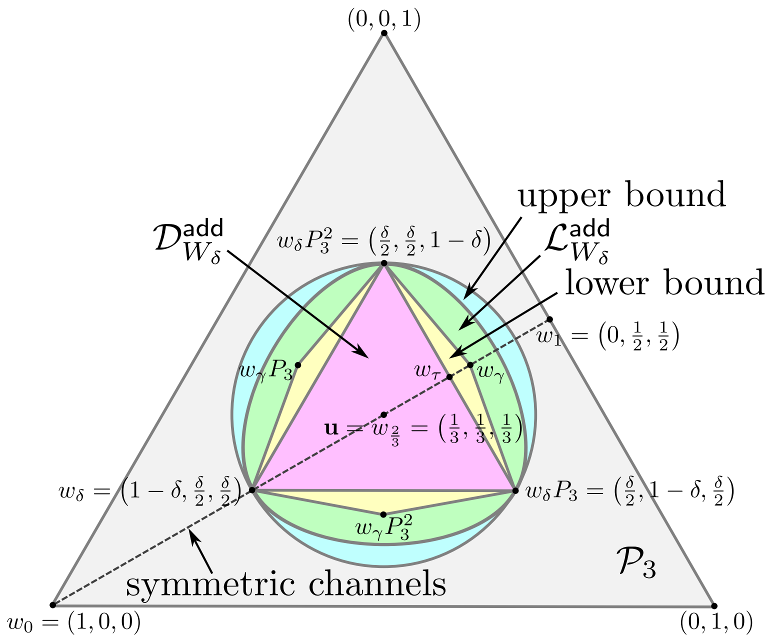

Fig. 2. Illustration of the additive less noisy domination region and additive degradation region for a q-ary symmetric channel when $q = 3$ and $\delta \in (0, 2/3)$: The gray triangle denotes the probability simplex of noise pmfs $\mathcal{P}_3$. The dotted line denotes the parametrized family of noise pmfs of 3-ary symmetric channels $\{w_\delta \in \mathcal{P}_3 : \delta \in [0,1]\}$; its noteworthy points are w_0 (corner of simplex, W_0 is less noisy than every channel), w_δ for some fixed $\delta \in (0, 2/3)$ (noise pmf of 3-ary symmetric channel W_δ under consideration), $w_{2/3} = \mathbf{u}$ (uniform pmf, $W_{2/3}$ is more noisy than every channel), w_τ with $\tau = 1-(\delta/2)$ (W_τ is the extremal symmetric channel that is degraded by W_δ), w_γ with $\gamma = (1-\delta)/(1-\delta+(\delta/4))$ (W_γ is a 3-ary symmetric channel that is not degraded by W_δ but $W_\delta \succeq_{\mathsf{ln}} W_\gamma$), and w_1 (edge of simplex). The magenta triangle denotes the additive degradation region $\mathsf{conv}\left(\{w_\delta, w_\delta P_3, w_\delta P_3^2\}\right)$ of W_δ. The green convex region denotes the additive less noisy domination region of W_δ, and the yellow region $\mathsf{conv}\left(\{w_\delta, w_\delta P_3, w_\delta P_3^2, w_\gamma, w_\gamma P_3, w_\gamma P_3^2\}\right)$ is its lower bound while the circular cyan region $\left\{v \in \mathcal{P}_3 : \|v - \mathbf{u}\|_{\ell^2} \leq \|w_\delta - \mathbf{u}\|_{\ell^2}\right\}$ (which is a hypersphere for general $q \geq 3$) is its upper bound. Note that we do not need to specify the underlying group because there is only one group of order 3.

The standard Dirichlet form is the usual choice for Dirichlet form comparison because its logarithmic Sobolev constant has been precisely computed in [25, Appendix, Theorem A.1]. So, we present Theorem 4 using $\mathcal{E}_{\mathsf{std}}$ rather than $\mathcal{E}_{W_\delta}$.

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:*

$$\forall f \in \mathbb{R}^q, \;\; \mathcal{E}_V\left(f,f\right) \geq \frac{q\delta}{q-1}\mathcal{E}_{\mathsf{std}}\left(f,f\right).$$

An extension of Theorem 4 is proved in section VII. The domination of Dirichlet forms shown in Theorem 4 has several useful consequences. A major consequence is that we can immediately establish Poincaré (spectral gap) inequalities and logarithmic Sobolev inequalities (LSIs) for the channel V using the corresponding inequalities for q-ary symmetric channels. For example, the LSI for $W_\delta \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ with $q > 2$ is:

$$D\left(f^2\mathbf{u}||\mathbf{u}\right) \leq \frac{(q-1)\log(q-1)}{(q-2)\delta}\,\mathcal{E}_{W_\delta}\left(f,f\right) \tag{19}$$

for all $f \in \mathbb{R}^q$ such that $\sum_{k=1}^{q} f_k^2 = q$, where we use (54) and the logarithmic Sobolev constant computed in part 1 of Proposition 12. As shown in Appendix B, (19) is easily established using the known logarithmic Sobolev constant corresponding to the standard Dirichlet form. Using the LSI for V that follows from (19) and Theorem 4, we immediately obtain guarantees on the convergence rate and *hypercontractivity* properties of the associated *Markov semigroup* $\{\exp(-t(I_q - V)) : t \geq 0\}$. We refer readers to [25] and [26] for comprehensive accounts of such topics.

E. Outline

We briefly outline the content of the ensuing sections. In section III, we study the structure of less noisy domination and degradation regions of channels. In section IV, we prove Theorem 1 and present some other equivalent characterizations of $\succeq_{\mathsf{ln}}$. We then derive several necessary and sufficient conditions for less noisy domination among additive noise channels in section V, which together with the results of section III, culminates in a proof of Theorem 3. Section VI provides a proof of Theorem 2, and section VII introduces LSIs and proves an extension of Theorem 4. Finally, we conclude our discussion in section VIII.

III. Less noisy domination and degradation regions

In this section, we focus on understanding the "geometric" aspects of less noisy domination and degradation by channels. We begin by deriving some simple characteristics of the sets of channels that are dominated by some fixed channel in the less noisy and degraded senses. We then specialize our results for additive noise channels, and this culminates in a complete characterization of $\mathcal{D}_{W_\delta}^{\textsf{add}}$ and derivations of certain properties of $\mathcal{L}_{W_\delta}^{\textsf{add}}$ presented in Theorem 3.

Let $W \in \mathbb{R}_{\textsf{sto}}^{q\times r}$ be a fixed channel with $q, r \in \mathbb{N}$, and define its *less noisy domination region*:

$$\mathcal{L}_W \triangleq \left\{V \in \mathbb{R}_{\textsf{sto}}^{q\times r} : W \succeq_{\textsf{ln}} V\right\} \tag{20}$$

as the set of all channels on the same input and output alphabets that are dominated by W in the less noisy sense. Moreover, we define the *degradation region* of W:

$$\mathcal{D}_W \triangleq \left\{V \in \mathbb{R}_{\textsf{sto}}^{q\times r} : W \succeq_{\textsf{deg}} V\right\} \tag{21}$$

as the set of all channels on the same input and output alphabets that are degraded versions of W. Then, $\mathcal{L}_W$ and $\mathcal{D}_W$ satisfy the properties delineated below.

Proposition 5 (Less Noisy Domination and Degradation Regions)**.** *Given the channel $W \in \mathbb{R}_{\textsf{sto}}^{q\times r}$, its less noisy domination region $\mathcal{L}_W$ and its degradation region $\mathcal{D}_W$ are non-empty, closed, convex, and output alphabet permutation symmetric (i.e. $V \in \mathcal{L}_W \Rightarrow VP \in \mathcal{L}_W$ and $V \in \mathcal{D}_W \Rightarrow VP \in \mathcal{D}_W$ for every permutation matrix $P \in \mathbb{R}^{r\times r}$).*

Proof.
Non-Emptiness of $\mathcal{L}_W$ and $\mathcal{D}_W$: $W \succeq_{\textsf{ln}} W \Rightarrow W \in \mathcal{L}_W$, and $W \succeq_{\textsf{deg}} W \Rightarrow W \in \mathcal{D}_W$. So, $\mathcal{L}_W$ and $\mathcal{D}_W$ are non-empty.
Closure of $\mathcal{L}_W$: Fix any two pmfs $P_X, Q_X \in \mathcal{P}_q$, and consider a sequence of channels $V_k \in \mathcal{L}_W$ such that $V_k \to V \in \mathbb{R}_{\textsf{sto}}^{q\times r}$ (with respect to the Frobenius norm). Then, we also have $P_X V_k \to P_X V$ and $Q_X V_k \to Q_X V$ (with respect to the ℓ^2-norm). Hence, we get:

$$\begin{aligned} D\left(P_X V || Q_X V\right) &\leq \liminf_{k\to\infty} D\left(P_X V_k || Q_X V_k\right) \\ &\leq D\left(P_X W || Q_X W\right) \end{aligned}$$

where the first line follows from the lower semicontinuity of KL divergence [27, Theorem 1], [28, Theorem 3.6, Section 3.5], and the second line holds because $V_k \in \mathcal{L}_W$. This implies that for any two pmfs $P_X, Q_X \in \mathcal{P}_q$, the set $\mathcal{S}\left(P_X, Q_X\right) = \left\{V \in \mathbb{R}_{\textsf{sto}}^{q\times r} : D\left(P_X W || Q_X W\right) \geq D\left(P_X V || Q_X V\right)\right\}$ is actually closed. Using Proposition 1, we have that:

$$\mathcal{L}_W = \bigcap_{P_X, Q_X \in \mathcal{P}_q} \mathcal{S}\left(P_X, Q_X\right).$$

So, $\mathcal{L}_W$ is closed since it is an intersection of closed sets [29].
Closure of $\mathcal{D}_W$: Consider a sequence of channels $V_k \in \mathcal{D}_W$ such that $V_k \to V \in \mathbb{R}_{\textsf{sto}}^{q\times r}$. Since each $V_k = W A_k$ for some channel $A_k \in \mathbb{R}_{\textsf{sto}}^{r\times r}$ belonging to the compact set $\mathbb{R}_{\textsf{sto}}^{r\times r}$, there exists a subsequence A_{k_m} that converges by (sequential) compactness [29]: $A_{k_m} \to A \in \mathbb{R}_{\textsf{sto}}^{r\times r}$. Hence, $V \in \mathcal{D}_W$ since $V_{k_m} = W A_{k_m} \to WA = V$, and $\mathcal{D}_W$ is a closed set.
Convexity of $\mathcal{L}_W$: Suppose $V_1, V_2 \in \mathcal{L}_W$, and let $\lambda \in [0,1]$ and $\bar{\lambda} = 1 - \lambda$. Then, for every $P_X, Q_X \in \mathcal{P}_q$, we have:

$$D(P_X W || Q_X W) \geq D(P_X(\lambda V_1 + \bar{\lambda} V_2) || Q_X(\lambda V_1 + \bar{\lambda} V_2))$$

by the convexity of KL divergence. Hence, $\mathcal{L}_W$ is convex.
Convexity of $\mathcal{D}_W$: If $V_1, V_2 \in \mathcal{D}_W$ so that $V_1 = W A_1$ and $V_2 = W A_2$ for some $A_1, A_2 \in \mathbb{R}_{\textsf{sto}}^{r\times r}$, then $\lambda V_1 + \bar{\lambda} V_2 = W(\lambda A_1 + \bar{\lambda} A_2) \in \mathcal{D}_W$ for all $\lambda \in [0,1]$, and $\mathcal{D}_W$ is convex.
Symmetry of $\mathcal{L}_W$: This is obvious from Proposition 1 because KL divergence is invariant to permutations of its input pmfs.
Symmetry of $\mathcal{D}_W$: Given $V \in \mathcal{D}_W$ so that $V = WA$ for some $A \in \mathbb{R}_{\textsf{sto}}^{r\times r}$, we have that $VP = WAP \in \mathcal{D}_W$ for every permutation matrix $P \in \mathbb{R}^{r\times r}$. This completes the proof. ∎

While the channels in $\mathcal{L}_W$ and $\mathcal{D}_W$ all have the same output alphabet as W, as defined in (20) and (21), we may extend the output alphabet of W by adding zero probability letters. So, separate less noisy domination and degradation regions can be defined for each output alphabet size that is at least as large as the original output alphabet size of W.

A. Less noisy domination and degradation regions for additive noise channels

Often in information theory, we are concerned with additive noise channels on an Abelian group $(\mathcal{X}, \oplus)$ with $\mathcal{X} = [q]$ and $q \in \mathbb{N}$, as defined in (9). Such channels are completely defined by a noise pmf $P_Z \in \mathcal{P}_q$ with corresponding channel transition probability matrix $\mathsf{circ}_{\mathcal{X}}(P_Z) \in \mathbb{R}_{\mathsf{sto}}^{q \times q}$. Suppose $W = \mathsf{circ}_{\mathcal{X}}(w) \in \mathbb{R}_{\mathsf{sto}}^{q \times q}$ is an additive noise channel with noise pmf $w \in \mathcal{P}_q$. Then, we are often only interested in the set of additive noise channels that are dominated by W. We define the *additive less noisy domination region* of W:

$$\mathcal{L}_W^{\mathsf{add}} \triangleq \{v \in \mathcal{P}_q : W \succeq_{\mathsf{ln}} \mathsf{circ}_{\mathcal{X}}(v)\} \tag{22}$$

as the set of all noise pmfs whose corresponding channel transition matrices are dominated by W in the less noisy sense. Likewise, we define the *additive degradation region* of W:

$$\mathcal{D}_W^{\mathsf{add}} \triangleq \{v \in \mathcal{P}_q : W \succeq_{\mathsf{deg}} \mathsf{circ}_{\mathcal{X}}(v)\} \tag{23}$$

as the set of all noise pmfs whose corresponding channel transition matrices are degraded versions of W. (These definitions generalize (14) and (15), and can also hold for any non-additive noise channel W.) The next proposition illustrates certain properties of $\mathcal{L}_W^{\mathsf{add}}$ and explicitly characterizes $\mathcal{D}_W^{\mathsf{add}}$.

Proposition 6 (Additive Less Noisy Domination and Degradation Regions)**.** *Given the additive noise channel $W = \mathsf{circ}_{\mathcal{X}}(w) \in \mathbb{R}_{\mathsf{sto}}^{q \times q}$ with noise pmf $w \in \mathcal{P}_q$, we have:*

1) *$\mathcal{L}_W^{\mathsf{add}}$ and $\mathcal{D}_W^{\mathsf{add}}$ are non-empty, closed, convex, and invariant under the permutations $\{P_x \in \mathbb{R}^{q \times q} : x \in \mathcal{X}\}$ defined in (4) (i.e. $v \in \mathcal{L}_W^{\mathsf{add}} \Rightarrow vP_x \in \mathcal{L}_W^{\mathsf{add}}$ and $v \in \mathcal{D}_W^{\mathsf{add}} \Rightarrow vP_x \in \mathcal{D}_W^{\mathsf{add}}$ for every $x \in \mathcal{X}$).*
2) *$\mathcal{D}_W^{\mathsf{add}} = \mathsf{conv}(\{wP_x : x \in \mathcal{X}\}) = \{v \in \mathcal{P}_q : w \succeq_{\mathcal{X}} v\}$, where $\succeq_{\mathcal{X}}$ denotes the group majorization preorder as defined in Appendix A.*

To prove Proposition 6, we will need the following lemma.

Lemma 1 (Additive Noise Channel Degradation)**.** *Given two additive noise channels $W = \mathsf{circ}_{\mathcal{X}}(w) \in \mathbb{R}_{\mathsf{sto}}^{q \times q}$ and $V = \mathsf{circ}_{\mathcal{X}}(v) \in \mathbb{R}_{\mathsf{sto}}^{q \times q}$ with noise pmfs $w, v \in \mathcal{P}_q$, $W \succeq_{\mathsf{deg}} V$ if and only if $V = W\mathsf{circ}_{\mathcal{X}}(z) = \mathsf{circ}_{\mathcal{X}}(z)W$ for some $z \in \mathcal{P}_q$ (i.e. for additive noise channels $W \succeq_{\mathsf{deg}} V$, the channel that degrades W to produce V is also an additive noise channel without loss of generality).*

Proof. Since $\mathcal{X}$-circulant matrices commute, we must have $W\mathsf{circ}_{\mathcal{X}}(z) = \mathsf{circ}_{\mathcal{X}}(z)W$ for every $z \in \mathcal{P}_q$. Furthermore, $V = W\mathsf{circ}_{\mathcal{X}}(z)$ for some $z \in \mathcal{P}_q$ implies that $W \succeq_{\mathsf{deg}} V$ by Definition 1. So, it suffices to prove that $W \succeq_{\mathsf{deg}} V$ implies $V = W\mathsf{circ}_{\mathcal{X}}(z)$ for some $z \in \mathcal{P}_q$. By Definition 1, $W \succeq_{\mathsf{deg}} V$ implies that $V = WR$ for some doubly stochastic channel $R \in \mathbb{R}_{\mathsf{sto}}^{q \times q}$ (as V and W are doubly stochastic). Let r with $r^T \in \mathcal{P}_q$ be the first column of R, and $s = Wr$ with $s^T \in \mathcal{P}_q$ be the first column of V. Then, it is straightforward to verify using (7) that:

$$\begin{aligned} V &= \begin{bmatrix} s & P_1 s & P_2 s & \cdots & P_{q-1} s \end{bmatrix} \\ &= \begin{bmatrix} Wr & P_1 Wr & P_2 Wr & \cdots & P_{q-1} Wr \end{bmatrix} \\ &= W \begin{bmatrix} r & P_1 r & P_2 r & \cdots & P_{q-1} r \end{bmatrix} \end{aligned}$$

where the third equality holds because $\{P_x : x \in \mathcal{X}\}$ are $\mathcal{X}$-circulant matrices which commute with W. Hence, V is the product of W and an $\mathcal{X}$-circulant stochastic matrix, i.e. $V = W\mathsf{circ}_{\mathcal{X}}(z)$ for some $z \in \mathcal{P}_q$. This concludes the proof. ∎

We emphasize that in Lemma 1, the channel that degrades W to produce V is only an additive noise channel without loss of generality. We can certainly have $V = WR$ with a non-additive noise channel R. Consider for instance, $V = W = \mathbf{1}\mathbf{1}^T/q$, where every doubly stochastic matrix R satisfies $V = WR$. However, when we consider $V = WR$ with an additive noise channel R, V corresponds to the channel W with an additional independent additive noise term associated with R. We now prove Proposition 6.

Proof of Proposition 6.
Part 1: Non-emptiness, closure, and convexity of $\mathcal{L}_W^{\mathsf{add}}$ and $\mathcal{D}_W^{\mathsf{add}}$ can be proved in exactly the same way as in Proposition 5, with the additional observation that the set of $\mathcal{X}$-circulant matrices is closed and convex. Moreover, for every $x \in \mathcal{X}$:

$$\begin{aligned} W &\succeq_{\mathsf{ln}} \ WP_x = \mathsf{circ}_{\mathcal{X}}(wP_x) \succeq_{\mathsf{ln}} \ W \\ W &\succeq_{\mathsf{deg}} WP_x = \mathsf{circ}_{\mathcal{X}}(wP_x) \succeq_{\mathsf{deg}} W \end{aligned}$$

where the equalities follow from (7). These inequalities and the transitive properties of $\succeq_{\mathsf{ln}}$ and $\succeq_{\mathsf{deg}}$ yield the invariance of $\mathcal{L}_W^{\mathsf{add}}$ and $\mathcal{D}_W^{\mathsf{add}}$ with respect to $\{P_x \in \mathbb{R}^{q\times q} : x \in \mathcal{X}\}$.

Part 2: Lemma 1 is equivalent to the fact that $v \in \mathcal{D}_W^{\mathsf{add}}$ if and only if $\mathsf{circ}_{\mathcal{X}}(v) = \mathsf{circ}_{\mathcal{X}}(w)\,\mathsf{circ}_{\mathcal{X}}(z)$ for some $z \in \mathcal{P}_q$. This implies that $v \in \mathcal{D}_W^{\mathsf{add}}$ if and only if $v = w\,\mathsf{circ}_{\mathcal{X}}(z)$ for some $z \in \mathcal{P}_q$ (due to (7) and the fact that $\mathcal{X}$-circulant matrices commute). Applying Proposition 14 from Appendix A completes the proof. ■

We remark that part 1 of Proposition 6 does not require W to be an additive noise channel. The proofs of closure, convexity, and invariance with respect to $\{P_x \in \mathbb{R}^{q\times q} : x \in \mathcal{X}\}$ hold for general $W \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$. Moreover, $\mathcal{L}_W^{\mathsf{add}}$ and $\mathcal{D}_W^{\mathsf{add}}$ are non-empty because $\mathbf{u} \in \mathcal{L}_W^{\mathsf{add}}$ and $\mathbf{u} \in \mathcal{D}_W^{\mathsf{add}}$.

B. Less noisy domination and degradation regions for symmetric channels

Since q-ary symmetric channels for $q \in \mathbb{N}$ are additive noise channels, Proposition 6 holds for symmetric channels. In this subsection, we deduce some simple results that are unique to symmetric channels. The first of these is a specialization of part 2 of Proposition 6 which states that the additive degradation region of a symmetric channel can be characterized by traditional majorization instead of group majorization.

Corollary 1 (Degradation Region of Symmetric Channel)**.** *The q-ary symmetric channel $W_\delta = \mathsf{circ}_{\mathcal{X}}(w_\delta) \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$ for $\delta \in [0,1]$ has additive degradation region:*

$$\mathcal{D}_{W_\delta}^{\mathsf{add}} = \{v \in \mathcal{P}_q : w_\delta \succeq_{\mathsf{maj}} v\} = \mathsf{conv}\left(\left\{w_\delta P_q^k : k \in [q]\right\}\right)$$

where $\succeq_{\mathsf{maj}}$ denotes the majorization preorder defined in Appendix A, and $P_q \in \mathbb{R}^{q\times q}$ is defined in (8).

Proof. From part 2 of Proposition 6, we have that:

$$\begin{aligned}\mathcal{D}_{W_\delta}^{\mathsf{add}} &= \mathsf{conv}\left(\{w_\delta P_x : x \in \mathcal{X}\}\right) = \mathsf{conv}\left(\left\{w_\delta P_q^k : k \in [q]\right\}\right) \\ &= \mathsf{conv}\left(\left\{w_\delta P : P \in \mathbb{R}^{q\times q} \text{ is a permutation matrix}\right\}\right) \\ &= \{v \in \mathcal{P}_q : w \succeq_{\mathsf{maj}} v\}\end{aligned}$$

where the second and third equalities hold regardless of the choice of group $(\mathcal{X}, \oplus)$, because the sets of all cyclic or regular permutations of $w_\delta = (1-\delta, \delta/(q-1), \ldots, \delta/(q-1))$ equal $\{w_\delta P_x : x \in \mathcal{X}\}$. The final equality follows from the definition of majorization in Appendix A. ■

With this geometric characterization of the additive degradation region, it is straightforward to find the extremal symmetric channel W_τ that is a degraded version of W_δ for some fixed $\delta \in [0,1]\backslash\left\{\frac{q-1}{q}\right\}$. Indeed, we compute τ by using the fact that the noise pmf $w_\tau \in \mathsf{conv}\left(\left\{w_\delta P_q^k : k = 1, \ldots, q-1\right\}\right)$:

$$w_\tau = \sum_{i=1}^{q-1} \lambda_i w_\delta P_q^i \tag{24}$$

for some $\lambda_1, \ldots, \lambda_{q-1} \in [0,1]$ such that $\lambda_1 + \cdots + \lambda_{q-1} = 1$. Solving (24) for τ and $\lambda_1, \ldots, \lambda_{q-1}$ yields:

$$\tau = 1 - \frac{\delta}{q-1} \tag{25}$$

and $\lambda_1 = \cdots = \lambda_{q-1} = \frac{1}{q-1}$, which means that:

$$w_\tau = \frac{1}{q-1} \sum_{i=1}^{q-1} w_\delta P_q^i. \tag{26}$$

This is illustrated in Figure 2 for the case where $\delta \in \left(0, \frac{q-1}{q}\right)$ and $\tau > \frac{q-1}{q} > \delta$. For $\delta \in \left(0, \frac{q-1}{q}\right)$, the symmetric channels that are degraded versions of W_δ are $\{W_\gamma : \gamma \in [\delta, \tau]\}$. In particular, for such $\gamma \in [\delta, \tau]$, $W_\gamma = W_\delta W_\beta$ with $\beta = (\gamma - \delta)/(1 - \delta - \frac{\delta}{q-1})$ using the proof of part 5 of Proposition 4 in Appendix B.

In the spirit of comparing symmetric and erasure channels as done in [15] for the binary input case, our next result shows that a q-ary symmetric channel can never be less noisy than a q-ary erasure channel.

Proposition 7 (Symmetric Channel $\not\succeq_{\mathsf{ln}}$ Erasure Channel)**.** *For $q \in \mathbb{N}\backslash\{1\}$, given a q-ary erasure channel $E_\epsilon \in \mathbb{R}_{\mathsf{sto}}^{q\times(q+1)}$ with erasure probability $\epsilon \in (0,1)$, there does not exist $\delta \in (0,1)$ such that the corresponding q-ary symmetric channel $W_\delta \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$ on the same input alphabet satisfies $W_\delta \succeq_{\mathsf{ln}} E_\epsilon$.*

Proof. For a q-ary erasure channel E_ϵ with $\epsilon \in (0,1)$, we always have $D(\mathbf{u}E_\epsilon || \Delta_0 E_\epsilon) = +\infty$ for $\mathbf{u}, \Delta_0 = (1,0,\dots,0) \in \mathcal{P}_q$. On the other hand, for any q-ary symmetric channel W_δ with $\delta \in (0,1)$, we have $D(P_X W_\delta || Q_X W_\delta) < +\infty$ for every $P_X, Q_X \in \mathcal{P}_q$. Thus, $W_\delta \not\succeq_{\mathsf{ln}} E_\epsilon$ for any $\delta \in (0,1)$. ■

In fact, the argument for Proposition 7 conveys that a symmetric channel $W_\delta \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$ with $\delta \in (0,1)$ satisfies $W_\delta \succeq_{\mathsf{ln}} V$ for some channel $V \in \mathbb{R}_{\mathsf{sto}}^{q\times r}$ only if $D(P_X V || Q_X V) < +\infty$ for every $P_X, Q_X \in \mathcal{P}_q$. Typically, we are only interested in studying q-ary symmetric channels with $q \geq 2$ and $\delta \in \left(0, \frac{q-1}{q}\right)$. For example, the BSC with crossover probability δ is usually studied for $\delta \in \left(0, \frac{1}{2}\right)$. Indeed, the less noisy domination characteristics of the extremal q-ary symmetric channels with $\delta = 0$ or $\delta = \frac{q-1}{q}$ are quite elementary. Given $q \geq 2$, $W_0 = I_q \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$ satisfies $W_0 \succeq_{\mathsf{ln}} V$, and $W_{(q-1)/q} = \mathbf{1u} \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$ satisfies $V \succeq_{\mathsf{ln}} W_{(q-1)/q}$, for every channel $V \in \mathbb{R}_{\mathsf{sto}}^{q\times r}$ on a common input alphabet. For the sake of completeness, we also note that for $q \geq 2$, the extremal q-ary erasure channels $E_0 \in \mathbb{R}_{\mathsf{sto}}^{q\times(q+1)}$ and $E_1 \in \mathbb{R}_{\mathsf{sto}}^{q\times(q+1)}$, with $\epsilon = 0$ and $\epsilon = 1$ respectively, satisfy $E_0 \succeq_{\mathsf{ln}} V$ and $V \succeq_{\mathsf{ln}} E_1$ for every channel $V \in \mathbb{R}_{\mathsf{sto}}^{q\times r}$ on a common input alphabet.

The result that the q-ary symmetric channel with uniform noise pmf $W_{(q-1)/q}$ is more noisy than every channel on the same input alphabet has an analogue concerning additive white Gaussian noise (AWGN) channels. Consider all additive noise channels of the form:

$$Y = X + Z \tag{27}$$

where $X, Y \in \mathbb{R}$, the input X is uncorrelated with the additive noise Z: $\mathbb{E}[XZ] = 0$, and the noise Z has power constraint $\mathbb{E}\left[Z^2\right] \leq \sigma_Z^2$ for some fixed $\sigma_Z > 0$. Let $X = X_{\mathsf{g}} \sim \mathcal{N}(0, \sigma_X^2)$ (Gaussian distribution with mean 0 and variance σ_X^2) for some $\sigma_X > 0$. Then, we have:

$$I\left(X_{\mathsf{g}}; X_{\mathsf{g}} + Z\right) \geq I\left(X_{\mathsf{g}}; X_{\mathsf{g}} + Z_{\mathsf{g}}\right) \tag{28}$$

where $Z_{\mathsf{g}} \sim \mathcal{N}(0, \sigma_Z^2)$, Z_{g} is independent of X_{g}, and equality occurs if and only if $Z = Z_{\mathsf{g}}$ in distribution [28, Section 4.7]. This states that Gaussian noise is the "worst case additive noise" for a Gaussian source. Hence, the AWGN channel is *not more capable* than any other additive noise channel with the same constraints. As a result, the AWGN channel is *not less noisy* than any other additive noise channel with the same constraints (using Proposition 3).

IV. Equivalent characterizations of less noisy preorder

Having studied the structure of less noisy domination and degradation regions of channels, we now consider the problem of verifying whether a channel W is less noisy than another channel V. Since using Definition 2 or Proposition 1 directly is difficult, we often start by checking whether V is a degraded version of W. When this fails, we typically resort to verifying van Dijk's condition in Proposition 2, cf. [12, Theorem 2]. In this section, we prove the equivalent characterization of the less noisy preorder in Theorem 1, and then present some useful corollaries of van Dijk's condition.

A. Characterization using χ^2-divergence

Recall the general measure theoretic setup and the definition of χ^2-divergence from subsection II-A. It is well-known that KL divergence is locally approximated by χ^2-divergence, e.g. [28, Section 4.2]. While this approximation sometimes fails globally, cf. [30], the following notable result was first shown by Ahlswede and Gács in the discrete case in [2], and then extended to general alphabets in [3, Theorem 3]:

$$\eta_{\mathsf{KL}}(W) = \eta_{\chi^2}(W) \triangleq \sup_{\substack{P_X, Q_X \\ 0 < \chi^2(P_X || Q_X) < +\infty}} \frac{\chi^2\left(P_X W || Q_X W\right)}{\chi^2\left(P_X || Q_X\right)} \tag{29}$$

for any Markov kernel $W : \mathcal{H}_1 \times \mathcal{X} \rightarrow [0,1]$, where $\eta_{\mathsf{KL}}(W)$ is defined as in (11), $\eta_{\chi^2}(W)$ is the contraction coefficient for χ^2-divergence, and the suprema in $\eta_{\mathsf{KL}}(W)$ and $\eta_{\chi^2}(W)$ are taken over all probability measures P_X and Q_X on $(\mathcal{X}, \mathcal{F})$. Since η_{KL} characterizes less noisy domination with respect to an erasure channel as mentioned in subsection I-D, (29) portrays that η_{χ^2} also characterizes this. We will now prove Theorem 1 from subsection II-A, which generalizes (29) and illustrates that χ^2-divergence actually characterizes less noisy domination by an arbitrary channel.

Proof of Theorem 1. In order to prove the forward direction, we recall the local approximation of KL divergence using χ^2-divergence from [28, Proposition 4.2], which states that for any two probability measures P_X and Q_X on $(\mathcal{X}, \mathcal{F})$:

$$\lim_{\lambda \rightarrow 0^+} \frac{2}{\lambda^2} D\left(\lambda P_X + \bar{\lambda} Q_X || Q_X\right) = \chi^2(P_X || Q_X) \tag{30}$$

where $\bar{\lambda} = 1 - \lambda$ for $\lambda \in (0,1)$, and both sides of (30) are finite or infinite together. Then, we observe that for any two probability measures P_X and Q_X, and any $\lambda \in [0,1]$, we have:

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

since $W \succeq_{\mathsf{ln}} V$. Scaling this inequality by $\frac{2}{\lambda^2}$ and letting $\lambda \to 0$ produces:

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

as shown in (30). This proves the forward direction.

To establish the converse direction, we recall an integral representation of KL divergence using χ^2-divergence presented in [3, Appendix A.2] (which can be distilled from the argument in [31, Theorem 1]):[5]

$$D(P_X || Q_X) = \int_0^{\infty} \frac{\chi^2(P_X || Q_X^t)}{t+1} \, dt \tag{31}$$

for any two probability measures P_X and Q_X on $(\mathcal{X}, \mathcal{F})$, where $Q_X^t = \frac{t}{1+t} P_X + \frac{1}{t+1} Q_X$ for $t \in [0, \infty)$, and both sides of (31) are finite or infinite together (as a close inspection of the proof in [3, Appendix A.2] reveals). Hence, for every P_X and Q_X, we have by assumption:

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

which implies that:

$$\int_0^{\infty} \frac{\chi^2(P_X W || Q_X^t W)}{t+1} \, dt \geq \int_0^{\infty} \frac{\chi^2(P_X V || Q_X^t V)}{t+1} \, dt$$
$$\Rightarrow \; D(P_X W || Q_X W) \geq D(P_X V || Q_X V) \, .$$

Hence, $W \succeq_{\mathsf{ln}} V$, which completes the proof. ∎

B. Characterizations via the Löwner partial order and spectral radius

We will use the finite alphabet setup of subsection I-B for the remaining discussion in this paper. In the finite alphabet setting, Theorem 1 states that $W \in \mathbb{R}_{\mathsf{sto}}^{q \times r}$ is less noisy than $V \in \mathbb{R}_{\mathsf{sto}}^{q \times s}$ if and only if for every $P_X, Q_X \in \mathcal{P}_q$:

$$\chi^2(P_X W || Q_X W) \geq \chi^2(P_X V || Q_X V) \, . \tag{32}$$

This characterization has the flavor of a Löwner partial order condition. Indeed, it is straightforward to verify that for any $P_X \in \mathcal{P}_q$ and $Q_X \in \mathcal{P}_q^{\circ}$, we can write their χ^2-divergence as:

$$\chi^2(P_X || Q_X) = J_X \mathsf{diag}(Q_X)^{-1} J_X^T . \tag{33}$$

where $J_X = P_X - Q_X$. Hence, we can express (32) as:

$$J_X W \mathsf{diag}(Q_X W)^{-1} W^T J_X^T \geq J_X V \mathsf{diag}(Q_X V)^{-1} V^T J_X^T \tag{34}$$

for every $J_X = P_X - Q_X$ such that $P_X \in \mathcal{P}_q$ and $Q_X \in \mathcal{P}_q^{\circ}$. This suggests that (32) is equivalent to:

$$W \mathsf{diag}(Q_X W)^{-1} W^T \succeq_{\mathsf{PSD}} V \mathsf{diag}(Q_X V)^{-1} V^T \tag{35}$$

for every $Q_X \in \mathcal{P}_q^{\circ}$. It turns out that (35) indeed characterizes $\succeq_{\mathsf{ln}}$, and this is straightforward to prove directly. The next proposition illustrates that (35) also follows as a corollary of van Dijk's characterization in Proposition 2, and presents an equivalent spectral characterization of $\succeq_{\mathsf{ln}}$.

Proposition 8 (Löwner and Spectral Characterizations of $\succeq_{\mathsf{ln}}$)**.** *For any pair of channels $W \in \mathbb{R}_{\mathsf{sto}}^{q \times r}$ and $V \in \mathbb{R}_{\mathsf{sto}}^{q \times s}$ on the same input alphabet $[q]$, the following are equivalent:*

1) $W \succeq_{\mathsf{ln}} V$
2) *For every $P_X \in \mathcal{P}_q^{\circ}$, we have:*

$$W \mathsf{diag}(P_X W)^{-1} W^T \succeq_{\mathsf{PSD}} V \mathsf{diag}(P_X V)^{-1} V^T$$

[5] Note that [3, Equation (78)], and hence [1, Equation (7)], are missing factors of $\frac{1}{t+1}$ inside the integrals.

3) *For every* $P_X \in \mathcal{P}_q^\circ$, *we have* $\mathcal{R}\big(V\,\textsf{diag}(P_X V)^{-1}\,V^T\big) \subseteq \mathcal{R}\big(W\,\textsf{diag}(P_X W)^{-1}\,W^T\big)$ *and:*[6]

$$\rho\bigg(\Big(W\,\textsf{diag}(P_X W)^{-1}\,W^T\Big)^{\dagger} V\,\textsf{diag}(P_X V)^{-1}\,V^T\bigg) = 1.$$

Proof. ($1 \Leftrightarrow 2$) Recall the functional $F : \mathcal{P}_q \to \mathbb{R}, F(P_X) = I(P_X, W_{Y|X}) - I(P_X, V_{Y|X})$ defined in Proposition 2, cf. [12, Theorem 2]. Since $F : \mathcal{P}_q \to \mathbb{R}$ is continuous on its domain $\mathcal{P}_q$, and twice differentiable on $\mathcal{P}_q^\circ$, F is concave if and only if its Hessian is negative semidefinite for every $P_X \in \mathcal{P}_q^\circ$ (i.e. $-\nabla^2 F\left(P_X\right) \succeq_{\textsf{PSD}} 0$ for every $P_X \in \mathcal{P}_q^\circ$) [32, Section 3.1.4]. The Hessian matrix of F, $\nabla^2 F : \mathcal{P}_q^\circ \to \mathbb{R}^{q\times q}_{\textsf{sym}}$, is defined entry-wise for every $x, x' \in [q]$ as:

$$\left[\nabla^2 F(P_X)\right]_{x,x'} = \frac{\partial^2 F}{\partial P_X(x) \partial P_X(x')}\left(P_X\right)$$

where we index the matrix $\nabla^2 F(P_X)$ starting at 0 rather than 1. Furthermore, a straightforward calculation shows that:

$$\nabla^2 F(P_X) = V\textsf{diag}(P_X V)^{-1}\,V^T - W\textsf{diag}(P_X W)^{-1}\,W^T$$

for every $P_X \in \mathcal{P}_q^\circ$. (Note that the matrix inverses here are well-defined because $P_X \in \mathcal{P}_q^\circ$). Therefore, F is concave if and only if for every $P_X \in \mathcal{P}_q^\circ$:

$$W\textsf{diag}(P_X W)^{-1}\,W^T \succeq_{\textsf{PSD}} V\textsf{diag}(P_X V)^{-1}\,V^T.$$

This establishes the equivalence between parts 1 and 2 due to van Dijk's characterization of $\succeq_{\textsf{ln}}$ in Proposition 2.

($2 \Leftrightarrow 3$) We now derive the spectral characterization of $\succeq_{\textsf{ln}}$ using part 2. Recall the well-known fact (see [33, Theorem 1 parts (a),(f)] and [19, Theorem 7.7.3 (a)]):
Given positive semidefinite matrices $A, B \in \mathbb{R}^{q\times q}_{\succeq 0}$, $A \succeq_{\textsf{PSD}} B$ *if and only if* $\mathcal{R}(B) \subseteq \mathcal{R}(A)$ *and* $\rho\left(A^\dagger B\right) \leq 1$.
Since $W\textsf{diag}(P_X W)^{-1}\,W^T$ and $V\textsf{diag}(P_X V)^{-1}\,V^T$ are positive semidefinite for every $P_X \in \mathcal{P}_q^\circ$, applying this fact shows that part 2 holds if and only if for every $P_X \in \mathcal{P}_q^\circ$, we have $\mathcal{R}\big(V\textsf{diag}(P_X V)^{-1}\,V^T\big) \subseteq \mathcal{R}\big(W\textsf{diag}(P_X W)^{-1}\,W^T\big)$ and:

$$\rho\left(\left(W\textsf{diag}(P_X W)^{-1}\,W^T\right)^{\dagger} V\textsf{diag}(P_X V)^{-1}\,V^T\right) \leq 1.$$

To prove that this inequality is actually an equality, for any $P_X \in \mathcal{P}_q^\circ$, let $A = W\textsf{diag}(P_X W)^{-1}\,W^T$ and $B = V\textsf{diag}(P_X V)^{-1}\,V^T$. It suffices to prove that: $\mathcal{R}(B) \subseteq \mathcal{R}(A)$ and $\rho\left(A^\dagger B\right) \leq 1$ if and only if $\mathcal{R}(B) \subseteq \mathcal{R}(A)$ and $\rho\left(A^\dagger B\right) = 1$. The converse direction is trivial, so we only establish the forward direction. Observe that $P_X A = \mathbf{1}^T$ and $P_X B = \mathbf{1}^T$. This implies that $\mathbf{1}^T A^\dagger B = P_X(AA^\dagger)B = P_X B = \mathbf{1}^T$, where $(AA^\dagger)B = B$ because $\mathcal{R}(B) \subseteq \mathcal{R}(A)$ and $AA^\dagger$ is the orthogonal projection matrix onto $\mathcal{R}(A)$. Since $\rho\left(A^\dagger B\right) \leq 1$ and $A^\dagger B$ has an eigenvalue of 1, we have $\rho\left(A^\dagger B\right) = 1$. Thus, we have proved that part 2 holds if and only if for every $P_X \in \mathcal{P}_q^\circ$, we have $\mathcal{R}\big(V\textsf{diag}(P_X V)^{-1}\,V^T\big) \subseteq \mathcal{R}\big(W\textsf{diag}(P_X W)^{-1}\,W^T\big)$ and:

$$\rho\left(\left(W\textsf{diag}(P_X W)^{-1}\,W^T\right)^{\dagger} V\textsf{diag}(P_X V)^{-1}\,V^T\right) = 1.$$

This completes the proof. ∎

The Löwner characterization of $\succeq_{\textsf{ln}}$ in part 2 of Proposition 8 will be useful for proving some of our ensuing results. We remark that the equivalence between parts 1 and 2 can be derived by considering several other functionals. For instance, for any fixed pmf $Q_X \in \mathcal{P}_q^\circ$, we may consider the functional $F_2 : \mathcal{P}_q \to \mathbb{R}$ defined by:

$$F_2(P_X) = D(P_X W || Q_X W) - D(P_X V || Q_X V) \tag{36}$$

which has Hessian matrix, $\nabla^2 F_2 : \mathcal{P}_q^\circ \to \mathbb{R}^{q\times q}_{\textsf{sym}}$, $\nabla^2 F_2(P_X) = W\textsf{diag}(P_X W)^{-1}\,W^T - V\textsf{diag}(P_X V)^{-1}\,V^T$, that does not depend on Q_X. Much like van Dijk's functional F, F_2 is *convex* (for all $Q_X \in \mathcal{P}_q^\circ$) if and only if $W \succeq_{\textsf{ln}} V$. This is reminiscent of Ahlswede and Gács' technique to prove (29), where the convexity of a similar functional is established [2].

As another example, for any fixed pmf $Q_X \in \mathcal{P}_q^\circ$, consider the functional $F_3 : \mathcal{P}_q \to \mathbb{R}$ defined by:

$$F_3(P_X) = \chi^2(P_X W || Q_X W) - \chi^2(P_X V || Q_X V) \tag{37}$$

which has Hessian matrix, $\nabla^2 F_3 : \mathcal{P}_q^\circ \to \mathbb{R}^{q\times q}_{\textsf{sym}}$, $\nabla^2 F_3(P_X) = 2\,W\textsf{diag}(Q_X W)^{-1}\,W^T - 2\,V\textsf{diag}(Q_X V)^{-1}\,V^T$, that does not depend on P_X. Much like F and F_2, F_3 is *convex* for all $Q_X \in \mathcal{P}_q^\circ$ if and only if $W \succeq_{\textsf{ln}} V$.

[6] Note that [1, Theorem 1 part 4] neglected to mention the inclusion relation $\mathcal{R}\big(V\textsf{diag}(P_X V)^{-1}\,V^T\big) \subseteq \mathcal{R}\big(W\textsf{diag}(P_X W)^{-1}\,W^T\big)$.

Finally, we also mention some specializations of the spectral radius condition in part 3 of Proposition 8. If $q \geq r$ and W has full column rank, the expression for spectral radius in the proposition statement can be simplified to:

$$\rho\left((W^{\dagger})^T \mathsf{diag}(P_X W)\, W^{\dagger} V \mathsf{diag}(P_X V)^{-1} V^T\right) = 1 \tag{38}$$

using basic properties of the Moore-Penrose pseudoinverse. Moreover, if $q = r$ and W is non-singular, then the Moore-Penrose pseudoinverses in (38) can be written as inverses, and the inclusion relation between the ranges in part 3 of Proposition 8 is trivially satisfied (and can be omitted from the proposition statement). We have used the spectral radius condition in this latter setting to (numerically) compute the additive less noisy domination region in Figure 2.

V. Conditions for less noisy domination over additive noise channels

We now turn our attention to deriving several conditions for determining when q-ary symmetric channels are less noisy than other channels. Our interest in q-ary symmetric channels arises from their analytical tractability; Proposition 4 from subsection I-C, Proposition 12 from section VII, and [34, Theorem 4.5.2] (which conveys that q-ary symmetric channels have uniform capacity achieving input distributions) serve as illustrations of this tractability. We focus on additive noise channels in this section, and on general channels in the next section.

A. Necessary conditions

We first present some straightforward necessary conditions for when an additive noise channel $W \in \mathbb{R}_{\mathsf{sto}}^{q \times q}$ with $q \in \mathbb{N}$ is less noisy than another additive noise channel $V \in \mathbb{R}_{\mathsf{sto}}^{q \times q}$ on an Abelian group $(\mathcal{X}, \oplus)$. These conditions can obviously be specialized for less noisy domination by symmetric channels.

Proposition 9 (Necessary Conditions for $\succeq_{\mathsf{ln}}$ Domination over Additive Noise Channels)**.** *Suppose $W = \mathsf{circ}_{\mathcal{X}}(w)$ and $V = \mathsf{circ}_{\mathcal{X}}(v)$ are additive noise channels with noise pmfs $w, v \in \mathcal{P}_q$ such that $W \succeq_{\mathsf{ln}} V$. Then, the following are true:*

1) *(Circle Condition) $\|w - \mathbf{u}\|_{\ell^2} \geq \|v - \mathbf{u}\|_{\ell^2}$.*
2) *(Contraction Condition) $\eta_{\mathsf{KL}}(W) \geq \eta_{\mathsf{KL}}(V)$.*
3) *(Entropy Condition) $H(v) \geq H(w)$, where $H : \mathcal{P}_q \to \mathbb{R}^+$ is the Shannon entropy function.*

Proof.
Part 1: Letting $P_X = (1, 0, \ldots, 0)$ and $Q_X = \mathbf{u}$ in the χ^2-divergence characterization of $\succeq_{\mathsf{ln}}$ in Theorem 1 produces:

$$q\,\|w - \mathbf{u}\|_{\ell^2}^2 = \chi^2(w||\mathbf{u}) \geq \chi^2(v||\mathbf{u}) = q\,\|v - \mathbf{u}\|_{\ell^2}^2$$

since $\mathbf{u}W = \mathbf{u}V = \mathbf{u}$, and $P_X W = w$ and $P_X V = v$ using (7). (This result can alternatively be proved using part 2 of Proposition 8 and Fourier analysis.)
Part 2: This easily follows from Proposition 1 and (11).
Part 3: Letting $P_X = (1, 0, \ldots, 0)$ and $Q_X = \mathbf{u}$ in the KL divergence characterization of $\succeq_{\mathsf{ln}}$ in Proposition 1 produces:

$$\log(q) - H(w) = D(w||\mathbf{u}) \geq D(v||\mathbf{u}) = \log(q) - H(v)$$

via the same reasoning as part 1. This completes the proof. ■

We remark that the aforementioned necessary conditions have many generalizations. Firstly, if $W, V \in \mathbb{R}_{\mathsf{sto}}^{q \times q}$ are doubly stochastic matrices, then the generalized circle condition holds:

$$\left\|W - W_{\frac{q-1}{q}}\right\|_{\mathsf{Fro}} \geq \left\|V - W_{\frac{q-1}{q}}\right\|_{\mathsf{Fro}} \tag{39}$$

where $W_{(q-1)/q} = \mathbf{1u}$ is the q-ary symmetric channel whose conditional pmfs are all uniform, and $\|\cdot\|_{\mathsf{Fro}}$ denotes the Frobenius norm. Indeed, letting $P_X = \Delta_x = (0, \ldots, 1, \ldots, 0)$ for $x \in [q]$, which has unity in the $(x+1)$th position, in the proof of part 1 and then adding the inequalities corresponding to every $x \in [q]$ produces (39). Secondly, the contraction condition in Proposition 9 actually holds for any pair of general channels $W \in \mathbb{R}_{\mathsf{sto}}^{q \times r}$ and $V \in \mathbb{R}_{\mathsf{sto}}^{q \times s}$ on a common input alphabet (not necessarily additive noise channels). Moreover, we can start with Theorem 1 and take the suprema of the ratios in $\chi^2(P_X W||Q_X W)/\chi^2(P_X||Q_X) \geq \chi^2(P_X V||Q_X V)/\chi^2(P_X||Q_X)$ over all P_X $(\neq Q_X)$ to get:

$$\rho_{\mathsf{max}}(Q_X, W) \geq \rho_{\mathsf{max}}(Q_X, V) \tag{40}$$

for any $Q_X \in \mathcal{P}_q$, where $\rho_{\mathsf{max}}(\cdot)$ denotes *maximal correlation* which is defined later in part 3 of Proposition 12, cf. [35], and we use [36, Theorem 3] (or the results of [37]). A similar result also holds for the contraction coefficient for KL divergence with fixed input pmf (see e.g. [36, Definition 1] for a definition).

B. Sufficient conditions

We next portray a sufficient condition for when an additive noise channel $V \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$ is a degraded version of a symmetric channel $W_\delta \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$. By Proposition 3, this is also a sufficient condition for $W_\delta \succeq_{\mathsf{ln}} V$.

Proposition 10 (Degradation by Symmetric Channels)**.** *Given an additive noise channel $V = \textsf{circ}_{\mathcal{X}}(v)$ with noise pmf $v \in \mathcal{P}_q$ and minimum probability $\tau = \min\{[V]_{i,j} : 1 \leq i,j \leq q\}$, we have:*

$$0 \leq \delta \leq (q-1)\,\tau \;\;\Rightarrow\;\; W_\delta \succeq_{\textsf{deg}} V$$

where $W_\delta \in \mathbb{R}_{\textsf{sto}}^{q\times q}$ is a q-ary symmetric channel.

Proof. Using Corollary 1, it suffices to prove that the noise pmf $w_{(q-1)\tau} \succeq_{\mathsf{maj}} v$. Since $0 \leq \tau \leq \frac{1}{q}$, we must have $0 \leq (q-1)\tau \leq \frac{q-1}{q}$. So, all entries of $w_{(q-1)\tau}$, except (possibly) the first, are equal to its minimum entry of τ. As $v \geq \tau$ (entry-wise), $w_{(q-1)\tau} \succeq_{\mathsf{maj}} v$ because the conditions of part 3 in Proposition 13 in Appendix A are satisfied. ∎

It is compelling to find a sufficient condition for $W_\delta \succeq_{\mathsf{ln}} V$ that does not simply ensure $W_\delta \succeq_{\mathsf{deg}} V$ (such as Proposition 10 and Theorem 2). The ensuing proposition elucidates such a sufficient condition for additive noise channels. The general strategy for finding such a condition for additive noise channels is to identify a noise pmf that belongs to $\mathcal{L}_{W_\delta}^{\mathsf{add}} \backslash \mathcal{D}_{W_\delta}^{\mathsf{add}}$. One can then use Proposition 6 to explicitly construct a set of noise pmfs that is a subset of $\mathcal{L}_{W_\delta}^{\mathsf{add}}$ but strictly includes $\mathcal{D}_{W_\delta}^{\mathsf{add}}$. The proof of Proposition 11 finds such a noise pmf (that corresponds to a q-ary symmetric channel).

Proposition 11 (Less Noisy Domination by Symmetric Channels)**.** *Given an additive noise channel $V = \textsf{circ}_{\mathcal{X}}(v)$ with noise pmf $v \in \mathcal{P}_q$ and $q \geq 2$, if for $\delta \in \left[0, \frac{q-1}{q}\right]$ we have:*

$$v \in \textsf{conv}\left(\left\{w_\delta P_q^k : k \in [q]\right\} \cup \left\{w_\gamma P_q^k : k \in [q]\right\}\right)$$

then $W_\delta \succeq_{\textsf{ln}} V$, where $P_q \in \mathbb{R}^{q\times q}$ is defined in (8), and:

$$\gamma = \frac{1-\delta}{1-\delta+\frac{\delta}{(q-1)^2}} \in \left[1-\frac{\delta}{q-1}, 1\right].$$

Proof. Due to Proposition 6 and $\{w_\gamma P_x : x \in \mathcal{X}\} = \{w_\gamma P_q^k : k \in [q]\}$, it suffices to prove that $W_\delta \succeq_{\mathsf{ln}} W_\gamma$. Since $\delta = 0 \Rightarrow \gamma = 1$ and $\delta = \frac{q-1}{q} \Rightarrow \gamma = \frac{q-1}{q}$, $W_\delta \succeq_{\mathsf{ln}} W_\gamma$ is certainly true for $\delta \in \left\{0, \frac{q-1}{q}\right\}$. So, we assume that $\delta \in \left(0, \frac{q-1}{q}\right)$, which implies that:

$$\gamma = \frac{1-\delta}{1-\delta+\frac{\delta}{(q-1)^2}} \in \left(\frac{q-1}{q}, 1\right).$$

Since our goal is to show $W_\delta \succeq_{\mathsf{ln}} W_\gamma$, we prove the equivalent condition in part 2 of Proposition 8 that for every $P_X \in \mathcal{P}_q^\circ$:

$$\begin{aligned}
&W_\delta\, \mathsf{diag}(P_X W_\delta)^{-1}\, W_\delta^T \succeq_{\mathsf{PSD}} W_\gamma\, \mathsf{diag}(P_X W_\gamma)^{-1}\, W_\gamma^T \\
\Leftrightarrow\; & W_\gamma^{-1} \mathsf{diag}(P_X W_\gamma)\, W_\gamma^{-1} \succeq_{\mathsf{PSD}} W_\delta^{-1} \mathsf{diag}(P_X W_\delta)\, W_\delta^{-1} \\
\Leftrightarrow\; & \mathsf{diag}(P_X W_\gamma) \succeq_{\mathsf{PSD}} W_\gamma W_\delta^{-1} \mathsf{diag}(P_X W_\delta)\, W_\delta^{-1} W_\gamma \\
\Leftrightarrow\; & I_q \succeq_{\mathsf{PSD}} \mathsf{diag}(P_X W_\gamma)^{-\frac{1}{2}} W_\tau \mathsf{diag}(P_X W_\delta) W_\tau \mathsf{diag}(P_X W_\gamma)^{-\frac{1}{2}} \\
\Leftrightarrow\; & 1 \geq \left\| \mathsf{diag}(P_X W_\gamma)^{-\frac{1}{2}} W_\tau \mathsf{diag}(P_X W_\delta) W_\tau \mathsf{diag}(P_X W_\gamma)^{-\frac{1}{2}} \right\|_{\mathsf{op}} \\
\Leftrightarrow\; & 1 \geq \left\| \mathsf{diag}(P_X W_\gamma)^{-\frac{1}{2}}\, W_\tau \mathsf{diag}(P_X W_\delta)^{\frac{1}{2}} \right\|_{\mathsf{op}}
\end{aligned}$$

where the second equivalence holds because W_δ and W_γ are symmetric and invertible (see part 4 of Proposition 4 and [19, Corollary 7.7.4]), the third and fourth equivalences are non-singular $*$-congruences with $W_\tau = W_\delta^{-1} W_\gamma = W_\gamma W_\delta^{-1}$ and:

$$\tau = \frac{\gamma - \delta}{1-\delta-\frac{\delta}{q-1}} > 0$$

which can be computed as in the proof of Proposition 15 in Appendix C, and $\|\cdot\|_{\mathsf{op}}$ denotes the spectral or operator norm.[7]

[7]Note that we cannot use the strict Löwner partial order $\succ_{\mathsf{PSD}}$ (for $A, B \in \mathbb{R}_{\mathsf{sym}}^{q\times q}$, $A \succ_{\mathsf{PSD}} B$ if and only if $A - B$ is positive definite) for these equivalences as $\mathbf{1}^T W_\gamma^{-1} \mathsf{diag}(P_X W_\gamma)\, W_\gamma^{-1} \mathbf{1} = \mathbf{1}^T W_\delta^{-1} \mathsf{diag}(P_X W_\delta)\, W_\delta^{-1} \mathbf{1}$.

It is instructive to note that if $W_\tau \in \mathbb{R}_{\textsf{sto}}^{q\times q}$, then the *divergence transition matrix* $\textsf{diag}(P_X W_\gamma)^{-\frac{1}{2}} W_\tau \textsf{diag}(P_X W_\delta)^{\frac{1}{2}}$ has right singular vector $\sqrt{P_X W_\delta}^T$ and left singular vector $\sqrt{P_X W_\gamma}^T$ corresponding to its maximum singular value of unity (where the square roots are applied entry-wise)–see [36] and the references therein. So, $W_\tau \in \mathbb{R}_{\textsf{sto}}^{q\times q}$ is a sufficient condition for $W_\delta \succeq_{\textsf{ln}} W_\gamma$. Since $W_\tau \in \mathbb{R}_{\textsf{sto}}^{q\times q}$ if and only if $0 \leq \tau \leq 1$ if and only if $\delta \leq \gamma \leq 1 - \frac{\delta}{q-1}$, the latter condition also implies that $W_\delta \succeq_{\textsf{ln}} W_\gamma$. However, we recall from (25) in subsection III-B that $W_\delta \succeq_{\textsf{deg}} W_\gamma$ for $\delta \leq \gamma \leq 1 - \frac{\delta}{q-1}$, while we seek some $1 - \frac{\delta}{q-1} < \gamma \leq 1$ for which $W_\delta \succeq_{\textsf{ln}} W_\gamma$. When $q = 2$, we only have:

$$\gamma = \frac{1-\delta}{1-\delta+\frac{\delta}{(q-1)^2}} = 1 - \frac{\delta}{q-1} = 1-\delta$$

which implies that $W_\delta \succeq_{\textsf{deg}} W_\gamma$ is true for $q = 2$. On the other hand, when $q \geq 3$, it is straightforward to verify that:

$$\gamma = \frac{1-\delta}{1-\delta+\frac{\delta}{(q-1)^2}} \in \left(1 - \frac{\delta}{q-1}, 1\right)$$

since $\delta \in \left(0, \frac{q-1}{q}\right)$.

From the preceding discussion, it suffices to prove for $q \geq 3$ that for every $P_X \in \mathcal{P}_q^\circ$:

$$\left\|\textsf{diag}(P_X W_\gamma)^{-\frac{1}{2}} W_\tau \textsf{diag}(P_X W_\delta) W_\tau \textsf{diag}(P_X W_\gamma)^{-\frac{1}{2}}\right\|_{\textsf{op}} \leq 1.$$

Since $\tau > 0$, and $0 \leq \tau \leq 1$ does not produce $\gamma > 1 - \frac{\delta}{q-1}$, we require that $\tau > 1$ ($\Leftrightarrow \gamma > 1 - \frac{\delta}{q-1}$) so that W_τ has strictly negative entries along the diagonal. Notice that:

$$\forall x \in [q], \;\; \textsf{diag}(\Delta_x W_\gamma) \succeq_{\textsf{PSD}} W_\gamma W_\delta^{-1} \textsf{diag}(\Delta_x W_\delta)\, W_\delta^{-1} W_\gamma$$

where $\Delta_x = (0, \dots, 1, \dots, 0) \in \mathcal{P}_q$ denotes the Kronecker delta pmf with unity at the $(x+1)$th position, implies that:

$$\textsf{diag}(P_X W_\gamma) \succeq_{\textsf{PSD}} W_\gamma W_\delta^{-1} \textsf{diag}(P_X W_\delta)\, W_\delta^{-1} W_\gamma$$

for every $P_X \in \mathcal{P}_q^\circ$, because convex combinations preserve the Löwner relation. So, it suffices to prove that for every $x \in [q]$:

$$\left\|\textsf{diag}\big(w_\gamma P_q^x\big)^{-\frac{1}{2}} W_\tau \textsf{diag}\big(w_\delta P_q^x\big) W_\tau \textsf{diag}\big(w_\gamma P_q^x\big)^{-\frac{1}{2}}\right\|_{\textsf{op}} \leq 1$$

where $P_q \in \mathbb{R}^{q\times q}$ is defined in (8), because $\Delta_x M$ extracts the $(x+1)$th row of a matrix $M \in \mathbb{R}^{q\times q}$. Let us define $A_x \triangleq \textsf{diag}\big(w_\gamma P_q^x\big)^{-\frac{1}{2}} W_\tau \textsf{diag}\big(w_\delta P_q^x\big) W_\tau \textsf{diag}\big(w_\gamma P_q^x\big)^{-\frac{1}{2}}$ for each $x \in [q]$. Observe that for every $x \in [q]$, $A_x \in \mathbb{R}_{\succeq 0}^{q\times q}$ is orthogonally diagonalizable by the real spectral theorem [38, Theorem 7.13], and has a strictly positive eigenvector $\sqrt{w_\gamma P_q^x}$ corresponding to the eigenvalue of unity:

$$\forall x \in [q], \;\; \sqrt{w_\gamma P_q^x} A_x = \sqrt{w_\gamma P_q^x}$$

so that all other eigenvectors of A_x have some strictly negative entries since they are orthogonal to $\sqrt{w_\gamma P_q^x}$. Suppose A_x is entry-wise non-negative for every $x \in [q]$. Then, the largest eigenvalue (known as the Perron-Frobenius eigenvalue) and the spectral radius of each A_x is unity by the Perron-Frobenius theorem [19, Theorem 8.3.4], which proves that $\|A_x\|_{\textsf{op}} \leq 1$ for every $x \in [q]$. Therefore, it is sufficient to prove that A_x is entry-wise non-negative for every $x \in [q]$. Equivalently, we can prove that $W_\tau \textsf{diag}\big(w_\delta P_q^x\big) W_\tau$ is entry-wise non-negative for every $x \in [q]$, since $\textsf{diag}\big(w_\gamma P_q^x\big)^{-\frac{1}{2}}$ scales the rows or columns of the matrix it is pre- or post-multiplied with using strictly positive scalars.

We now show the equivalent condition below that the minimum possible entry of $W_\tau \textsf{diag}\big(w_\delta P_q^x\big) W_\tau$ is non-negative:

$$\begin{aligned} 0 &\leq \min_{\substack{x\in[q] \\ 1\leq i,j\leq q}} \underbrace{\sum_{r=1}^{q} [W_\tau]_{i,r} [W_\delta]_{x+1,r} [W_\tau]_{r,j}}_{= \; \left[W_\tau \textsf{diag}\left(w_\delta P_q^x\right) W_\tau\right]_{i,j}} \\ &= \frac{\tau(1-\delta)(1-\tau)}{q-1} + \frac{\delta\tau(1-\tau)}{(q-1)^2} + (q-2)\frac{\delta\tau^2}{(q-1)^3}. \end{aligned} \tag{41}$$

The above equality holds because for $i \neq j$:

$$\frac{\delta}{q-1} \sum_{r=1}^{q} \underbrace{[W_\tau]_{i,r} [W_\tau]_{r,i}}_{= [W_\tau]_{i,r}^2 \geq 0} \geq \frac{\delta}{q-1} \sum_{r=1}^{q} [W_\tau]_{i,r} [W_\tau]_{r,j}$$

is clearly true (using, for example, the rearrangement inequality in [39, Section 10.2]), and adding $\left(1-\delta-\frac{\delta}{q-1}\right)[W_\tau]_{i,k}^2 \geq 0$ (regardless of the value of $1 \leq k \leq q$) to the left summation increases its value, while adding $\left(1-\delta-\frac{\delta}{q-1}\right)[W_\tau]_{i,p}[W_\tau]_{p,j} < 0$ (which exists for an appropriate value $1 \leq p \leq q$ as $\tau > 1$) to the right summation decreases its value. As a result, the minimum possible entry of $W_\tau \mathsf{diag}\left(w_\delta P_q^x\right) W_\tau$ can be achieved with $x+1 = i \neq j$ or $i \neq j = x+1$. We next substitute $\tau = (\gamma-\delta)/\left(1-\delta-\frac{\delta}{q-1}\right)$ into (41) and simplify the resulting expression to get:

$$0 \leq (\gamma-\delta)\left(\left(1-\frac{\delta}{q-1}-\gamma\right)\left(1-\delta+\frac{\delta}{q-1}\right)+\frac{(q-2)\,\delta\,(\gamma-\delta)}{(q-1)^2}\right).$$

The right hand side of this inequality is quadratic in γ with roots $\gamma = \delta$ and $\gamma = \frac{1-\delta}{1-\delta+(\delta/(q-1)^2)}$. Since the coefficient of γ^2 in this quadratic is strictly negative:

$$\underbrace{\frac{(q-2)\,\delta}{(q-1)^2}-\left(1-\delta+\frac{\delta}{q-1}\right)}_{\text{coefficient of } \gamma^2} < 0 \Leftrightarrow 1-\delta+\frac{\delta}{(q-1)^2} > 0$$

the minimum possible entry of $W_\tau \mathsf{diag}\left(w_\delta P_q^x\right) W_\tau$ is non-negative if and only if:

$$\delta \leq \gamma \leq \frac{1-\delta}{1-\delta+\frac{\delta}{(q-1)^2}}$$

where we use the fact that $\frac{1-\delta}{1-\delta+(\delta/(q-1)^2)} \geq 1-\frac{\delta}{q-1} \geq \delta$. Therefore, $\gamma = \frac{1-\delta}{1-\delta+(\delta/(q-1)^2)}$ produces $W_\delta \succeq_{\mathsf{ln}} W_\gamma$, which completes the proof. ∎

Heretofore we have derived results concerning less noisy domination and degradation regions in section III, and proven several necessary and sufficient conditions for less noisy domination of additive noise channels by symmetric channels in this section. We finally have all the pieces in place to establish Theorem 3 from section II. In closing this section, we indicate the pertinent results that coalesce to justify it.

Proof of Theorem 3. The first equality follows from Corollary 1. The first set inclusion is obvious, and its strictness follows from the proof of Proposition 11. The second set inclusion follows from Proposition 11. The third set inclusion follows from the circle condition (part 1) in Proposition 9. Lastly, the properties of $\mathcal{L}_{W_\delta}^{\mathsf{add}}$ are derived in Proposition 6. ∎

VI. Sufficient conditions for degradation over general channels

While Propositions 10 and 11 present sufficient conditions for a symmetric channel $W_\delta \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$ to be less noisy than an additive noise channel, our more comprehensive objective is to find the maximum $\delta \in \left[0, \frac{q-1}{q}\right]$ such that $W_\delta \succeq_{\mathsf{ln}} V$ for any given general channel $V \in \mathbb{R}_{\mathsf{sto}}^{q\times r}$ on a common input alphabet. We may formally define this maximum δ (that characterizes the extremal symmetric channel that is less noisy than V) as:

$$\delta^\star(V) \triangleq \sup\left\{\delta \in \left[0, \frac{q-1}{q}\right] : W_\delta \succeq_{\mathsf{ln}} V\right\} \tag{42}$$

and for every $0 \leq \delta < \delta^\star(V)$, $W_\delta \succeq_{\mathsf{ln}} V$. Alternatively, we can define a non-negative (less noisy) *domination factor* function for any channel $V \in \mathbb{R}_{\mathsf{sto}}^{q\times r}$:

$$\mu_V(\delta) \triangleq \sup_{\substack{P_X, Q_X \in \mathcal{P}_q: \\ 0 < D(P_X W_\delta || Q_X W_\delta) < +\infty}} \frac{D(P_X V || Q_X V)}{D(P_X W_\delta || Q_X W_\delta)} \tag{43}$$

with $\delta \in \left[0, \frac{q-1}{q}\right)$, which is analogous to the contraction coefficient for KL divergence since $\mu_V(0) \triangleq \eta_{\mathsf{KL}}(V)$. Indeed, we may perceive $P_X W_\delta$ and $Q_X W_\delta$ in the denominator of (43) as pmfs inside the "shrunk" simplex $\mathsf{conv}(\{w_\delta P_q^k : k \in [q]\})$, and (43) represents a contraction coefficient of V where the supremum is taken over this "shrunk" simplex.[8] For simplicity, consider a channel $V \in \mathbb{R}_{\mathsf{sto}}^{q\times r}$ that is strictly positive entry-wise, and has domination factor function $\mu_V : \left(0, \frac{q-1}{q}\right) \to \mathbb{R}^+$, where the domain excludes zero because μ_V is only interesting for $\delta \in \left(0, \frac{q-1}{q}\right)$, and this exclusion also affords us some analytical simplicity. It is shown in Proposition 15 of Appendix C that μ_V is always

[8]Pictorially, the "shrunk" simplex is the magenta triangle in Figure 2 while the simplex itself is the larger gray triangle.

finite on $\left(0, \frac{q-1}{q}\right)$, continuous, convex, strictly increasing, and has a vertical asymptote at $\delta = \frac{q-1}{q}$. Since for every $P_X, Q_X \in \mathcal{P}_q$:

$$\mu_V(\delta)\, D(P_X W_\delta || Q_X W_\delta) \geq D(P_X V || Q_X V) \tag{44}$$

we have $\mu_V(\delta) \leq 1$ if and only if $W_\delta \succeq_{\mathsf{ln}} V$. Hence, using the strictly increasing property of $\mu_V : \left(0, \frac{q-1}{q}\right) \to \mathbb{R}^+$, we can also characterize $\delta^\star(V)$ as:

$$\delta^\star(V) = \mu_V^{-1}(1) \tag{45}$$

where μ_V^{-1} denotes the inverse function of μ_V, and unity is in the range of μ_V by Theorem 2 since V is strictly positive entry-wise.

We next briefly delineate how one might computationally approximate $\delta^\star(V)$ for a given general channel $V \in \mathbb{R}^{q\times r}_{\mathsf{sto}}$. From part 2 of Proposition 8, it is straightforward to obtain the following minimax characterization of $\delta^\star(V)$:

$$\delta^\star(V) = \inf_{P_X \in \mathcal{P}_q^\circ} \sup_{\delta \in \mathcal{S}(P_X)} \delta \tag{46}$$

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 in (46) can be naïvely approximated by sampling several $P_X \in \mathcal{P}_q^\circ$. The supremum in (46) can be estimated by verifying collections of rational (ratio of polynomials) inequalities in δ. This is because the positive semidefiniteness of a matrix is equivalent to the non-negativity of all its principal minors by *Sylvester's criterion* [19, Theorem 7.2.5]. Unfortunately, this procedure appears to be rather cumbersome.

Since analytically computing $\delta^\star(V)$ also seems intractable, we now prove Theorem 2 from section II. Theorem 2 provides a sufficient condition for $W_\delta \succeq_{\mathsf{deg}} V$ (which implies $W_\delta \succeq_{\mathsf{ln}} V$ using Proposition 3) by restricting its attention to the case where $V \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ with $q \geq 2$. Moreover, it can be construed as a lower bound on $\delta^\star(V)$:

$$\delta^\star(V) \geq \frac{\nu}{1 - (q-1)\nu + \frac{\nu}{q-1}} \tag{47}$$

where $\nu = \min\{[V]_{i,j} : 1 \leq i, j \leq q\}$ is the minimum conditional probability in V.

Proof of Theorem 2. Let the channel $V \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ have the conditional pmfs $v_1, \ldots, v_q \in \mathcal{P}_q$ as its rows:

$$V = \begin{bmatrix} v_1^T & v_2^T & \cdots & v_q^T \end{bmatrix}^T.$$

From the proof of Proposition 10, we know that $w_{(q-1)\nu} \succeq_{\mathsf{maj}} v_i$ for every $i \in \{1, \ldots, q\}$. Using part 1 of Proposition 13 in Appendix A (and the fact that the set of all permutations of $w_{(q-1)\nu}$ is exactly the set of all cyclic permutations of $w_{(q-1)\nu}$), we can write this as:

$$\forall i \in \{1, \ldots, q\}, \;\; v_i = \sum_{j=1}^{q} p_{i,j}\, w_{(q-1)\nu} P_q^{j-1}$$

where $P_q \in \mathbb{R}^{q\times q}$ is given in (8), and $\{p_{i,j} \geq 0 : 1 \leq i, j \leq q\}$ are the convex weights such that $\sum_{j=1}^q p_{i,j} = 1$ for every $i \in \{1, \ldots, q\}$. Defining $P \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ entry-wise as $[P]_{i,j} = p_{i,j}$ for every $1 \leq i, j \leq q$, we can also write this equation as $V = P W_{(q-1)\nu}$.[9] Observe that:

$$P = \sum_{1 \leq j_1, \ldots, j_q \leq q} \left(\prod_{i=1}^{q} p_{i,j_i}\right) E_{j_1,\ldots,j_q}$$

where $\{\prod_{i=1}^q p_{i,j_i} : 1 \leq j_1, \ldots, j_q \leq q\}$ form a product pmf of convex weights, and for every $1 \leq j_1, \ldots, j_q \leq q$:

$$E_{j_1,\ldots,j_q} \triangleq \begin{bmatrix} e_{j_1} & e_{j_2} & \cdots & e_{j_q} \end{bmatrix}^T$$

where $e_i \in \mathbb{R}^q$ is the ith standard basis (column) vector that has unity at the ith entry and zero elsewhere. Hence, we get:

$$V = \sum_{1 \leq j_1, \ldots, j_q \leq q} \left(\prod_{i=1}^{q} p_{i,j_i}\right) E_{j_1,\ldots,j_q} W_{(q-1)\nu}.$$

[9] Matrices of the form $V = P W_{(q-1)\nu}$ with $P \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ are *not* necessarily degraded versions of $W_{(q-1)\nu}$: $W_{(q-1)\nu} \not\succeq_{\mathsf{deg}} V$ (although we certainly have input-output degradation: $W_{(q-1)\nu} \succeq_{\mathsf{iod}} V$). As a counterexample, consider $W_{1/2}$ for $q = 3$, and $P = [1\,0\,0;\, 1\,0\,0;\, 0\,1\,0]$, where the colons separate the rows of the matrix. If $W_{1/2} \succeq_{\mathsf{deg}} P W_{1/2}$, then there exists $A \in \mathbb{R}^{3\times 3}_{\mathsf{sto}}$ such that $P W_{1/2} = W_{1/2} A$. However, $A = W_{1/2}^{-1} P W_{1/2} = (1/4)\,[3\,0\,1;\, 3\,0\,1;\, -1\,4\,1]$ has a strictly negative entry, which leads to a contradiction.

Suppose there exists $\delta \in \left[0, \frac{q-1}{q}\right]$ such that for all $j_1, \ldots, j_q \in \{1, \ldots, q\}$:

$$\exists M_{j_1,\ldots,j_q} \in \mathbb{R}^{q \times q}_{\mathsf{sto}}, \;\; E_{j_1,\ldots,j_q} W_{(q-1)\nu} = W_\delta M_{j_1,\ldots,j_q}$$

i.e. $W_\delta \succeq_{\mathsf{deg}} E_{j_1,\ldots,j_q} W_{(q-1)\nu}$. Then, we would have:

$$V = W_\delta \underbrace{\sum_{1 \leq j_1,\ldots,j_q \leq q} \left(\prod_{i=1}^{q} p_{i,j_i} \right) M_{j_1,\ldots,j_q}}_{\text{stochastic matrix}}$$

which implies that $W_\delta \succeq_{\mathsf{deg}} V$.

We will demonstrate that for every $j_1, \ldots, j_q \in \{1, \ldots, q\}$, there exists $M_{j_1,\ldots,j_q} \in \mathbb{R}^{q \times q}_{\mathsf{sto}}$ such that $E_{j_1,\ldots,j_q} W_{(q-1)\nu} = W_\delta M_{j_1,\ldots,j_q}$ when $0 \leq \delta \leq \nu / \big(1 - (q-1)\nu + \frac{\nu}{q-1}\big)$. Since $0 \leq \nu \leq \frac{1}{q}$, the preceding inequality implies that $0 \leq \delta \leq \frac{q-1}{q}$, where $\delta = \frac{q-1}{q}$ is possible if and only if $\nu = \frac{1}{q}$. When $\nu = \frac{1}{q}$, $V = W_{(q-1)/q}$ is the channel with all uniform conditional pmfs, and $W_{(q-1)/q} \succeq_{\mathsf{deg}} V$ clearly holds. Hence, we assume that $0 \leq \nu < \frac{1}{q}$ so that $0 \leq \delta < \frac{q-1}{q}$, and establish the equivalent condition that for every $j_1, \ldots, j_q \in \{1, \ldots, q\}$:

$$M_{j_1,\ldots,j_q} = W_\delta^{-1} E_{j_1,\ldots,j_q} W_{(q-1)\nu}$$

is a valid stochastic matrix. Recall that $W_\delta^{-1} = W_\tau$ with $\tau = \frac{-\delta}{1-\delta-(\delta/(q-1))}$ using part 4 of Proposition 4. Clearly, all the rows of each $M_{j_1,\ldots,j_q}$ sum to unity. So, it remains to verify that each $M_{j_1,\ldots,j_q}$ has non-negative entries. For any $j_1, \ldots, j_q \in \{1, \ldots, q\}$ and any $i, j \in \{1, \ldots, q\}$:

$$\left[M_{j_1,\ldots,j_q}\right]_{i,j} \geq \nu\,(1-\tau) + \tau\,(1 - (q-1)\,\nu)$$

where the right hand side is the minimum possible entry of any $M_{j_1,\ldots,j_q}$ (with equality when $j_1 > 1$ and $j_2 = j_3 = \cdots = j_q = 1$ for example) as $\tau < 0$ and $1 - (q-1)\,\nu > \nu$. To ensure each $M_{j_1,\ldots,j_q}$ is entry-wise non-negative, the minimum possible entry must satisfy:

$$\begin{aligned} & \nu\,(1-\tau) + \tau\,(1-(q-1)\,\nu) \geq 0 \\ \Leftrightarrow \;\; & \nu + \frac{\delta\nu}{1-\delta-\frac{\delta}{q-1}} - \frac{\delta\,(1-(q-1)\,\nu)}{1-\delta-\frac{\delta}{q-1}} \geq 0 \end{aligned}$$

and the latter inequality is equivalent to:

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

This completes the proof. ■

We remark that if $V = E_{2,1,\ldots,1} W_{(q-1)\nu} \in \mathbb{R}^{q \times q}_{\mathsf{sto}}$, then this proof illustrates that $W_\delta \succeq_{\mathsf{deg}} V$ if and only if $0 \leq \delta \leq \nu / \big(1 - (q-1)\nu + \frac{\nu}{q-1}\big)$. Hence, the condition in Theorem 2 is tight when no further information about V is known. It is worth juxtaposing Theorem 2 and Proposition 10. The upper bounds on δ from these results satisfy:

$$\underbrace{\frac{\nu}{1-(q-1)\nu + \frac{\nu}{q-1}}}_{\text{upper bound in Theorem 2}} \leq \underbrace{(q-1)\,\nu}_{\substack{\text{upper bound in} \\ \text{Proposition 10}}} \tag{48}$$

where we have equality if and only if $\nu = \frac{1}{q}$, and it is straightforward to verify that (48) is equivalent to $\nu \leq \frac{1}{q}$. Moreover, assuming that q is large and $\nu = o\,(1/q)$, the upper bound in Theorem 2 is $\nu / \big(1 + o\,(1) + o\,(1/q^2)\big) = \Theta\,(\nu)$, while the upper bound in Proposition 10 is $\Theta\,(q\nu)$.[10] (Note that both bounds are $\Theta\,(1)$ if $\nu = \frac{1}{q}$.) Therefore, when $V \in \mathbb{R}^{q \times q}_{\mathsf{sto}}$ is an additive noise channel, $\delta = O\,(q\nu)$ is enough for $W_\delta \succeq_{\mathsf{deg}} V$, but a general channel $V \in \mathbb{R}^{q \times q}_{\mathsf{sto}}$ requires $\delta = O\,(\nu)$ for such degradation. So, in order to account for q different conditional pmfs in the general case (as opposed to a single conditional pmf which characterizes the channel in the additive noise case), we loose a factor of q in the upper bound on δ. Furthermore, we can check using simulations that $W_\delta \in \mathbb{R}^{q \times q}_{\mathsf{sto}}$ is not in general less noisy than $V \in \mathbb{R}^{q \times q}_{\mathsf{sto}}$ for $\delta = (q-1)\nu$. Indeed, counterexamples can be easily obtained by letting $V = E_{j_1,\ldots,j_q} W_\delta$ for specific values of $1 \leq j_1, \ldots, j_q \leq q$, and computationally verifying that $W_\delta \not\succeq_{\mathsf{ln}} V + J \in \mathbb{R}^{q \times q}_{\mathsf{sto}}$ for appropriate choices of perturbation matrices $J \in \mathbb{R}^{q \times q}$ with sufficiently small Frobenius norm.

[10] We use the *Bachmann-Landau asymptotic notation* here. Consider the (strictly) positive functions $f : \mathbb{N} \to \mathbb{R}$ and $g : \mathbb{N} \to \mathbb{R}$. The little-$o$ notation is defined as: $f(q) = o\,(g(q)) \Leftrightarrow \lim_{q\to\infty} f(q)/g(q) = 0$. The big-$O$ notation is defined as: $f(q) = O\,(g(q)) \Leftrightarrow \limsup_{q\to\infty} |f(q)/g(q)| < +\infty$. Finally, the big-$\Theta$ notation is defined as: $f(q) = \Theta\,(g(q)) \Leftrightarrow 0 < \liminf_{q\to\infty} |f(q)/g(q)| \leq \limsup_{q\to\infty} |f(q)/g(q)| < +\infty$.

We have now proved Theorems 1, 2, and 3 from section II. The next section relates our results regarding less noisy and degradation preorders to LSIs, and proves Theorem 4.

VII. LESS NOISY DOMINATION AND LOGARITHMIC SOBOLEV INEQUALITIES

Logarithmic Sobolev inequalities (LSIs) are a class of functional inequalities that shed light on several important phenomena such as concentration of measure, and ergodicity and hypercontractivity of Markov semigroups. We refer readers to [40] and [41] for a general treatment of such inequalities, and more pertinently to [25] and [26], which present LSIs in the context of finite state-space Markov chains. In this section, we illustrate that proving a channel $W \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$ is less noisy than a channel $V \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$ allows us to translate an LSI for W to an LSI for V. Thus, important information about V can be deduced (from its LSI) by proving $W \succeq_{\mathsf{ln}} V$ for an appropriate channel W (such as a q-ary symmetric channel) that has a known LSI.

We commence by introducing some appropriate notation and terminology associated with LSIs. For fixed input and output alphabet $\mathcal{X} = \mathcal{Y} = [q]$ with $q \in \mathbb{N}$, we think of a channel $W \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$ as a Markov kernel on $\mathcal{X}$. We assume that the "time homogeneous" discrete-time Markov chain defined by W is *irreducible*, and has unique *stationary distribution* (or invariant measure) $\pi \in \mathcal{P}_q$ such that $\pi W = \pi$. Furthermore, we define the Hilbert space $\mathcal{L}^2(\mathcal{X},\pi)$ of all real functions with domain $\mathcal{X}$ endowed with the inner product:

$$\forall f, g \in \mathcal{L}^2(\mathcal{X},\pi), \;\; \langle f, g\rangle_\pi \triangleq \sum_{x\in\mathcal{X}} \pi(x) f(x) g(x) \tag{49}$$

and induced norm $\|\cdot\|_\pi$. We construe $W : \mathcal{L}^2(\mathcal{X},\pi) \to \mathcal{L}^2(\mathcal{X},\pi)$ as a conditional expectation operator that takes a function $f \in \mathcal{L}^2(\mathcal{X},\pi)$, which we can write as a column vector $f = [f(0)\cdots f(q-1)]^T \in \mathbb{R}^q$, to another function $Wf \in \mathcal{L}^2(\mathcal{X},\pi)$, which we can also write as a column vector $Wf \in \mathbb{R}^q$. Corresponding to the discrete-time Markov chain W, we may also define a continuous-time *Markov semigroup*:

$$\forall t \geq 0, \;\; H_t \triangleq \exp\left(-t\left(I_q - W\right)\right) \in \mathbb{R}_{\mathsf{sto}}^{q\times q} \tag{50}$$

where the "discrete-time derivative" $W - I_q$ is the *Laplacian operator* that forms the *generator* of the Markov semigroup. The unique stationary distribution of this Markov semigroup is also π, and we may interpret $H_t : \mathcal{L}^2(\mathcal{X},\pi) \to \mathcal{L}^2(\mathcal{X},\pi)$ as a conditional expectation operator for each $t \geq 0$ as well.

In order to present LSIs, we define the *Dirichlet form* $\mathcal{E}_W : \mathcal{L}^2(\mathcal{X},\pi) \times \mathcal{L}^2(\mathcal{X},\pi) \to \mathbb{R}$:

$$\forall f, g \in \mathcal{L}^2(\mathcal{X},\pi), \;\; \mathcal{E}_W(f,g) \triangleq \langle (I_q - W) f, g\rangle_\pi \tag{51}$$

which is used to study properties of the Markov chain W and its associated Markov semigroup $\left\{H_t \in \mathbb{R}_{\mathsf{sto}}^{q\times q} : t \geq 0\right\}$. ($\mathcal{E}_W$ is technically only a Dirichlet form when W is a *reversible* Markov chain, i.e. W is a self-adjoint operator, or equivalently, W and π satisfy the *detailed balance condition* [25, Section 2.3, page 705].) Moreover, the quadratic form defined by $\mathcal{E}_W$ represents the energy of its input function, and satisfies:

$$\forall f \in \mathcal{L}^2(\mathcal{X},\pi), \, \mathcal{E}_W(f,f) = \left\langle \left(I_q - \frac{W + W^*}{2}\right) f, f\right\rangle_\pi \tag{52}$$

where $W^* : \mathcal{L}^2(\mathcal{X},\pi) \to \mathcal{L}^2(\mathcal{X},\pi)$ is the adjoint operator of W. Finally, we introduce a particularly important Dirichlet form corresponding to the channel $W_{(q-1)/q} = \mathbf{1u}$, which has all uniform conditional pmfs and uniform stationary distribution $\pi = \mathbf{u}$, known as the *standard Dirichlet form*:

$$\begin{aligned} \mathcal{E}_{\mathsf{std}}(f,g) &\triangleq \mathcal{E}_{\mathbf{1u}}(f,g) = \mathbb{COV}_{\mathbf{u}}(f,g) \\ &= \sum_{x\in\mathcal{X}} \frac{f(x)g(x)}{q} - \left(\sum_{x\in\mathcal{X}} \frac{f(x)}{q}\right)\left(\sum_{x\in\mathcal{X}} \frac{g(x)}{q}\right) \end{aligned} \tag{53}$$

for any $f, g \in \mathcal{L}^2(\mathcal{X},\mathbf{u})$. The quadratic form defined by the standard Dirichlet form is presented in (17) in subsection II-D.

We now present the LSIs associated with the Markov chain W and the Markov semigroup $\left\{H_t \in \mathbb{R}_{\mathsf{sto}}^{q\times q} : t \geq 0\right\}$ it defines. The LSI for the Markov semigroup with constant $\alpha \in \mathbb{R}$ states that for every $f \in \mathcal{L}^2(\mathcal{X},\pi)$ such that $\|f\|_\pi = 1$, we have:

$$D\left(f^2\pi \,||\, \pi\right) = \sum_{x\in\mathcal{X}} \pi(x) f^2(x) \log\left(f^2(x)\right) \leq \frac{1}{\alpha}\mathcal{E}_W(f,f) \tag{54}$$

where we construe $\mu = f^2\pi \in \mathcal{P}_q$ as a pmf such that $\mu(x) = f(x)^2\pi(x)$ for every $x \in \mathcal{X}$, and f^2 behaves like the Radon-Nikodym derivative (or density) of μ with respect to π. The largest constant α such that (54) holds:

$$\alpha(W) \triangleq \inf_{\substack{f\in\mathcal{L}^2(\mathcal{X},\pi):\\ \|f\|_\pi=1\\ D(f^2\pi||\pi)\neq 0}} \frac{\mathcal{E}_W(f,f)}{D(f^2\pi\,||\,\pi)} \tag{55}$$

is called the *logarithmic Sobolev constant* (LSI constant) of the Markov chain W (or the Markov chain $(W + W^*)/2$). Likewise, the LSI for the discrete-time Markov chain with constant $\alpha \in \mathbb{R}$ states that for every $f \in \mathcal{L}^2(\mathcal{X}, \pi)$ such that $\|f\|_\pi = 1$, we have:

$$D\left(f^2\pi\,||\,\pi\right) \leq \frac{1}{\alpha}\mathcal{E}_{WW^*}(f,f) \tag{56}$$

where $\mathcal{E}_{WW^*} : \mathcal{L}^2(\mathcal{X},\pi) \times \mathcal{L}^2(\mathcal{X},\pi) \to \mathbb{R}$ is the "discrete" Dirichlet form. The largest constant α such that (56) holds is the LSI constant of the Markov chain WW^*, $\alpha(WW^*)$, and we refer to it as the *discrete logarithmic Sobolev constant* of the Markov chain W. As we mentioned earlier, there are many useful consequences of such LSIs. For example, if (54) holds with constant (55), then for every pmf $\mu \in \mathcal{P}_q$:

$$\forall t \geq 0,\;\; D(\mu H_t||\pi) \leq e^{-2\alpha(W)t} D(\mu||\pi) \tag{57}$$

where the exponent $2\alpha(W)$ can be improved to $4\alpha(W)$ if W is reversible [25, Theorem 3.6]. This is a measure of ergodicity of the semigroup $\left\{H_t \in \mathbb{R}^{q\times q}_{\mathsf{sto}} : t \geq 0\right\}$. Likewise, if (56) holds with constant $\alpha(WW^*)$, then for every pmf $\mu \in \mathcal{P}_q$:

$$\forall n \in \mathbb{N},\;\; D(\mu W^n||\pi) \leq (1 - \alpha(WW^*))^n\, D(\mu||\pi) \tag{58}$$

as mentioned in [25, Remark, page 725] and proved in [42]. This is also a measure of ergodicity of the Markov chain W.

Although LSIs have many useful consequences, LSI constants are difficult to compute analytically. Fortunately, the LSI constant corresponding to $\mathcal{E}_{\mathsf{std}}$ has been computed in [25, Appendix, Theorem A.1]. Therefore, using the relation in (18), we can compute LSI constants for q-ary symmetric channels as well. The next proposition collects the LSI constants for q-ary symmetric channels (which are irreducible for $\delta \in (0, 1]$) as well as some other related quantities.

Proposition 12 (Constants of Symmetric Channels)**.** *The q-ary symmetric channel $W_\delta \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ with $q \geq 2$ has:*

1) *LSI constant:*
$$\alpha(W_\delta) = \begin{cases} \delta, & q = 2 \\ \frac{(q-2)\delta}{(q-1)\log(q-1)}, & q > 2 \end{cases}$$
for $\delta \in (0, 1]$.
2) *discrete LSI constant:*
$$\alpha(W_\delta W_\delta^*) = \alpha(W_{\delta'}) = \begin{cases} 2\delta(1-\delta), & q = 2 \\ \frac{(q-2)(2q-2-q\delta)\delta}{(q-1)^2\log(q-1)}, & q > 2 \end{cases}$$
for $\delta \in (0, 1]$, where $\delta' = \delta\big(2 - \frac{q\delta}{q-1}\big)$.
3) *Hirschfeld-Gebelein-Rényi maximal correlation corresponding to the uniform stationary distribution $\boldsymbol{u} \in \mathcal{P}_q$:*
$$\rho_{\mathsf{max}}(\boldsymbol{u}, W_\delta) = \left|1 - \delta - \frac{\delta}{q-1}\right|$$
for $\delta \in [0, 1]$, where for any channel $W \in \mathbb{R}^{q\times r}_{\mathsf{sto}}$ and any source pmf $P_X \in \mathcal{P}_q$, we define the maximal correlation between the input random variable $X \in [q]$ and the output random variable $Y \in [r]$ (with joint pmf $P_{X,Y}(x, y) = P_X(x)W_{Y|X}(y|x)$) as:
$$\rho_{\mathsf{max}}(P_X, W) \triangleq \sup_{\substack{f:[q]\to\mathbb{R},\, g:[r]\to\mathbb{R}\\ \mathbb{E}[f(X)]=\mathbb{E}[g(Y)]=0\\ \mathbb{E}[f^2(X)]=\mathbb{E}[g^2(Y)]=1}} \mathbb{E}\left[f(X)g(Y)\right].$$
4) *contraction coefficient for KL divergence bounded by:*
$$\left(1 - \delta - \frac{\delta}{q-1}\right)^2 \leq \eta_{\mathsf{KL}}(W_\delta) \leq \left|1 - \delta - \frac{\delta}{q-1}\right|$$

for $\delta \in [0,1]$.

Proof. See Appendix B. ■

In view of Proposition 12 and the intractability of computing LSI constants for general Markov chains, we often "compare" a given irreducible channel $V \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ with a q-ary symmetric channel $W_\delta \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ to try and establish an LSI for it. We assume for the sake of simplicity that V is doubly stochastic and has uniform stationary pmf (just like q-ary symmetric channels). Usually, such a comparison between W_δ and V requires us to prove domination of Dirichlet forms, such as:

$$\forall f \in \mathcal{L}^2(\mathcal{X}, \mathbf{u}), \;\; \mathcal{E}_V(f,f) \geq \mathcal{E}_{W_\delta}(f,f) = \frac{q\delta}{q-1}\mathcal{E}_{\mathsf{std}}(f,f) \tag{59}$$

where we use (18). Such pointwise domination results immediately produce LSIs, (54) and (56), for V. Furthermore, they also lower bound the LSI constants of V; for example:

$$\alpha(V) \geq \alpha(W_\delta)\,. \tag{60}$$

This is turn begets other results such as (57) and (58) for the channel V (albeit with worse constants in the exponents since the LSI constants of W_δ are used instead of those for V). More general versions of Dirichlet form domination between Markov chains on different state spaces with different stationary distributions, and the resulting bounds on their LSI constants are presented in [25, Lemmata 3.3 and 3.4]. We next illustrate that the information theoretic notion of less noisy domination is a sufficient condition for various kinds of pointwise Dirichlet form domination.

Theorem 4′ (Domination of Dirichlet Forms)**.** *Let $W, V \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ be doubly stochastic channels, and $\pi = \boldsymbol{u}$ be the uniform stationary distribution. Then, the following are true:*

1) *If $W \succeq_{\mathsf{ln}} V$, then:*
$$\forall f \in \mathcal{L}^2(\mathcal{X}, \boldsymbol{u}), \;\; \mathcal{E}_{VV^*}(f,f) \geq \mathcal{E}_{WW^*}(f,f)\,.$$
2) *If $W \in \mathbb{R}^{q\times q}_{\succeq 0}$ is positive semidefinite, V is normal (i.e. $V^TV = VV^T$), and $W \succeq_{\mathsf{ln}} V$, then:*
$$\forall f \in \mathcal{L}^2(\mathcal{X}, \boldsymbol{u}), \;\; \mathcal{E}_V(f,f) \geq \mathcal{E}_W(f,f)\,.$$
3) *If $W = W_\delta \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ is any q-ary symmetric channel with $\delta \in \left[0, \frac{q-1}{q}\right]$ and $W_\delta \succeq_{\mathsf{ln}} V$, 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.
Part 1: First observe that:

$$\forall f \in \mathcal{L}^2(\mathcal{X}, \mathbf{u}), \;\; \mathcal{E}_{WW^*}(f,f) = \frac{1}{q}f^T\left(I_q - WW^T\right)f$$
$$\forall f \in \mathcal{L}^2(\mathcal{X}, \mathbf{u}), \;\; \mathcal{E}_{VV^*}(f,f) = \frac{1}{q}f^T\left(I_q - VV^T\right)f$$

where we use the facts that $W^T = W^*$ and $V^T = V^*$ because the stationary distribution is uniform. This implies that $\mathcal{E}_{VV^*}(f,f) \geq \mathcal{E}_{WW^*}(f,f)$ for every $f \in \mathcal{L}^2(\mathcal{X}, \mathbf{u})$ if and only if $I_q - VV^T \succeq_{\mathsf{PSD}} I_q - WW^T$, which is true if and only if $WW^T \succeq_{\mathsf{PSD}} VV^T$. Since $W \succeq_{\mathsf{ln}} V$, we get $WW^T \succeq_{\mathsf{PSD}} VV^T$ from part 2 of Proposition 8 after letting $P_X = \mathbf{u} = P_XW = P_XV$.
Part 2: Once again, we first observe using (52) that:

$$\forall f \in \mathcal{L}^2(\mathcal{X}, \mathbf{u}), \;\; \mathcal{E}_W(f,f) = \frac{1}{q}f^T\left(I_q - \frac{W + W^T}{2}\right)f,$$
$$\forall f \in \mathcal{L}^2(\mathcal{X}, \mathbf{u}), \;\; \mathcal{E}_V(f,f) = \frac{1}{q}f^T\left(I_q - \frac{V + V^T}{2}\right)f.$$

So, $\mathcal{E}_V(f,f) \geq \mathcal{E}_W(f,f)$ for every $f \in \mathcal{L}^2(\mathcal{X}, \mathbf{u})$ if and only if $\left(W + W^T\right)/2 \succeq_{\mathsf{PSD}} \left(V + V^T\right)/2$. Since $WW^T \succeq_{\mathsf{PSD}} VV^T$ from the proof of part 1, it is sufficient to prove that:

$$WW^T \succeq_{\mathsf{PSD}} VV^T \;\; \Rightarrow \;\; \frac{W + W^T}{2} \succeq_{\mathsf{PSD}} \frac{V + V^T}{2}. \tag{61}$$

Lemma 2 in Appendix C establishes the claim in (61) because $W \in \mathbb{R}^{q\times q}_{\succeq 0}$ and V is a normal matrix.
Part 3: We note that when V is a normal matrix, this result follows from part 2 because $W_\delta \in \mathbb{R}^{q\times q}_{\succeq 0}$ for $\delta \in \left[0, \frac{q-1}{q}\right]$,

as can be seen from part 2 of Proposition 4. For a general doubly stochastic channel V, we need to prove that $\mathcal{E}_V\left(f,f\right) \geq \mathcal{E}_{W_\delta}\left(f,f\right) = \frac{q\delta}{q-1}\mathcal{E}_{\mathsf{std}}(f,f)$ for every $f \in \mathcal{L}^2\left(\mathcal{X},\mathbf{u}\right)$ (where we use (18)). Following the proof of part 2, it is sufficient to prove (61) with $W = W_\delta$:[11]

$$W_\delta^2 \succeq_{\mathsf{PSD}} VV^T \;\;\Rightarrow\;\; W_\delta \succeq_{\mathsf{PSD}} \frac{V+V^T}{2}$$

where $W_\delta^2 = W_\delta W_\delta^T$ and $W_\delta = \left(W_\delta + W_\delta^T\right)/2$. Recall the *Löwner-Heinz theorem* [43], [44], (cf. [45, Section 6.6, Problem 17]), which states that for $A, B \in \mathbb{R}^{q\times q}_{\succeq 0}$ and $0 \leq p \leq 1$:

$$A \succeq_{\mathsf{PSD}} B \;\;\Rightarrow\;\; A^p \succeq_{\mathsf{PSD}} B^p \tag{62}$$

or equivalently, $f : [0,\infty) \to \mathbb{R}$, $f(x) = x^p$ is an *operator monotone* function for $p \in [0,1]$. Using (62) with $p = \frac{1}{2}$ (cf. [19, Corollary 7.7.4 (b)]), we have:

$$W_\delta^2 \succeq_{\mathsf{PSD}} VV^T \;\;\Rightarrow\;\; W_\delta \succeq_{\mathsf{PSD}} \left(VV^T\right)^{\frac{1}{2}}$$

because the Gramian matrix $VV^T \in \mathbb{R}^{q\times q}_{\succeq 0}$. (Here, $\left(VV^T\right)^{\frac{1}{2}}$ is the unique positive semidefinite square root matrix of VV^T.)

Let $VV^T = Q\Lambda Q^T$ and $(V+V^T)/2 = U\Sigma U^T$ be the spectral decompositions of VV^T and $(V+V^T)/2$, where Q and U are orthogonal matrices with eigenvectors as columns, and Λ and Σ are diagonal matrices of eigenvalues. Since VV^T and $(V+V^T)/2$ are both doubly stochastic, they both have the unit norm eigenvector $\mathbf{1}/\sqrt{q}$ corresponding to the maximum eigenvalue of unity. In fact, we have:

$$\left(VV^T\right)^{\frac{1}{2}}\frac{\mathbf{1}}{\sqrt{q}} = \frac{\mathbf{1}}{\sqrt{q}} \quad \text{and} \quad \left(\frac{V+V^T}{2}\right)\frac{\mathbf{1}}{\sqrt{q}} = \frac{\mathbf{1}}{\sqrt{q}}$$

where we use the fact that $(VV^T)^{\frac{1}{2}} = Q\Lambda^{\frac{1}{2}}Q^T$ is the spectral decomposition of $(VV^T)^{\frac{1}{2}}$. For any matrix $A \in \mathbb{R}^{q\times q}_{\mathsf{sym}}$, let $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_q(A)$ denote the eigenvalues of A in descending order. Without loss of generality, we assume that $[\Lambda]_{j,j} = \lambda_j(VV^T)$ and $[\Sigma]_{j,j} = \lambda_j((V+V^T)/2)$ for every $1 \leq j \leq q$. So, $\lambda_1((VV^T)^{\frac{1}{2}}) = \lambda_1((V+V^T)/2) = 1$, and the first columns of both Q and U are equal to $\mathbf{1}/\sqrt{q}$.

From part 2 of Proposition 4, we have $W_\delta = QDQ^T = UDU^T$, where D is the diagonal matrix of eigenvalues such that $[D]_{1,1} = \lambda_1(W_\delta) = 1$ and $[D]_{j,j} = \lambda_j(W_\delta) = 1 - \delta - \frac{\delta}{q-1}$ for $2 \leq j \leq q$. Note that we may use either of the eigenbases, Q or U, because they both have first column $\mathbf{1}/\sqrt{q}$, which is the eigenvector of W_δ corresponding to $\lambda_1(W_\delta) = 1$ since W_δ is doubly stochastic, and the remaining eigenvector columns are permitted to be any orthonormal basis of $\mathsf{span}(\mathbf{1}/\sqrt{q})^\perp$ as $\lambda_j(W_\delta) = 1 - \delta - \frac{\delta}{q-1}$ for $2 \leq j \leq q$. Hence, we have:

$$W_\delta \succeq_{\mathsf{PSD}} \left(VV^T\right)^{\frac{1}{2}} \Leftrightarrow QDQ^T \succeq_{\mathsf{PSD}} Q\Lambda^{\frac{1}{2}}Q^T \Leftrightarrow D \succeq_{\mathsf{PSD}} \Lambda^{\frac{1}{2}},$$
$$W_\delta \succeq_{\mathsf{PSD}} \frac{V+V^T}{2} \Leftrightarrow UDU^T \succeq_{\mathsf{PSD}} U\Sigma U^T \Leftrightarrow D \succeq_{\mathsf{PSD}} \Sigma.$$

In order to show that $D \succeq_{\mathsf{PSD}} \Lambda^{\frac{1}{2}} \Rightarrow D \succeq_{\mathsf{PSD}} \Sigma$, it suffices to prove that $\Lambda^{\frac{1}{2}} \succeq_{\mathsf{PSD}} \Sigma$. Recall from [45, Corollary 3.1.5] that for any matrix $A \in \mathbb{R}^{q\times q}$, we have:[12]

$$\forall i \in \{1,\ldots,q\}\,,\;\; \lambda_i\left(\left(AA^T\right)^{\frac{1}{2}}\right) \geq \lambda_i\left(\frac{A+A^T}{2}\right). \tag{63}$$

Hence, $\Lambda^{\frac{1}{2}} \succeq_{\mathsf{PSD}} \Sigma$ is true, cf. [46, Lemma 2.5]. This completes the proof. ∎

Theorem 4′ includes Theorem 4 from section II as part 3, and also provides two other useful pointwise Dirichlet form domination results. Part 1 of Theorem 4′ states that less noisy domination implies discrete Dirichlet form domination. In particular, if we have $W_\delta \succeq_{\mathsf{ln}} V$ for some irreducible q-ary symmetric channel $W_\delta \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ and irreducible doubly stochastic channel $V \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$, then part 1 implies that:

$$\forall n \in \mathbb{N},\;\; D\left(\mu V^n || \mathbf{u}\right) \leq \left(1 - \alpha(W_\delta W_\delta^*)\right)^n D\left(\mu || \mathbf{u}\right) \tag{64}$$

[11]Note that (61) trivially holds for $W = W_\delta$ with $\delta = (q-1)/q$, because $W_{(q-1)/q} = W^2_{(q-1)/q} = \mathbf{1u} \succeq_{\mathsf{PSD}} VV^T$ implies that $V = W_{(q-1)/q}$.

[12]This states that for any matrix $A \in \mathbb{R}^{q\times q}$, the ith largest eigenvalue of the symmetric part of A is less than or equal to the ith largest singular value of A (which is the ith largest eigenvalue of the unique positive semidefinite part $(AA^T)^{1/2}$ in the polar decomposition of A) for every $1 \leq i \leq q$.

for all pmfs $\mu \in \mathcal{P}_q$, where $\alpha(W_\delta W_\delta^*)$ is computed in part 2 of Proposition 12. However, it is worth mentioning that (58) for W_δ and Proposition 1 directly produce (64). So, such ergodicity results for the discrete-time Markov chain V do not require the full power of the Dirichlet form domination in part 1. Regardless, Dirichlet form domination results, such as in parts 2 and 3, yield several functional inequalities (like Poincaré inequalities and LSIs) which have many other potent consequences as well.

Parts 2 and 3 of Theorem $4'$ convey that less noisy domination also implies the usual (continuous) Dirichlet form domination under regularity conditions. We note that in part 2, the channel W is more general than that in part 3, but the channel V is restricted to be normal (which includes the case where V is an additive noise channel). The proofs of these parts essentially consist of two segments. The first segment uses part 1, and the second segment illustrates that pointwise domination of discrete Dirichlet forms implies pointwise domination of Dirichlet forms (as shown in (59)). This latter segment is encapsulated in Lemma 2 of Appendix C for part 2, and requires a slightly more sophisticated proof pertaining to q-ary symmetric channels in part 3.

VIII. Conclusion

In closing, we briefly reiterate our main results by delineating a possible program for proving LSIs for certain Markov chains. Given an arbitrary irreducible doubly stochastic channel $V \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$ with minimum entry $\nu = \min\{[V]_{i,j} : 1 \leq i,j \leq q\} > 0$ and $q \geq 2$, we can first use Theorem 2 to generate a q-ary symmetric channel $W_\delta \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$ with $\delta = \nu/\big(1-(q-1)\nu + \frac{\nu}{q-1}\big)$ such that $W_\delta \succeq_{\mathsf{deg}} V$. This also means that $W_\delta \succeq_{\mathsf{ln}} V$, using Proposition 3. Moreover, the δ parameter can be improved using Theorem 3 (or Propositions 10 and 11) if V is an additive noise channel. We can then use Theorem $4'$ to deduce a pointwise domination of Dirichlet forms. Since W_δ satisfies the LSIs (54) and (56) with corresponding LSI constants given in Proposition 12, Theorem $4'$ establishes the following LSIs for V:

$$D\left(f^2\mathbf{u}\,||\,\mathbf{u}\right) \leq \frac{1}{\alpha(W_\delta)}\mathcal{E}_V\left(f,f\right) \tag{65}$$

$$D\left(f^2\mathbf{u}\,||\,\mathbf{u}\right) \leq \frac{1}{\alpha(W_\delta W_\delta^*)}\mathcal{E}_{VV^*}(f,f) \tag{66}$$

for every $f \in \mathcal{L}^2(\mathcal{X},\mathbf{u})$ such that $\|f\|_{\mathbf{u}} = 1$. These inequalities can be used to derive a myriad of important facts about V. We note that the equivalent characterizations of the less noisy preorder in Theorem 1 and Proposition 8 are particularly useful for proving some of these results. Finally, we accentuate that Theorems 2 and 3 address our motivation in subsection I-D by providing analogs of the relationship between less noisy domination by q-ary erasure channels and contraction coefficients in the context of q-ary symmetric channels.

Appendix A
Basics of Majorization Theory

Since we use some majorization arguments in our analysis, we briefly introduce the notion of *group majorization* over row vectors in $\mathbb{R}^q$ (with $q \in \mathbb{N}$) in this appendix. Given a group $\mathcal{G} \subseteq \mathbb{R}^{q\times q}$ of matrices (with the operation of matrix multiplication), we may define a preorder called $\mathcal{G}$-majorization over row vectors in $\mathbb{R}^q$. For two row vectors $x, y \in \mathbb{R}^q$, we say that x $\mathcal{G}$-majorizes y if $y \in \mathsf{conv}(\{xG : G \in \mathcal{G}\})$, where $\{xG : G \in \mathcal{G}\}$ is the orbit of x under the group $\mathcal{G}$. Group majorization intuitively captures a notion of "spread" of vectors. So, x $\mathcal{G}$-majorizes y when x is more spread out than y with respect to $\mathcal{G}$. We refer readers to [9, Chapter 14, Section C] and the references therein for a thorough treatment of group majorization. If we let $\mathcal{G}$ be the symmetric group of all permutation matrices in $\mathbb{R}^{q\times q}$, then $\mathcal{G}$-majorization corresponds to traditional majorization of vectors in $\mathbb{R}^q$ as introduced in [39]. The next proposition collects some results about traditional majorization.

Proposition 13 (Majorization [9], [39])**.** *Given row vectors $x = (x_1,\dots,x_q), y = (y_1,\dots,y_q) \in \mathbb{R}^q$, let $x_{(1)} \leq \cdots \leq x_{(q)}$ and $y_{(1)} \leq \cdots \leq y_{(q)}$ denote the re-orderings of x and y in ascending order. Then, the following are equivalent:*

1) *x majorizes y, or equivalently, y resides in the convex hull of all permutations of x.*
2) *$y = xD$ for some doubly stochastic matrix $D \in \mathbb{R}_{\mathsf{sto}}^{q\times q}$.*
3) *The entries of x and y satisfy:*

$$\sum_{i=1}^{k} x_{(i)} \leq \sum_{i=1}^{k} y_{(i)}\,, \text{ for } k = 1,\dots,q-1\,,$$
$$\text{and } \sum_{i=1}^{q} x_{(i)} = \sum_{i=1}^{q} y_{(i)}\,.$$

When these conditions are true, we will write $x \succeq_{\mathsf{maj}} y$.

In the context of subsection I-C, given an Abelian group $(\mathcal{X}, \oplus)$ of order q, another useful notion of $\mathcal{G}$-majorization can be obtained by letting $\mathcal{G} = \{P_x \in \mathbb{R}^{q\times q} : x \in \mathcal{X}\}$ be the group of permutation matrices defined in (4) that is isomorphic to $(\mathcal{X}, \oplus)$. For such choice of $\mathcal{G}$, we write $x \succeq_{\mathcal{X}} y$ when x $\mathcal{G}$-majorizes (or $\mathcal{X}$-majorizes) y for any two row vectors $x, y \in \mathbb{R}^q$. We will only require one fact about such group majorization, which we present in the next proposition.

Proposition 14 (Group Majorization)**.** *Given two row vectors $x, y \in \mathbb{R}^q$, $x \succeq_{\mathcal{X}} y$ if and only if there exists $\lambda \in \mathcal{P}_q$ such that $y = x\, \mathsf{circ}_{\mathcal{X}}(\lambda)$.*

Proof. Observe that:

$$\begin{aligned} x \succeq_{\mathcal{X}} y &\Leftrightarrow y \in \mathsf{conv}\left(\{xP_z : z \in \mathcal{X}\}\right) \\ &\Leftrightarrow y = \lambda\, \mathsf{circ}_{\mathcal{X}}(x) \text{ for some } \lambda \in \mathcal{P}_q \\ &\Leftrightarrow y = x\, \mathsf{circ}_{\mathcal{X}}(\lambda) \text{ for some } \lambda \in \mathcal{P}_q \end{aligned}$$

where the second step follows from (7), and the final step follows from the commutativity of $\mathcal{X}$-circular convolution. ■

Proposition 14 parallels the equivalence between parts 1 and 2 of Proposition 13, because $\mathsf{circ}_{\mathcal{X}}(\lambda)$ is a doubly stochastic matrix for every pmf $\lambda \in \mathcal{P}_q$. In closing this appendix, we mention a well-known special case of such group majorization. When $(\mathcal{X}, \oplus)$ is the cyclic Abelian group $\mathbb{Z}/q\mathbb{Z}$ of integers with addition modulo q, $\mathcal{G} = \left\{I_q, P_q, P_q^2, \dots, P_q^{q-1}\right\}$ is the group of all cyclic permutation matrices in $\mathbb{R}^{q\times q}$, where $P_q \in \mathbb{R}^{q\times q}$ is defined in (8). The corresponding notion of $\mathcal{G}$-majorization is known as *cyclic majorization*, cf. [47].

APPENDIX B
PROOFS OF PROPOSITIONS 4 AND 12

Proof of Proposition 4.
Part 1: This is obvious from (10).
Part 2: Since the DFT matrix jointly diagonalizes all circulant matrices, it diagonalizes every W_δ for $\delta \in \mathbb{R}$ (using part 1). The corresponding eigenvalues are all real because W_δ is symmetric. To explicitly compute these eigenvalues, we refer to [19, Problem 2.2.P10]. Observe that for any row vector $x = (x_0, \dots, x_{q-1}) \in \mathbb{R}^q$, the corresponding circulant matrix satisfies:

$$\begin{aligned} \mathsf{circ}_{\mathbb{Z}/q\mathbb{Z}}(x) &= \sum_{k=0}^{q-1} x_k P_q^k = F_q \left(\sum_{k=0}^{q-1} x_k D_q^k\right) F_q^H \\ &= F_q\, \mathsf{diag}(\sqrt{q}\, x F_q)\, F_q^H \end{aligned}$$

where the first equality follows from (6) for the group $\mathbb{Z}/q\mathbb{Z}$ [19, Section 0.9.6], $D_q = \mathsf{diag}((1, \omega, \omega^2, \dots, \omega^{q-1}))$, and $P_q = F_q D_q F_q^H \in \mathbb{R}^{q\times q}$ is defined in (8). Hence, we have:

$$\begin{aligned} \lambda_j\left(W_\delta\right) &= \sum_{k=1}^{q} \left(w_\delta\right)_k \omega^{(j-1)(k-1)} \\ &= \begin{cases} 1, & j = 1 \\ 1 - \delta - \frac{\delta}{q-1}, & j = 2, \dots, q \end{cases} \end{aligned}$$

where $w_\delta = (1 - \delta, \delta/(q-1), \dots, \delta/(q-1))$.
Part 3: This is also obvious from (10)–recall that a square stochastic matrix is doubly stochastic if and only if its stationary distribution is uniform [19, Section 8.7].
Part 4: For $\delta \neq \frac{q-1}{q}$, we can verify that $W_\tau W_\delta = I_q$ when $\tau = \frac{-\delta}{1-\delta-\frac{\delta}{q-1}}$ by direct computation:

$$\begin{aligned} \left[W_\tau W_\delta\right]_{j,j} &= (1-\tau)(1-\delta) + (q-1)\left(\frac{\tau}{q-1}\right)\left(\frac{\delta}{q-1}\right) \\ &= 1\,, \text{ for } j = 1, \dots, q\,, \\ \left[W_\tau W_\delta\right]_{j,k} &= \frac{\delta(1-\tau)}{q-1} + \frac{\tau(1-\delta)}{q-1} + (q-2)\frac{\tau\delta}{(q-1)^2} \\ &= 0\,, \text{ for } j \neq k \text{ and } 1 \leq j, k \leq q\,. \end{aligned}$$

The $\delta = \frac{q-1}{q}$ case follows from (10).

Part 5: The set $\{W_\delta : \delta \in \mathbb{R}\backslash\{\frac{q-1}{q}\}\}$ is closed over matrix multiplication. Indeed, for $\epsilon, \delta \in \mathbb{R}\backslash\{\frac{q-1}{q}\}$, we can straightforwardly verify that $W_\epsilon W_\delta = W_\tau$ with $\tau = \epsilon + \delta - \epsilon\delta - \frac{\epsilon\delta}{q-1}$. Moreover, $\tau \neq \frac{q-1}{q}$ because W_τ is invertible (since W_ϵ and W_δ are invertible using part 4). The set also includes the identity matrix as $W_0 = I_q$, and multiplicative inverses (using part 4). Finally, the associativity of matrix multiplication and the commutativity of circulant matrices proves that $\{W_\delta : \delta \in \mathbb{R}\backslash\{\frac{q-1}{q}\}\}$ is an Abelian group. ■

Proof of Proposition 12.

Part 1: We first recall from [25, Appendix, Theorem A.1] that the Markov chain $\mathbf{1u} \in \mathbb{R}^{q\times q}_{\mathsf{sto}}$ with uniform stationary distribution $\pi = \mathbf{u} \in \mathcal{P}_q$ has LSI constant:

$$\alpha(\mathbf{1u}) = \inf_{\substack{f\in\mathcal{L}^2(\mathcal{X},\mathbf{u}):\\ \|f\|_{\mathbf{u}}=1\\ D(f^2\mathbf{u}||\mathbf{u})\neq 0}} \frac{\mathcal{E}_{\mathsf{std}}(f,f)}{D(f^2\mathbf{u}\,||\,\mathbf{u})} = \begin{cases} \frac{1}{2}, & q=2 \\ \frac{1-\frac{2}{q}}{\log(q-1)}, & q>2 \end{cases} .$$

Now using (18), $\alpha(W_\delta) = \frac{q\delta}{q-1}\alpha(\mathbf{1u})$, which proves part 1.

Part 2: Observe that $W_\delta W_\delta^* = W_\delta W_\delta^T = W_\delta^2 = W_{\delta'}$, where the first equality holds because W_δ has uniform stationary pmf, and $\delta' = \delta\big(2 - \frac{q\delta}{q-1}\big)$ using the proof of part 5 of Proposition 4. As a result, the discrete LSI constant $\alpha(W_\delta W_\delta^*) = \alpha(W_{\delta'})$, which we can calculate using part 1 of this proposition.

Part 3: It is well-known in the literature that $\rho_{\mathsf{max}}(\mathbf{u}, W_\delta)$ equals the second largest singular value of the divergence transition matrix $\mathsf{diag}\big(\sqrt{\mathbf{u}}\big)^{-1} W_\delta\, \mathsf{diag}\big(\sqrt{\mathbf{u}}\big) = W_\delta$ (see [36, Subsection I-B] and the references therein). Hence, from part 2 of Proposition 4, we have $\rho_{\mathsf{max}}(\mathbf{u}, W_\delta) = \big|1 - \delta - \frac{\delta}{q-1}\big|$.

Part 4: First recall the *Dobrushin contraction coefficient* (for total variation distance) for any channel $W \in \mathbb{R}^{q\times r}_{\mathsf{sto}}$:

$$\eta_{\mathsf{TV}}(W) \triangleq \sup_{\substack{P_X,Q_X\in\mathcal{P}_q:\\ P_X\neq Q_X}} \frac{\|P_X W - Q_X W\|_{\ell^1}}{\|P_X - Q_X\|_{\ell^1}} \tag{67}$$

$$= \frac{1}{2}\max_{x,x'\in[q]} \big\|W_{Y|X}(\cdot|x) - W_{Y|X}(\cdot|x')\big\|_{\ell^1} \tag{68}$$

where $\|\cdot\|_{\ell^1}$ denotes the ℓ^1-norm, and the second equality is Dobrushin's two-point characterization of η_{TV} [48]. Using this characterization, we have:

$$\eta_{\mathsf{TV}}(W_\delta) = \frac{1}{2}\max_{x,x'\in[q]} \left\|w_\delta P_q^x - w_\delta P_q^{x'}\right\|_{\ell^1} = \left|1 - \delta - \frac{\delta}{q-1}\right|$$

where w_δ is the noise pmf of W_δ for $\delta \in [0,1]$, and $P_q \in \mathbb{R}^{q\times q}$ is defined in (8). It is well-known in the literature (see e.g. the introduction of [49] and the references therein) that:

$$\rho_{\mathsf{max}}(\mathbf{u}, W_\delta)^2 \leq \eta_{\mathsf{KL}}(W_\delta) \leq \eta_{\mathsf{TV}}(W_\delta)\,. \tag{69}$$

Hence, the value of $\eta_{\mathsf{TV}}(W_\delta)$ and part 3 of this proposition establish part 4. This completes the proof. ■

Appendix C
Auxiliary Results

Proposition 15 (Properties of Domination Factor Function)**.** *Given a channel $V \in \mathbb{R}^{q\times r}_{\mathsf{sto}}$ that is strictly positive entry-wise, its domination factor function $\mu_V : \big(0, \frac{q-1}{q}\big) \rightarrow \mathbb{R}^+$ is continuous, convex, and strictly increasing. Moreover, we have $\lim_{\delta\to\frac{q-1}{q}} \mu_V(\delta) = +\infty$.*

Proof. We first prove that μ_V is finite on $\big(0, \frac{q-1}{q}\big)$. For any $P_X, Q_X \in \mathcal{P}_q$ and any $\delta \in \big(0, \frac{q-1}{q}\big)$, we have:

$$D(P_X V||Q_X V) \leq \chi^2(P_X V||Q_X V) \leq \frac{\|(P_X - Q_X)V\|_{\ell^2}^2}{\nu}$$

$$\leq \frac{\|P_X - Q_X\|_{\ell^2}^2\,\|V\|_{\mathsf{op}}^2}{\nu}$$

where the first inequality is well-known (see e.g. [36, Lemma 8]) and $\nu = \min\{[V]_{i,j} : 1 \leq i \leq q, 1 \leq j \leq r\}$, and:

$$\begin{aligned} D(P_X W_\delta || Q_X W_\delta) &\geq \frac{1}{2} \left\| (P_X - Q_X) W_\delta \right\|_{\ell^2}^2 \\ &\geq \frac{1}{2} \left\| P_X - Q_X \right\|_{\ell^2}^2 \left(1 - \delta - \frac{\delta}{q-1}\right)^2 \end{aligned}$$

where the first inequality follows from Pinsker's inequality (see e.g. [36, Proof of Lemma 6]), and the second inequality follows from part 2 of Proposition 4. Hence, we get:

$$\forall \delta \in \left(0, \frac{q-1}{q}\right), \ \ \mu_V(\delta) \leq \frac{2 \left\| V \right\|_{\mathsf{op}}^2}{\nu \left(1 - \delta - \frac{\delta}{q-1}\right)^2}. \tag{70}$$

To prove that μ_V is strictly increasing, observe that $W_{\delta'} \succeq_{\mathsf{deg}} W_\delta$ for $0 < \delta' < \delta < \frac{q-1}{q}$, because $W_\delta = W_{\delta'} W_p$ with:

$$\begin{aligned} p &= \delta - \frac{\delta'}{1 - \delta' - \frac{\delta'}{q-1}} + \frac{\delta\delta'}{1 - \delta' - \frac{\delta'}{q-1}} + \frac{\frac{\delta\delta'}{q-1}}{1 - \delta' - \frac{\delta'}{q-1}} \\ &= \frac{\delta - \delta'}{1 - \delta' - \frac{\delta'}{q-1}} \in \left(0, \frac{q-1}{q}\right) \end{aligned}$$

where we use part 4 of Proposition 4, the proof of part 5 of Proposition 4 in Appendix B, and the fact that $W_p = W_{\delta'}^{-1} W_\delta$. As a result, we have for every $P_X, Q_X \in \mathcal{P}_q$:

$$D(P_X W_\delta || Q_X W_\delta) \leq \eta_{\mathsf{KL}}(W_p)\, D(P_X W_{\delta'} || Q_X W_{\delta'})$$

using the SDPI for KL divergence, where part 4 of Proposition 12 reveals that $\eta_{\mathsf{KL}}(W_p) \in (0,1)$ since $p \in \left(0, \frac{q-1}{q}\right)$. Hence, we have for $0 < \delta' < \delta < \frac{q-1}{q}$:

$$\mu_V(\delta') \leq \eta_{\mathsf{KL}}(W_p)\, \mu_V(\delta) \tag{71}$$

using (43), and the fact that $0 < D(P_X W_{\delta'} || Q_X W_{\delta'}) < +\infty$ if and only if $0 < D(P_X W_\delta || Q_X W_\delta) < +\infty$. This implies that μ_V is strictly increasing.

We next establish that μ_V is convex and continuous. For any fixed $P_X, Q_X \in \mathcal{P}_q$ such that $P_X \neq Q_X$, consider the function $\delta \mapsto D(P_X V || Q_X V) / D(P_X W_\delta || Q_X W_\delta)$ with domain $\left(0, \frac{q-1}{q}\right)$. This function is convex, because $\delta \mapsto D(P_X W_\delta || Q_X W_\delta)$ is convex by the convexity of KL divergence, and the reciprocal of a non-negative convex function is convex. Therefore, μ_V is convex since (43) defines it as a pointwise supremum of a collection of convex functions. Furthermore, we note that μ_V is also continuous since a convex function is continuous on the interior of its domain.

Finally, observe that:

$$\begin{aligned} \liminf_{\delta \to \frac{q-1}{q}} \mu_V(\delta) &\geq \sup_{\substack{P_X, Q_X \in \mathcal{P}_q \\ P_X \neq Q_X}} \liminf_{\delta \to \frac{q-1}{q}} \frac{D(P_X V || Q_X V)}{D(P_X W_\delta || Q_X W_\delta)} \\ &= \sup_{\substack{P_X, Q_X \in \mathcal{P}_q \\ P_X \neq Q_X}} \frac{D(P_X V || Q_X V)}{\limsup_{\delta \to \frac{q-1}{q}} D(P_X W_\delta || Q_X W_\delta)} \\ &= +\infty \end{aligned}$$

where the first inequality follows from the minimax inequality and (43) (note that $0 < D(P_X W_\delta || Q_X W_\delta) < +\infty$ for $P_X \neq Q_X$ and δ close to $\frac{q-1}{q}$), and the final equality holds because $P_X W_{(q-1)/q} = \mathbf{u}$ for every $P_X \in \mathcal{P}_q$. ■

Lemma 2 (Gramian Löwner Domination implies Symmetric Part Löwner Domination). *Given $A \in \mathbb{R}_{\succeq 0}^{q \times q}$ and $B \in \mathbb{R}^{q \times q}$ that is normal, we have:*

$$A^2 = AA^T \succeq_{\mathsf{PSD}} BB^T \ \Rightarrow \ A = \frac{A + A^T}{2} \succeq_{\mathsf{PSD}} \frac{B + B^T}{2}.$$

Proof. Since $AA^T \succeq_{\mathsf{PSD}} BB^T \succeq_{\mathsf{PSD}} 0$, using the Löwner-Heinz theorem (presented in (62)) with $p = \frac{1}{2}$, we get:

$$A = \left(AA^T\right)^{\frac{1}{2}} \succeq_{\mathsf{PSD}} \left(BB^T\right)^{\frac{1}{2}} \succeq_{\mathsf{PSD}} 0$$

where the first equality holds because $A \in \mathbb{R}_{\succeq 0}^{q\times q}$. It suffices to now prove that $(BB^T)^{1/2} \succeq_{\mathsf{PSD}} \left(B + B^T\right)/2$, as the transitive property of $\succeq_{\mathsf{PSD}}$ will produce $A \succeq_{\mathsf{PSD}} \left(B + B^T\right)/2$. Since B is normal, $B = UDU^H$ by the complex spectral theorem [38, Theorem 7.9], where U is a unitary matrix and D is a complex diagonal matrix. Using this unitary diagonalization, we have:

$$U|D|U^H = \left(BB^T\right)^{\frac{1}{2}} \succeq_{\mathsf{PSD}} \frac{B + B^T}{2} = U\,\mathrm{Re}\{D\}\,U^H$$

since $|D| \succeq_{\mathsf{PSD}} \mathrm{Re}\{D\}$, where $|D|$ and $\mathrm{Re}\{D\}$ denote the element-wise magnitude and real part of D, respectively. This completes the proof. ■

ACKNOWLEDGMENT

We would like to thank an anonymous reviewer and the Associate Editor, Chandra Nair, for bringing the reference [12] to our attention. Y.P. would like to thank Dr. Ziv Goldfeld for discussions on secrecy capacity.

REFERENCES

[1] A. Makur and Y. Polyanskiy, "Less noisy domination by symmetric channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, June 25-30 2017, pp. 2463–2467.
[2] 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.
[3] Y. Polyanskiy and Y. Wu, "Strong data-processing inequalities for channels and Bayesian networks," in *Convexity and Concentration*, ser. The IMA Volumes in Mathematics and its Applications, E. Carlen, M. Madiman, and E. M. Werner, Eds., vol. 161. New York: Springer, 2017, pp. 211–249.
[4] C. E. Shannon, "A note on a partial ordering for communication channels," *Information and Control*, vol. 1, pp. 390–397, 1958.
[5] ——, "The zero error capacity of a noisy channel," *IRE Trans. Inform. Theory*, vol. 2, no. 3, pp. 706–715, September 1956.
[6] J. H. Cohen, J. H. B. Kemperman, and G. Zbăganu, *Comparisons of Stochastic Matrices with Applications in Information Theory, Statistics, Economics and Population Sciences*. Ann Arbor: Birkhäuser, 1998.
[7] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 1, pp. 2–14, January 1972.
[8] P. P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. IT-19, no. 2, pp. 197–207, March 1973.
[9] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, 2nd ed., ser. Springer Series in Statistics. New York: Springer, 2011.
[10] 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.
[11] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. New York: Cambridge University Press, 2011.
[12] M. van Dijk, "On a special class of broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. 43, no. 2, pp. 712–714, March 1997.
[13] A. El Gamal, "The capacity of a class of broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 2, pp. 166–169, March 1979.
[14] 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.
[15] 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.
[16] 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].
[17] F. Unger, "Better gates can make fault-tolerant computation impossible," *Electronic Colloquium on Computational Complexity (ECCC)*, no. 164, November 2010.
[18] M. Artin, *Algebra*, 2nd ed. New Jersey: Pearson Prentice Hall, 2011.
[19] R. A. Horn and C. R. Johnson, *Matrix Analysis*, 2nd ed. New York: Cambridge University Press, 2013.
[20] P. Diaconis, *Group Representations in Probability and Statistics*, S. S. Gupta, Ed. USA: Inst. Math. Stat. Monogr., 1988, vol. 11.
[21] Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," *IEEE Trans. Inform. Theory*, vol. 59, no. 5, pp. 2576–2595, May 2013.
[22] 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.
[23] 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.
[24] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. IT-24, no. 3, pp. 339–348, May 1978.
[25] P. Diaconis and L. Saloff-Coste, "Logarithmic Sobolev inequalities for finite Markov chains," *Ann. Appl. Probab.*, vol. 6, no. 3, pp. 695–750, 1996.
[26] R. Montenegro and P. Tetali, *Mathematical Aspects of Mixing Times in Markov Chains*, ser. Found. Trends Theor. Comput. Sci., M. Sudan, Ed. Boston-Delft: now Publishers Inc., 2006, vol. 1, no. 3.
[27] E. C. Posner, "Random coding strategies for minimum entropy," *IEEE Trans. Inform. Theory*, vol. IT-21, no. 4, pp. 388–391, July 1975.
[28] Y. Polyanskiy and Y. Wu, "Lecture notes on information theory," August 2017, Lecture Notes 6.441, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA, USA.
[29] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed., ser. International Series in Pure and Applied Mathematics. New York: McGraw-Hill, Inc., 1976.
[30] V. Anantharam, A. A. Gohari, S. Kamath, and C. Nair, "On hypercontractivity and a data processing inequality," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, June 29-July 4 2014, pp. 3022–3026.

[31] M.-D. Choi, M. B. Ruskai, and E. Seneta, "Equivalence of certain entropy contraction coefficients," *Linear Algebra Appl.*, vol. 208-209, pp. 29–36, September 1994.
[32] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York: Cambridge University Press, 2004.
[33] C. Stępniak, "Ordering of nonnegative definite matrices with application to comparison of linear models," *Linear Algebra Appl.*, vol. 70, pp. 67–71, October 1985.
[34] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, Inc., 1968.
[35] A. Rényi, "On measures of dependence," *Acta Math. Hungar.*, vol. 10, no. 3-4, pp. 441–451, 1959.
[36] A. Makur and L. Zheng, "Bounds between contraction coefficients," in *Proc. 53rd Allerton Conference*, Allerton House, UIUC, Illinois, USA, September 29-October 2 2015, pp. 1422–1429.
[37] O. V. Sarmanov, "Maximal correlation coefficient (non-symmetric case)," *Dokl. Akad. Nauk*, vol. 121, no. 1, pp. 52–55, 1958.
[38] S. Axler, *Linear Algebra Done Right*, 2nd ed., ser. Undergraduate Texts in Mathematics. New York: Springer, 2004.
[39] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, 1st ed. London: Cambridge University Press, 1934.
[40] M. Ledoux, "Concentration of measure and logarithmic Sobolev inequalities," in *Séminaire de Probabilités XXXIII*, ser. Lecture Notes in Mathematics, J. Azéma, M. Émery, M. Ledoux, and M. Yor, Eds., vol. 1709. Berlin, Heidelberg: Springer, 1999, pp. 120–216.
[41] D. Bakry, "Functional inequalities for Markov semigroups," in *Probability Measures on Groups: Recent Directions and Trends*, ser. Proceedings of the CIMPA-TIFR School, Tata Institute of Fundamental Research, Mumbai, India, 2002, S. G. Dani and P. Graczyk, Eds. New Delhi, India: Narosa Publishing House, 2006, pp. 91–147.
[42] L. Miclo, "Remarques sur l'hypercontractivité at l'évolution de éntropie pour des chaînes de Markov finies," *Séminaire de probabilités de Strasbourg*, vol. 31, pp. 136–167, 1997.
[43] K. Löwner, "Über monotone matrixfunktionen," *Math. Z.*, vol. 38, no. 1, pp. 177–216, 1934.
[44] E. Heinz, "Beiträge zur störungstheorie der spektralzerlegung," *Math. Ann.*, vol. 123, pp. 415–438, 1951.
[45] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. New York: Cambridge University Press, 1991.
[46] P. Diaconis and L. Saloff-Coste, "Nash inequalities for finite Markov chains," *J. Theoret. Probab.*, vol. 9, no. 2, pp. 459–510, 1996.
[47] A. Giovagnoli and H. P. Wynn, "Cyclic majorization and smoothing operators," *Linear Algebra Appl.*, vol. 239, pp. 215–225, May 1996.
[48] R. L. Dobrushin, "Central limit theorem for nonstationary Markov chains. I," *Theory Probab. Appl.*, vol. 1, no. 1, pp. 65–80, 1956.
[49] Y. Polyanskiy and Y. Wu, "Dissipation of information in channels with input constraints," *IEEE Trans. Inform. Theory*, vol. 62, no. 1, pp. 35–55, January 2016.