The Dispersion of Slepian-Wolf Coding

Vincent Y. F. Tan∗†, Oliver Kosut†

∗ Institute for Infocomm Research, A*STAR (Email: tanyfv@i2r.a-star.edu.sg)
† Department of Electrical and Computer Engineering, National University of Singapore
‡ Stochastic Systems Group (SSG), Massachusetts Institute of Technology, (Email: okosut@mit.edu)

Abstract—We characterize second-order coding rates (or dispersions) for distributed lossless source coding (the Slepian-Wolf problem). We introduce a fundamental quantity known as the entropy dispersion matrix, which is analogous to scalar dispersion quantities. We show that if this matrix is positive-definite, the optimal rate region under the constraint of a fixed blocklength and non-zero error probability has a curved boundary compared to being polyhedral for the Slepian-Wolf case. In addition, the entropy dispersion matrix governs the rate of convergence of the non-asymptotic region to the asymptotic one. As a by-product of our analyses, we develop a general universal achievability procedure for dispersion analysis of some other network information theory problems such as the multiple-access channel. Numerical examples show how the region given by Gaussian approximations compares to the Slepian-Wolf region.

Index Terms—Slepian-Wolf, Dispersion, Second-order Rates

I. INTRODUCTION

Distributed lossless source coding consists in separately encoding two (or more) correlated sources \( (X_1^n, X_2^n) \) into a pair of rate-limited messages \( (M_1, M_2) \). Subsequently, given these compressed versions of the sources, a decoder seeks to reconstruct \( (X_1^n, X_2^n) \). One of the most remarkable results in information theory, proved by Slepian and Wolf [1], states that the set of achievable rate pairs \( (R_1, R_2) \) is equal to that when each of the encoders is given knowledge of the other source, i.e., encoder 1 knows \( X_2^n \) and vice versa. The optimal rate region \( \mathcal{R}^* \) is the polyhedron

\[
\begin{align*}
R_1 &\geq H(X_1|X_2) \\
R_2 &\geq H(X_2|X_1) \\
R_1 + R_2 &\geq H(X_1, X_2).
\end{align*}
\]

As with most other statements in information theory, this result is asymptotic in nature. In this paper, we take a step towards non-asymptotic results by analyzing the second-order coding rates of the Slepian-Wolf (SW) problem.

An two-sender SW code is characterized by four parameters; the blocklength \( n \), the rates of the first and second sources \( (R_1, R_2) \) and the probability of error defined as

\[
P_e^{(n)} := P(\hat{X}_1^n \neq X_1^n, \hat{X}_2^n \neq X_2^n),
\]

where \( \hat{X}_1^n \) and \( \hat{X}_2^n \) are the reconstructed versions of \( X_1^n \) and \( X_2^n \) respectively. Traditionally, we require \( P_e^{(n)} \rightarrow 0 \) as \( n \rightarrow \infty \). In this paper, we fix \( n \) and require the code to be such that \( P_e^{(n)} \leq \epsilon \). We then ask what the set of achievable pairs of rates as a function of \((n, \epsilon)\) is. The main tool that we use is a multidimensional version of the Berry-Esseen theorem [2].

A. Main Contributions

This paper characterizes the \((n, \epsilon)\)-optimal rate region for the SW problem \( \mathcal{R}^*(n, \epsilon) \) up to an \( O(\log \frac{n}{\epsilon}) \) factor. In the course of doing so, we introduce a fundamental quantity called the entropy dispersion matrix of \( P_{X_1, X_2} \) and show that if this matrix is non-singular, the boundary of \( \mathcal{R}^*(n, \epsilon) \) is, unlike that of SW, a smooth curve. We also demonstrate numerically how our region compares to the SW region and to the problem of finite blocklength source coding with side information also at the encoder. While the SW problem is the focus of this paper, our achievability technique is general enough to be applicable to multi-terminal channel coding problems such as the multiple-access, broadcast and interference channels. The results for these other problems are not included this paper. The interested reader may refer to [3] for more details.

B. Related Work

The redundancy of SW coding was discussed in [4]–[6]. However, the authors considered a single source \( X_1 \) to be compressed and side information \( X_2 \) available only at the decoder. Thus, \( X_2 \) is neither coded nor estimated. They showed that a scalar dispersion quantity governs the second-order coding rate. He et al. [5] also analyzed a variable-length variant of the SW problem and showed that the dispersion is smaller than in the fixed-length setting. This dispersion is similar to that for channel coding. Sarvotham et al. [7] considered the SW problem with two sources to be compressed but limited their setting to the case the sources are symmetric. This work generalizes their setting in that we consider all discrete sources. This paper is a network information theory analogue of the works on second-order coding rates [8], [9] and finite blocklength analysis [10]–[13]. We employ the information spectrum method [14] in our converse proof. This was also done in [9].

II. PROBLEM STATEMENT AND MAIN RESULTS

A. Notation

Random variables and the values they take on will be denoted by upper case (e.g., \( X \)) and lower case (e.g., \( x \)) respectively. Types (empirical distributions) will be denoted by upper case (e.g., \( P \)) and distributions by lower case (e.g., \( p \)). For a sequence \( x^n \in X^n \), the type is denoted as \( P_{x^n} \) and conditional types are denoted similarly. The entropy...
and conditional entropy are denoted as $H(X_1) = H(p_{X_1})$ and $H(X_2|X_1) = H(p_{X_2|X_1})$. For a pair of sequences $x_1^n, x_2^n$, the notations $\hat{H}(x_1^n) := H(P_{x_1^n})$ and $\hat{H}(x_2^n|x_1^n) := H(P_{x_2^n|x_1^n})$ denote, respectively, the empirical marginal and conditional entropies. For two vectors $u, v \in \mathbb{R}^d$, the notation $u \leq v$ means $u_t \leq v_t$ for all $t = 1, \ldots, d$. We also use the notation $[2^nR] := \{1, \ldots, [2^nR]\}$.

### B. Definitions

Let $(X_1, X_2, P_{X_1X_2}(x_1, x_2))$ be a discrete memoryless multiple source (DMMS). This means that $(X_1^n, X_2^n) \sim \prod_{k=1}^n p_{X_1X_2}(x_{1k}, x_{2k})$. The alphabets $X_1, X_2$ are finite.

**Definition 1.** An $(n, 2^nR_1, 2^nR_2, \epsilon)$-SW code consists of two encoders $f_{j,n} : X_1^n \rightarrow M_j := [2^nR_j], j = 1, 2$, and a decoder $\varphi_n : M_1 \times M_2 \rightarrow X_1^n \times X_2^n$ such that the error probability in (2) (with $(X_1^n, X_2^n) := \varphi_n(f_{1,n}(X_1^n), f_{2,n}(X_2^n))$) does not exceed $\epsilon$. The rates are defined as $R_j := \frac{1}{n} \log |M_j|$.\)

**Definition 2.** A rate pair $(R_1, R_2)$ is $(n, \epsilon)$-achievable if there exists an $(n, 2^nR_1, 2^nR_2, \epsilon)$-SW code for the DMMS $P_{X_1X_2}(x_1, x_2)$. The $(n, \epsilon)$-optimal rate region $\mathcal{R}^*(n, \epsilon) \subset \mathbb{R}^2$ is the set of all $(n, \epsilon)$-achievable rate pairs.

For a positive-semidefinite symmetric matrix $V \succeq 0$, let the random vector $Z \sim N(0, V)$. Define the set

$$\mathcal{S}(V, \epsilon) := \{z \in \mathbb{R}^3 : \Pr(Z \leq z) \geq 1 - \epsilon\}. \tag{3}$$

Note that $\mathcal{S}(V, \epsilon) \subset \mathbb{R}^3$ and is analogous to the cumulative distribution function of a zero-mean Gaussian with covariance matrix $V$. If $\epsilon \leq \frac{1}{2}$, $\mathcal{S}(V, \epsilon)$ is a convex, unbounded set in the positive orthant. The boundary of $\mathcal{S}(V, \epsilon)$ is a differentiable manifold if $V$ is positive-definite ($V > 0$).

**Definition 3.** The entropy density vector is defined as

$$h(X_1, X_2) := \begin{bmatrix} -\log p_{X_1X_2}(X_1X_2) \\ -\log p_{X_1X_2}(X_1X_2) \\ -\log p_{X_1X_2}(X_1X_2) \end{bmatrix}. \tag{4}$$

The mean of the entropy density vector is $\mathbb{E}[h(X_1, X_2)] = H(P_{X_1X_2}) := [H(X_1|X_2), H(X_2|X_1), H(X_1, X_2)]^T$.

**Definition 4.** The entropy dispersion matrix $V(p_{X_1X_2})$ is the covariance of the random vector $h(X_1, X_2)$.

We abbreviate the deterministic quantities $H(p_{X_1X_2})$ and $V(p_{X_1X_2})$ as $H$ and $V$ respectively. Observe that $V$ is an analogue of the scalar dispersion quantities that have gained attention in recent years [10]–[13]. We will find it convenient to define the rate vector $R := [R_1, R_2, R_1 + R_2]^T \in \mathbb{R}^3$.

**Definition 5.** Define the region $\mathcal{R}_{in}(n, \epsilon) \subset \mathbb{R}^2$ to be the set of rate pairs $(R_1, R_2)$ that satisfy

$$R \in H + \frac{1}{\sqrt{n}} \mathcal{S}(V, \epsilon) + \frac{\nu \log n}{n} 1, \tag{5}$$

where $\nu := \|X_1\|\|X_2\| + 1$ and $1 := (1,1,1)^T$. Also let $\mathcal{R}_{out}(n, \epsilon) \subset \mathbb{R}^2$ be the set of rate pairs $(R_1, R_2)$ that satisfy

$$R \in H + \frac{1}{\sqrt{n}} \mathcal{S}(V, \epsilon) - \frac{\log n}{n} 1. \tag{6}$$

An illustration is provided in Fig. 1. Henceforth, $\epsilon \in (0, 1)$.

### C. Main Result and Interpretation

**Theorem 1.** The $(n, \epsilon)$-optimal rate region $\mathcal{R}^*(n, \epsilon)$ satisfies

$$\mathcal{R}_{in}(n, \epsilon) \subset \mathcal{R}^*(n, \epsilon) \subset \mathcal{R}_{out}(n, \epsilon). \tag{7}$$

for all $n$ sufficiently large.

This theorem is proved for $V > 0$ in Section III. Sources for which $V$ is singular include those which are (i) independent, i.e., $I(X_1; X_2) = 0$, (ii) either $X_1$ or $X_2$ is uniform over their alphabets. The authors in [7] dealt with the specific case where $X_1, X_2 \in \mathbb{F}_2$, $X_1 = \text{Bern}(\frac{1}{2})$, $X_2 = X_1 \oplus N$ with $N = \text{Bern}(q), q \in (0, \frac{1}{2})$, i.e., a discrete symmetric binary source (DSBS). In Section IV, we comment on how the proof can be adapted to derive $\mathcal{R}^*(n, \epsilon)$ for a DSBS and all $V \succeq 0$.

The direct part of Theorem 1 is proved using the usual random binning argument together with a multidimensional Berry-Esséen theorem [2]. The converse is proved using an information spectrum technique by Han [14]. Theorem 1 extends to the case where there are more than two senders.

By examining $\mathcal{R}_{in}(n, \epsilon)$ and $\mathcal{R}_{out}(n, \epsilon)$, it can be seen that we have characterized the $(n, \epsilon)$-rate region up to an $O(\log \frac{n}{\epsilon})$ factor. This residual is a consequence of (i) universal decoding for the direct part and (ii) approximations resulting from using the multidimensional Berry-Esséen theorem [2]. Observe that as $n \rightarrow \infty$, the $(n, \epsilon)$-rate regions approach the SW region $[1]$ at a rate of $O(\frac{1}{\sqrt{n}})$. This follows from the multidimensional central limit theorem. However, somewhat unexpectedly, if $V \succeq 0$, the $(n, \epsilon)$-rate region is not-polyhedral [cf. (1)]. Its boundary is a smooth curve in $\mathbb{R}^2$. This curve, given by $V$, is due to the fact that the three empirical entropies $\hat{H}(X_1^n|X_2^n)$, $\hat{H}(X_2^n|X_1^n)$ and $\hat{H}(X_1^n, X_2^n)$ have to be jointly smaller than some rate vector. By Taylor’s theorem, we see that the empirical entropy vector behaves like a multivariate Gaussian with mean $H$ and covariance $V$.

### III. PROOF OF THEOREM 1

#### A. Achievability (Inner Bound)

**Proof:** Let $(R_1, R_2)$ be a rate pair such that $R$ belongs to the inner bound $\mathcal{R}_{in}(n, \epsilon)$, defined in (5).

**Codebook Generation:** For $j = 1, 2$, randomly and independently assign an index $f_{1,n}(x^n_j) \in [2^nR_j]$ to each sequence $x^n_j \in X^n_j$ according to a uniform pmf. The sequences of the same index form a bin, i.e., $B_j(m_j) := \{x^n_j \in X^n_j : f_{1,n}(x^n_j) = m_j\}$. Note that $B_j(m_j), m_j \in [2^nR_j]$ are random sets. The bin assignments are revealed to all parties. In particular, the decoder knows the bin rates $R_j$.

**Encoding:** Given $x^n_j \in X^n_j$, encoder $j$ transmits the bin index $f_{j,n}(x^n_j)$. Hence, for length-$n$ sequence, the rates of $m_1$ and $m_2$ are $R_1$ and $R_2$ respectively.

**Decoding:** The decoder, upon receipt of the bin indices $(m_1, m_2)$ finds the unique sequence pair $(\hat{x}^{\mu}_{1}, \hat{x}^{\mu}_{2}) \in B_1(m_1) \times B_2(m_2)$ such that the empirical entropy vector

$$\hat{H}(\hat{x}^{\mu}_{1}, \hat{x}^{\mu}_{2}) := \begin{bmatrix} \hat{H}(\hat{x}^{\mu}_{1}|\hat{x}^{\mu}_{2}) \\ \hat{H}(\hat{x}^{\mu}_{2}|\hat{x}^{\mu}_{1}) \\ \hat{H}(\hat{x}^{\mu}_{1}, \hat{x}^{\mu}_{2}) \end{bmatrix} \leq R - \delta n 1, \tag{8}$$

An illustration is provided in Fig. 1. Henceforth, $\epsilon \in (0, 1)$.
where \( \delta_n := (|X_1||X_2| + \frac{1}{2}) \log(n+1) \). Define the empirical entropy typical set \( \mathcal{T}(\mathbf{R}, \delta_n) := \{z \in \mathbb{R}^2 : z \leq \mathbf{R} - \delta_n \} \). Then, (8) is equivalent to \( \mathbf{H}(x_1^n, x_2^n) \in \mathcal{T}(\mathbf{R}, \delta_n) \). If there is more than one pair or no such pair in \( B_1(m_1) \times B_2(m_2) \), declare a decoding error. Note that our decoding scheme is universal \([15]\), i.e., the decoder does not depend on knowledge of the true distribution \( p_{X_1, X_2} \).

**Analysis of error probability:** Let the sequences sent by the two users be \((X_1^n, X_2^n)\) and let their corresponding bin indices be \((M_1, M_2)\). We bound the probability of error averaged over the random code construction. Clearly, the ensemble probability of error is bounded above by the sum of the probabilities of the following four events:

\[
\mathcal{E}_1 := \{ \mathbf{H}(X_1^n, X_2^n) \notin \mathcal{T}(\mathbf{R}, \delta_n) \}
\]

\[
\mathcal{E}_2 := \{ \exists x_1^n \in B_1(M_1) \setminus \{X_1^n\} : \mathbf{H}(x_1^n, X_2^n) \in \mathcal{T}(\mathbf{R}, \delta_n) \}
\]

\[
\mathcal{E}_3 := \{ \exists x_2^n \in B_2(M_2) \setminus \{X_2^n\} : \mathbf{H}(X_1^n, x_2^n) \in \mathcal{T}(\mathbf{R}, \delta_n) \}
\]

\[
\mathcal{E}_4 := \{ \exists x_1^n, x_2^n \in B_1(M_1) \times B_2(M_2) \setminus \{X_1^n, X_2^n\} : \mathbf{H}(x_1^n, x_2^n) \in \mathcal{T}(\mathbf{R}, \delta_n) \}
\]

(9)

We bound each of these in turn. Consider

\[
P(\mathcal{E}_1) = 1 - P(\mathbf{H}(P_{X_1^n, X_2^n} \in \mathcal{T}(\mathbf{R}, \delta_n))
\]

(10)

where we made the dependence of the empirical entropy vector on the type explicit. We now bound the probability in (10). Let \( \text{vec}(p_{X_1, X_2}) \in \mathbb{R}^{|X_1||X_2|} \) be a vectorized version of the joint distribution \( p_{X_1, X_2} \). Consider the Taylor series expansion:

\[
\mathbf{H}(P_{X_1^n, X_2^n}) = \mathbf{H}(p_{X_1, X_2}) + \text{vec}(P_{X_1^n, X_2^n} - p_{X_1, X_2}) + \Delta.
\]

(11)

where the Jacobian \( J \in \mathbb{R}^{3|X_1||X_2|} \) is defined entry-wise as

\[
[J]_{t, (x_1, x_2)} = \frac{\partial g_t(p_{X_1, X_2})}{\partial p_{X_1, X_2}(x_1, x_2)}|_{p_{X_1, X_2}(x_1, x_2)},
\]

(12)

where \( g_t(p_{X_1, X_2}) := H(X_1|X_2), g_2(p_{X_1, X_2}) := H(X_2|X_1) \) and \( g_3(p_{X_1, X_2}) := H(X_1, X_2). \) Because the \( g_i \)s are twice continuously differentiable, each entry of the second order correction term \( \Delta \in \mathbb{R}^3 \) in (11) is bounded above by

\[
C \text{ vec}(P_{X_1^n, X_2^n} - p_{X_1, X_2})^2
\]

for some constant \( C > 0 \). Let \( [J]t \) be the \( t \)-th row of the matrix \( J \). Now, note that

\[
[J]t, \text{vec}(P_{X_1^n, X_2^n}) = \sum_{x_1, x_2} P_{X_1^n, X_2^n}(x_1, x_2)[J]_{t, (x_1, x_2)}
\]

\[
= \sum_{x_1, x_2} P_{X_1^n, X_2^n}(x_1, x_2)[J]_{t, (x_1, x_2, x_2)}
\]

(13)

because the joint type \( P_{X_1^n, X_2^n} \) places a probability mass \( \frac{1}{n} \) on each sample \( (X_1, X_2) \). Define the random vector \( J_k := ([J]_1, (X_1, X_2, X_2), [J]_2, (X_1, X_2, X_2), [J]_3, (X_1, X_2, X_2))^T \). On account of (10), (11) and (13), we have

\[
P(\mathcal{E}_1) \leq P \left( \mathbf{H} + \frac{1}{n} \sum_{k=1}^n (J_k - \mathbf{E}[J_k]) \leq \mathbf{R} - \delta_n \right) \]

(14)

where (a) follows from the definition \( \mathcal{T}(\mathbf{R}, \delta_n) \) and (b) follows from the probability relation

\[
P(\mathbf{W} + \Delta \leq \mathbf{R}) \geq P(\mathbf{P}(\mathbf{W} \leq \mathbf{R} - c_n 1) - P(\|\Delta\| \geq c_n).
\]

As is shown in [3], \( P(\|\Delta\| \geq c_n) \leq 1/n^2 \) if \( c_n = O(1/n) \). With this choice of \( c_n \),

\[
P(\mathcal{E}_1) \geq P \left( \frac{1}{n} \sum_{k=1}^n \mathbf{E}[J_k] \right) \leq \mathbf{R} - \delta_n
\]

\[
- P(\|\Delta\| \geq c_n). \quad (15)
\]

because \( \mathbf{R} - \mathbf{H} = \frac{\mathbf{Z}}{n} + \frac{\log n}{n} \) for some \( \mathbf{Z} \) such that \( P(\mathbf{Z} \leq \mathbf{z}) \geq 1 - \epsilon \) for \( \mathbf{Z} \sim N(0, \mathbf{V}) \) [cf. definition of \( \mathcal{T}(\mathbf{V}, \epsilon) \)]. Since \( \mathbf{V} \) is bounded, the second term \( \mathbf{H}(x_1^n, x_2^n) = 1/2 \) (the coefficient of \( \delta_n \)), we have

\[
P(\mathcal{E}_1) \geq P \left( \frac{1}{n} \sum_{k=1}^n (J_k - \mathbf{E}[J_k]) \leq \frac{\mathbf{z}}{\sqrt{n}} + \psi_n \right)
\]

(16)

where \( \psi_n = O(\log n/n) \). Note now that the bounds above are i.i.d. random vectors. These random vectors have zero mean, covariance matrix \( \mathbf{V} \succ 0 \) and finite third moment \( \mathcal{E} := E[|x_1^n, x_2^n|]^3 \) because \( x_1, x_2 \) are finite sets. Since the set integrated over in (16) is convex, by the multidimensional Berry-Esseen theorem [2] (dimension \( d = 3 \)),

\[
P(\mathcal{E}_1) \geq P(\mathbf{Z} \leq \mathbf{z} + \psi_n) - \frac{400^d}{\lambda_{\min}(\mathbf{V})^{3/2}} \frac{1}{\sqrt{n}} \quad (\geq 1 - \epsilon + O(\psi_n) - \frac{530^d}{\lambda_{\min}(\mathbf{V})^{3/2}} \frac{1}{\sqrt{n}} \quad (17)
\]

where (a) follows from Taylor’s approximation theorem. Because \( \psi_n = O(\log n/n) \) dominates the \( O(1/n^2) \) term resulting from the Berry-Esseen approximation, we conclude that

\[
P(\mathcal{E}_1) \leq e - O\left( \frac{\log n}{\sqrt{n}} \right).
\]

(18)

For the second event, by symmetry and uniformity, \( P(\mathcal{E}_2) = P(\mathcal{E}_2) \frac{X_1^n}{B_1(1)} \). Now consider the chain of inequalities:

\[
P(\mathcal{E}_2) \leq P(\mathcal{E}_2) \leq P(\mathcal{E}_2)
\]

\[
\mathbf{H}(\hat{x}_1^n, x_2^n) \in \mathcal{T}(\mathbf{R}, \delta_n) \setminus \{X_1^n, X_2^n\} = (x_1^n, x_2^n, X_1^n) \in B_1(1)
\]

(19)

\[
\leq P(\hat{x}_1^n \in B_1(1))
\]

(20)

\[
\leq P(\hat{x}_1^n \in B_1(1))
\]

(21)

\[
\leq P(\hat{x}_1^n \in B_1(1))
\]

(22)

\[
\leq P(\hat{x}_1^n \in B_1(1))
\]

(23)

\[
\leq P(\hat{x}_1^n \in B_1(1))
\]

(24)

\[
\leq P(\hat{x}_1^n \in B_1(1))
\]

(25)

\[
\leq P(\hat{x}_1^n \in B_1(1))
\]

(26)

\[
\leq P(\hat{x}_1^n \in B_1(1))
\]

(27)
for all $P$. Uniting (20) and (21) yields
\[\text{non-universal one by comparing the entropy density vector}
\]
of (4) evaluated at $x_n$ in $X_n$ is not empty (denoted as $V \in \mathcal{V}(Q_{X_2})$). In (f) we upper bounded the cardinality of the $V$-shell as $|\mathcal{V}(x_n)| \leq 2^nH(V|P_{x_n}^2)$ [15, Lem. 1.2.5]. In (g), we used the Type Counting Lemma [15, Eq. (2.5.1)]. By using the definition of $\delta_n$, (19) gives $P(\mathcal{E}_2) \leq n^{-1/2}$. Similarly $P(\mathcal{E}_3)$ is upper bounded by $n^{-1/2}$.

Combining this with (18), the error probability averaged over the random binning is $P(\mathcal{E}) \leq \epsilon$. Hence, there is a deterministic code whose error probability is no greater than $\epsilon$ if the rate pair $(R_1, R_2)$ belongs to $\mathcal{A}_n(n, \epsilon)$.

### B. Converse (Outer Bound)

**Proof:** For the outer bound, [14, Lemma 7.2.2] asserts that every $(n, 2^{nR_1}, 2^{nR_2}, P_e^{\text{SW}})$-SW code must satisfy
\[P_e^{(n)} \geq 1 - P \left[ \frac{1}{n} \text{h}(X_1^n, X_2^n) - H \leq R + \epsilon 1 - 3(2^{-n\gamma}), \right] \quad (20)
\]
for all $n$ and for any $\gamma > 0$. Recall that $\text{h}(X_1^n, X_2^n)$ is the entropy density vector in (4) evaluated at $(X_1^n, X_2^n)$. Suppose that, to the contrary, there exists a rate pair $(R_1, R_2)$ such that $R \notin \mathcal{A}_n(n, \epsilon)$ but $(R_1, R_2)$ is $(n, \epsilon)$-achievable. Then, by (6), $\mathcal{Z} := \sqrt{n}(R - H + 2^{-\gamma/2}) \notin \mathcal{V}(V, \epsilon)$. By the definition of $\mathcal{V}(V, \epsilon)$ in (3), $\mathcal{Z} \in \mathbb{R}^2$ is such that $P(Z \leq z) < 1 - \epsilon$. Now consider the probability in (20), denoted as $s_n$:
\[s_n^{(a)} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} (\text{h}(X_{1k}, X_{2k}) - H) \leq z - \left( \frac{\log n}{\sqrt{n}} - \sqrt{n}\gamma \right) 1]
\]
\[\leq P\left[ Z \leq z - \left( \frac{\log n}{\sqrt{n}} - \sqrt{n}\gamma \right) 1 \right] + \frac{530\epsilon}{\lambda_{\text{min}}(V)^{3/2}} \sqrt{n}
\]
\[\leq 1 - \epsilon - O\left( \frac{\log n}{\sqrt{n}} \right) + O\left( \frac{1}{\sqrt{n}} \right)
\]
\[\leq 1 - \epsilon - O\left( \frac{\log n}{\sqrt{n}} \right) + O\left( \frac{1}{\sqrt{n}} \right)
\]
\[\leq 1 - \epsilon - O\left( \frac{\log n}{\sqrt{n}} \right) + O\left( \frac{1}{\sqrt{n}} \right)
\]
where (a) follows from the definition of $Z$, (b) follows from the multidimensional Berry-Esseen theorem [2] and (c) follows by taking $\gamma := \log n$ and using Taylor’s approximation theorem.

Uniting (20) and (21) yields $P_e^{(n)} > \epsilon$, contradicting the $(n, \epsilon)$-achievability of $(R_1, R_2)$ for all $n$ sufficiently large.

### C. Comments on the proof

Instead of the universal decoder in (8), one could use a non-universal one by comparing the entropy density vector with the rate vector. This is likened to maximum-likelihood decoding. Taylor expansion in (11) would not be required. Under this decoding strategy, there is symmetry between the error probabilities in the direct and converse parts. Also see [14, Lem. 7.2.1-2]. The rate penalty of using a universal decoder is of the order $O\left( \frac{\log n}{n} \right)$. This is insignificant compared to the dispersion term which is of the order $O\left( \frac{1}{n} \right)$.

### IV. SINGULAR ENTROPY DISPERSION MATRICES

When $V$ is rank-deficient, consider the set $\mathcal{V}(V, \epsilon)$. Suppose for the moment that $\text{rank}(V) = 1$. This is the case considered in [7] where the source is a DSBS $(q)$. For such a DSBS, $V = v_{3 \times 3}$ for $v = \text{Var}(\log p_{X_1|X_2}(X_1|X_2)) = \text{Var}(\log p_{X_1|X_2}(X_2|X_1)) = \text{Var}(\log p_{X_1|X_2}(X_2|X_1))$. As such, all the probability mass of the degenerate Gaussian $\mathcal{N}(0, V)$ lies in a subspace of dimension one. Therefore, the set $\mathcal{V}(V, \epsilon) = \{ z \in \mathbb{R}^3 : z \geq \sqrt{n}Q^{-1}(\epsilon) 1 \}$ is axis-aligned. The quantity $\sqrt{n}Q^{-1}(\epsilon)$ is the rate redundancy [4]–[7] for fixed-length SW coding in the finite blocklength regime for a DMMs for which $\text{rank}(V) = 1$. In this case, the bounds in (5) and (6) (up to $O\left( \frac{\log n}{n} \right)$ factors) degenerate to
\[R \geq H + \frac{(1-q)(\log((1-q)/q))^2}{2}\]
where the scalar dispersion $v := q(1-q)[\log((1-q)/q)^2]$. This reduces to results in previous works [4]–[7]. Our analysis, of course, applies to all sources. Furthermore, we improve on the residual term, which is now of the order $O\left( \frac{\log n}{n} \right)$. The case where $\text{rank}(V) = 2$ follows analogously. All the probability mass of $\mathcal{N}(0, V)$ is concentrated on a two-dimensional subspace in $\mathbb{R}^3$ and the boundary of the set $\mathcal{V}(V, \epsilon)$ are not differentiable. As such only one of the “corners” of $\mathcal{V}(V, \epsilon)$ will be curved and this will be reflected in a result similar to (22). This argument can be formalized and is done in the extended version of this work [3].

### V. NUMERICAL EXAMPLES

In this section, we present examples to illustrate $\mathcal{A}_n(n, \epsilon)$. We neglect the $O\left( \frac{\log n}{n} \right)$ terms throughout; thus we are just
concerned about Gaussian approximations. The source is taken to be \( p_{X_1,X_2} = [1-3a; a; a] \) where \( a = 0.1 \). This source has a positive-definite dispersion. In Fig. 1, we plot the boundaries of the SW region [1] and the boundary of \( \mathcal{R}^*(n, \epsilon) \) for \( \epsilon = 0.01 \). We also plot the boundary of the \((n, \epsilon)\)-region for coding with side information at encoders and decoder (SI-ED). This region \( \mathcal{R}_{SI-ED}^*(n, \epsilon) \subset \mathbb{R}^2 \) is the set of \((R_1, R_2)\) satisfying
\[
R \geq \mathbf{H} + \sqrt{\frac{\text{diag}(\mathbf{V}(p_{X_1,X_2}))}{n}} Q^{-1}(\epsilon). \tag{23}
\]

From Fig. 1, we see that \( \mathcal{R}^*(n, \epsilon) \) has a curved boundary, reflecting the correlations among the entropy densities. Also, it approaches the SW boundary as \( n \) grows. The boundaries of \( \mathcal{R}^*(n, \epsilon) \) and \( \mathcal{R}_{SI-ED}^*(n, \epsilon) \) coincide if \( R_2 \) meets the condition in (23) with equality and \( R_1 \) is large (and vice versa).

There are two interesting “slices” of the plots in Fig. 1. These are the equal rate slice (along the 45° line) and the slice passing through the origin and a corner point \((R_{1,n}^*, R_{2,n}^*)\) of \( \mathcal{R}_{SI-ED}^*(n, \epsilon) \), defined as follows:
\[
R_{2,n}^* := \inf \{ R_2 : (R_1, R_2) \in \mathcal{R}_{SI-ED}^*(n, \epsilon) \text{ for some } R_1 \} \\
R_{1,n}^* := \inf \{ R_1 : (R_1, R_{2,n}^*) \in \mathcal{R}_{SI-ED}^*(n, \epsilon) \}. \tag{24}
\]

These two slices are indicated by the markers (×, ▲) in Fig. 1. The sum rates along both slices are plotted as functions of \( n \) in Figs. 2 and 3 respectively. We observe from Fig. 2 that the two sum rates on the 45° equal rate line approach each other as \( n \) grows. Moreover, empirically we observe (and can prove) that their difference decays as \( \exp(-\Theta(n)) \), which is subsumed by the \( O\left(\frac{\log n}{n}\right) \) term, i.e., the dispersions are the same. Thus, when \( n \geq 10^3 \), there is essentially no loss in performing SW coding versus cooperative encoding if we wish to optimize the sum rate. On the other hand, from Fig. 3, we see that the corresponding difference in corner points decays at a much slower rate of \( \Theta(n^{-1/2}) \). Thus, the corner rate dispersions are different and if we wish to operate at this point, SW loses second-order coding rate relative to the cooperative scenario. See [3] for further analysis of this point.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we quantified the second-order coding rates of the Slepian-Wolf problem. We showed that these rates are governed by a so-called entropy dispersion matrix. Admittedly, our results cannot be described as being finite blocklength. We seek to work towards such results in the future and to compare the accuracy of the Gaussian approximation in Theorem 1 to upper and lower bounds on the blocklength required to achieve a target error probability.

REFERENCES