Compression for Quadratic Similarity Queries

Amir Ingber, *Member, IEEE*, Thomas Courtade, *Member, IEEE*, and Tsachy Weissman, *Fellow, IEEE*

***Abstract*— The problem of performing similarity queries on compressed data is considered. We focus on the quadratic similarity measure, and study the fundamental tradeoff between compression rate, sequence length, and reliability of queries performed on the compressed data. For a Gaussian source, we show that the queries can be answered reliably if and only if the compression rate exceeds a given threshold—the *identification rate*— which we explicitly characterize. Moreover, when compression is performed at a rate greater than the identification rate, responses to queries on the compressed data can be made exponentially reliable. We give a complete characterization of this exponent, which is analogous to the error and excess-distortion exponents in channel and source coding, respectively. For a general source, we prove that, as with classical compression, the Gaussian source requires the largest compression rate among sources with a given variance. Moreover, a robust scheme is described that attains this maximal rate for any source distribution.**

***Index Terms*— Compression, search, databases, error exponent, identification rate.**

I. INTRODUCTION

FOR a database consisting of many long sequences, it is natural to perform queries of the form: *which sequences in the database are similar to a given sequence* $\mathbf{y}$? In this paper, we study the problem of compressing this database so that queries about the original data can be answered reliably given only the compressed version. This goal stands in contrast to the traditional compression paradigm, where data is compressed so that it can be reconstructed – either exactly or approximately – from its compressed form.

Specifically, for each sequence $\mathbf{x}$ in the database we only keep a short *signature*, denoted $T(\mathbf{x})$, where $T(\cdot)$ is a signature assignment function. Queries are performed using only $\mathbf{y}$ and $T(\mathbf{x})$ as input, rather than the original (uncompressed) sequence $\mathbf{x}$. This setting is illustrated in Fig. 1.

As alluded to above, we generally do not require that the original data be reproducible from the signatures.

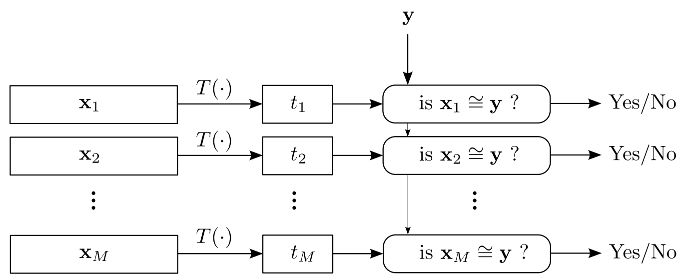

Fig. 1. Answering a query from compressed data.

Therefore the set of signatures is not meant to replace the database itself. Nevertheless, there are many instances where such compression is desirable. For example, the set of signatures can be thought of as a cached version of the original database (possibly hosted at many locations due to its relatively small size). By performing queries only on the cached (i.e., compressed) database, query latency can be reduced and the computational burden on the server hosting the uncompressed database can be lessened.

In many scenarios (e.g,. querying a criminal forensic database), query responses which are false negatives are not acceptable. A false negative occurs if a query performed on $T(\mathbf{x})$ and $\mathbf{y}$ indicates that $\mathbf{x}$ and $\mathbf{y}$ are not similar, but they are in truth. Therefore, we impose the restriction in our model that *false negatives are not permitted*. With this in mind, we regard the query responses from the compressed data as either "`no`" or "`maybe`". Since minimizing the probability that a query returns `maybe` is equivalent to minimizing the probability of returning a false positive,[1] any good compression scheme will have a corresponding query function which returns `maybe` with small probability. We note briefly that a false positive does not cause an error *per se*. Rather, it only introduces a computational burden due to the need for further verification.

In our setting we assume that the query and database sequences are independent from one another, and all entries are drawn i.i.d. according to a given distribution. The setting is closely related to the problem considered by Ahlswede et al. [1], where the focus was only on discrete sources. In [1], the authors attempt to attack the more general problem where both false positives and false negatives are allowed. In this general case, it was demonstrated in [1] that the question of 'achievable rate' is uninteresting and only the error exponent is studied (in the current paper, where false

Manuscript received July 24, 2013; revised October 17, 2014; accepted November 18, 2014. Date of publication February 11, 2015; date of current version April 17, 2015. This work was supported by the NSF Center for Science of Information under Grant CCF-0939370. This paper was presented at the 2013 Data Compression Conference.

A. Ingber was with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA. He is now with Yahoo! Labs, Sunnyvale, CA 94089 USA (e-mail: ingber@yahoo-inc.com).

T. Courtade was with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA. He is now with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94720 USA (e-mail: courtade@eecs.berkeley.edu).

T. Weissman is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: tsachy@stanford.edu).

Communicated by Y. Oohama, Associate Editor for Source Coding.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2015.2402972

[1]Complementary to false negatives, a false positive occurs if a query performed on $T(\mathbf{x})$ and $\mathbf{y}$ indicates that $\mathbf{x}$ and $\mathbf{y}$ are similar (i.e., returns `maybe`), but they are not in truth.

negatives are not allowed, we show that the rate question becomes interesting again). We should also note that the error exponent results in [1] are parameterized by an auxiliary random variable with unbounded alphabet cardinality, rendering those quantities incomputable, and therefore of limited practical use. Another closely related work is the one by Tuncel et al. [2], where the search accuracy was addressed by a reconstruction requirement with a single-letter distortion measure that is side-information dependent (and the tradeoff between compression and accuracy is that of a Wyner-Ziv [3] type). In contrast, in the current paper the search accuracy is measured directly by the accuracy of the query answers.

A different line of work attempting to identify the fundamental performance limits of database retrieval includes [4] and [5], which characterized the maximum rate of entries that can be reliably identified in a database. This line of work was extended independently in [6] and [7] allowing compression of the database, and in [8] to the case where sequence reconstruction is also required. In each of these works, the underlying assumption is that the original sequences are corrupted by noise before their enrollment in the database, the query sequence is *one of those original sequences*, and the objective is to identify which one. There are two fundamental differences between this line of work and ours. First, in our case the query sequence is random (i.e. generated by nature) and does not need to be a sequence that has already been enrolled in the database. Second, in our problem we attempt to identify sequences that are *similar* to the query sequence, rather than an exact match.

We should also note that in out setting, the query sequence is assumed to be statistically independent of the database sequences (this is the model that is primarily studied in [1]). In this case, the connection between the query and database sequences is modeled via the distributions themselves. Statistical dependence between the query and the database sequences would imply that the query sequence is related to *all* the sequences in the database, which is usually not the case. Moreover, since our formulation does not permit false negatives with probability 1, any database entries that are correlated with the query sequence and are similar will be automatically flagged for retrieval.

Other related ideas in the literature include Bloom filters [9] (with many subsequent improvements, see [10]), which are efficient data structures enabling queries without false negatives. The Bloom filter only applies for exact matches (where here we are interested in similarity queries) so it is not applicable to our problem. Nevertheless, as surveyed in [11], Bloom filters demonstrate the potential of answering queries from compressed data.

Another related notion is that of Locality Sensitive Hashing (LSH), which is a framework for data structures and algorithms for finding similar items in a given set (see [12] for a survey). LSH trades off accuracy with computational complexity and space, and false negatives are allowed. Two fundamental points are different in our approach. First, we study the information-theoretic aspect of the problem, i.e., we concentrate on space only (compression rate) and ignore computational complexity in an attempt to understand the amount of information relevant to querying that can be stored in the short signatures. Second, we do not allow false negatives, which, as discussed above, are inherent for LSH.

Other approaches for similarity search from compressed data involve dimensionality reduction techniques that preserve distances, namely those based on Johnson-Lindenstrauss-type embeddings [13] (see also *sketching*, [14]). A recent interesting application of this approach involves image retrieval for an augmented reality setting [15]. However, note that such mappings generally depend on the elements in the database; the distance preservation property cannot apply to *any* query element outside the database, making the guarantee for zero false negatives impossible without further assumptions.

This paper is organized as follows. In the next section we formally define the problem and the quantities we study (i.e., the identification rate and the identification exponent). In Section III we state and discuss our main results. Section IV provides the proofs of these results, and Section V delivers concluding remarks.

II. Problem Formulation

Throughout this paper, boldface notation $\mathbf{x}$ denotes a column vector of elements $[x_1, \ldots x_n]^T$. Capital letters denote random variables (e.g. X, Y), and $\mathbf{X}, \mathbf{Y}$ denote random vectors. Throughout the paper $\log(\cdot)$ denotes the base-2 logarithm, while $\ln(\cdot)$ is used for the usual natural logarithm.

We focus on the basic notion of quadratic similarity (sometimes called mean square error, or MSE). To this end, for any length-n real sequences $\mathbf{x}$ and $\mathbf{y}$ define

$$d(\mathbf{x}, \mathbf{y}) \triangleq \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2 = \frac{1}{n} \|\mathbf{x} - \mathbf{y}\|^2, \tag{1}$$

where $\|\cdot\|$ denotes the standard Euclidean norm. We say that $\mathbf{x}$ and $\mathbf{y}$ are *D-similar* when $d(\mathbf{x}, \mathbf{y}) \leq D$, or simply *similar* when D is clear from context.

A rate-R identification system (T, g) consists of a *signature assignment*

$$T : \mathbb{R}^n \to \{1, 2, \ldots, 2^{nR}\} \tag{2}$$

and a *query function*

$$g : \{1, 2, \ldots, 2^{nR}\} \times \mathbb{R}^n \to \{\texttt{no}, \texttt{maybe}\}. \tag{3}$$

A system (T, g) is said to be *D-admissible*, if for any $\mathbf{x}, \mathbf{y}$ satisfying $d(\mathbf{x}, \mathbf{y}) \leq D$, we have

$$g(T(\mathbf{x}), \mathbf{y}) = \texttt{maybe}. \tag{4}$$

This notion of D-admissibility motivates the use of "no" and "maybe" in describing the output of g:

- If $g(T(\mathbf{x}), \mathbf{y}) = \texttt{no}$, then $\mathbf{x}$ and $\mathbf{y}$ can not be D-similar.
- If $g(T(\mathbf{x}), \mathbf{y}) = \texttt{maybe}$, then $\mathbf{x}$ and $\mathbf{y}$ are possibly D-similar.

Stated another way, a D-admissible system (T, g) does not produce false negatives, i.e., indicate that $\mathbf{x}$ and $\mathbf{y}$ are not similar, when they are in truth. Thus, a natural figure of merit for a D-admissible system (T, g) is the frequency at which false positives occur (i.e., where $g(T(\mathbf{x}), \mathbf{y}) = \texttt{maybe}$

and $d(\mathbf{x},\mathbf{y}) > D$). To this end, let P_X and P_Y be probability distributions on $\mathbb{R}$, and assume $(\mathbf{X},\mathbf{Y}) \sim \prod_{i=1}^n P_X(x_i)P_Y(y_i)$. That is, the vectors $\mathbf{X}$ and $\mathbf{Y}$ are independent of each other and drawn i.i.d. according to P_X and P_Y respectively. Define the *false positive event*

$$\mathcal{E} = \{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}, d(\mathbf{X},\mathbf{Y}) > D\}, \tag{5}$$

and note that, for any D-admissible system (T, g), we have

$$\begin{aligned}
&\Pr\{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}\}\\
&= \Pr\{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}|d(\mathbf{X},\mathbf{Y}) \le D\}\Pr\{d(\mathbf{X},\mathbf{Y}) \le D\}\\
&\quad + \Pr\{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}, d(\mathbf{X},\mathbf{Y}) > D\} \qquad (6)\\
&= \Pr\{d(\mathbf{X},\mathbf{Y}) \le D\} + \Pr\{\mathcal{E}\}, \qquad (7)
\end{aligned}$$

where (7) follows since $\Pr\{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}|d(\mathbf{X},\mathbf{Y}) \le D\} = 1$ by D-admissibility of (T, g). Since $\Pr\{d(\mathbf{X},\mathbf{Y}) \le D\}$ does not depend on what scheme is employed, minimizing the false positive probability $\Pr\{\mathcal{E}\}$ over all D-admissible schemes (T, g) is equivalent to minimizing $\Pr\{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}\}$. Also note, that the only interesting case is when $\Pr\{d(\mathbf{X},\mathbf{Y}) \le D\} \to 0$ as n grows, since otherwise almost all the sequences in the database will be similar to the query sequence, making the problem degenerate (since almost all the database needs to be retrieved, regardless of the compression). In this case, it is easy to see that $\Pr\{\mathcal{E}\}$ vanishes if and only if the conditional probability

$$\Pr\{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}|d(\mathbf{X},\mathbf{Y}) > D\} \tag{8}$$

vanishes as well. In view of the above, we henceforth restrict our attention to the behavior of $\Pr\{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}\}$. In particular, we study the tradeoff between the rate R and $\Pr\{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}\}$.

This motivates the following definitions:

*Definition 1: For given distributions P_X, P_Y and a similarity threshold D, a rate R is said to be D-*achievable *if there exists a sequence of rate-R admissible schemes $(T^{(n)}, g^{(n)})$ satisfying*

$$\lim_{n\to\infty} \Pr\left\{g^{(n)}\left(T^{(n)}(\mathbf{X}),\mathbf{Y}\right) = \texttt{maybe}\right\} = 0. \tag{9}$$

Definition 2: For given distributions P_X, P_Y and a similarity threshold D, the identification rate $R_{\mathrm{ID}}(D, P_X, P_Y)$ *is the infimum of D-achievable rates. That is,*

$$R_{\mathrm{ID}}(D, P_X, P_Y) \triangleq \inf\{R : R \text{ is } D\text{-achievable}\}, \tag{10}$$

where an infimum over the empty set is equal to ∞.

The above definitions are in the same spirit of the rate distortion function (the rate above which a vanishing probability for excess distortion is achievable), and also in the spirit of the channel capacity (the rate below which a vanishing probability of error can be obtained). See, for example, Gallager [16].[2]

Having defined $R_{\mathrm{ID}}(D, P_X, P_Y)$, the rate at which $\Pr\{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}\}$ vanishes is also of significant interest. We expect the vanishing rate to be exponential as in the traditional source coding setting, motivating the following definition:

[2]See, for example, Cover and Thomas [17] for the alternative approach based on average distortion rather than excess distortion probability.

Definition 3: Fix $R \ge R_{\mathrm{ID}}(D, P_X, P_Y)$. The identification exponent *is defined as*

$$\mathbf{E}_{\mathrm{ID}}(R, D, P_X, P_Y) \triangleq \limsup_{n\to\infty} -\frac{1}{n}\log \inf_{g^{(n)},T^{(n)}} \Pr\left\{g^{(n)}\left(T^{(n)}(\mathbf{X}),\mathbf{Y}\right) = \texttt{maybe}\right\}, \tag{11}$$

where the infimum is over all D-admissible systems $(g^{(n)}, T^{(n)})$ of rate R and blocklength n.

The analogous quantity in source coding is the excess distortion exponent, first studied by Marton [18] for discrete sources and by Ihara and Kubo [19] for the Gaussian source (see also [20], [21] for other sources).

We pause to make a few additional remarks on the connection between $\Pr\{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}\}$ and $\Pr\{\mathcal{E}\}$, where $\mathcal{E}$ is the false positive event defined in (5). If P_X and P_Y have identical means and finite variances σ_X^2 and σ_Y^2, respectively, then the weak law of large numbers implies

$$\lim_{n\to\infty} \Pr\{d(\mathbf{X},\mathbf{Y}) \le D\} = 0 \tag{12}$$

when $D < \sigma_X^2 + \sigma_Y^2$. Thus, the relation (7) implies that vanishing $\Pr\{\mathcal{E}\}$ is attainable if and only if $R > R_{\mathrm{ID}}(D, P_X, P_Y)$ when $D < \sigma_X^2 + \sigma_Y^2$. Finally, observe that (7) implies the relationship

$$\mathbf{E}_{\mathrm{ID}}(R, D, P_X, P_Y) = \limsup_{n\to\infty} -\frac{1}{n}\log\max\left[\Pr\{d(\mathbf{X},\mathbf{Y}) \le D\}, \inf_{g^{(n)},T^{(n)}} \Pr\left\{\mathcal{E}^{(n)}\right\}\right], \tag{13}$$

where $\mathcal{E}^{(n)}$ is the false positive event defined via (5) for the system $(g^{(n)}, T^{(n)})$, and the infimum is taken over all D-admissible systems $(g^{(n)}, T^{(n)})$ of rate R and blocklength n.

III. Main Results

This section delivers our main results; all proofs are given in Section IV. The Gaussian distribution plays a prominent role in this section, therefore we use the shorthand notation $P_X = N(\mu, \sigma^2)$ to denote that P_X is the Gaussian distribution on $\mathbb{R}$ with mean μ and variance σ^2.

A. The Identification Rate for Gaussian Sources

Theorem 1: If $P_X = N(\mu, \sigma_X^2)$ and $P_Y = N(\mu, \sigma_Y^2)$, then

$$R_{\mathrm{ID}}(D, P_X, P_Y) = \begin{cases} 0 & \text{for } 0 \le D < (\sigma_X - \sigma_Y)^2 \\ \log\frac{2\sigma_X\sigma_Y}{\sigma_X^2+\sigma_Y^2-D} & \text{for } (\sigma_X - \sigma_Y)^2 \le D < \sigma_X^2 + \sigma_Y^2 \\ \infty & \text{for } D \ge \sigma_X^2 + \sigma_Y^2. \end{cases} \tag{14}$$

Before proceeding, we make a few observations about the behavior of $R_{\mathrm{ID}}(D, P_X, P_Y)$ under the assumptions of Theorem 1. First, the fact that $R_{\mathrm{ID}}(D, P_X, P_Y) = \infty$ for $D \ge \sigma_X^2 + \sigma_Y^2$ is not surprising. Indeed, if $D \ge \sigma_X^2 + \sigma_Y^2$, then $\mathbf{X}$ and $\mathbf{Y}$ are inherently D-similar. That is, $\Pr\{d(\mathbf{X},\mathbf{Y}) \le D\}$ is bounded away from zero (it actually converges to 1), and therefore (13) reveals that $\Pr\{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}\}$ can never vanish, regardless of what scheme is used.

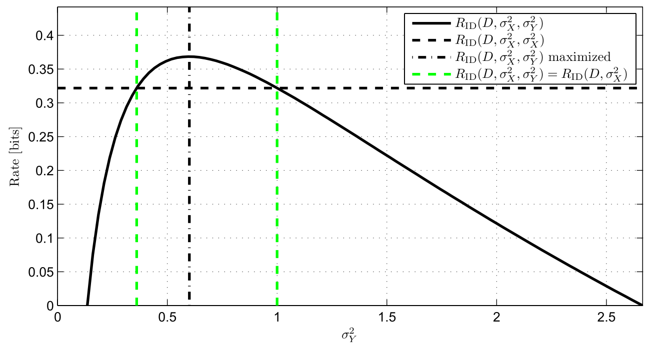

Fig. 2. The identification rate $R_{\mathrm{ID}}(D, \sigma_X^2, \sigma_Y^2) \triangleq R_{\mathrm{ID}}(D, N(\mu, \sigma_X^2), N(\mu, \sigma_Y^2))$ for different values of σ_Y^2. Here $\sigma_X^2 = 1$ and $D = 0.4$. $R_{\mathrm{ID}}(D, \sigma_X^2)$ is shorthand for $R_{\mathrm{ID}}(D, \sigma_X^2, \sigma_X^2)$.

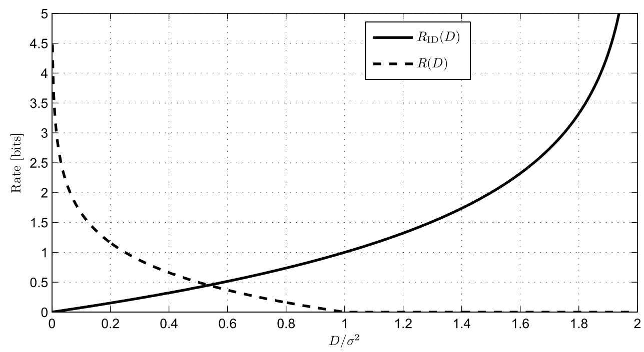

Fig. 3. The identification rate $R_{\mathrm{ID}}(D) \triangleq R_{\mathrm{ID}}(D, N(\mu, \sigma^2), N(\mu, \sigma^2))$ and the rate distortion function $R(D)$ for a Gaussian source with variance σ^2.

Second, (14) is symmetric with respect to σ_X^2 and σ_Y^2. Though this might be expected, it is not obviously true from the outset. Finally, for fixed σ_X^2 and $D < \sigma_X^2$, the function $R_{\mathrm{ID}}(D, P_X, P_Y)$ given by (14) is maximized when $\sigma_Y^2 = \sigma_X^2 - D$. In Fig. 2 we plot (14) for different values of σ_Y^2 in order to illustrate some of its properties.

As an immediate corollary to Theorem 1, we obtain the following concise result for the symmetric case of $P_X = P_Y = N(\mu, \sigma^2)$.

Corollary 1: If $P_X = P_Y = N(\mu, \sigma^2)$, then

$$R_{\mathrm{ID}}(D, P_X, P_Y) = \begin{cases} \log\left(\frac{2\sigma^2}{2\sigma^2 - D}\right) & \text{for } 0 \le D < 2\sigma^2 \\ \infty & \text{for } D \ge 2\sigma^2. \end{cases} \tag{15}$$

We remark that (15) is reminiscent of the Gaussian rate distortion function $R(D) = \left[\frac{1}{2}\log\frac{\sigma^2}{D}\right]^+$ (see [17]). The identification rate $R_{\mathrm{ID}}(D, N(\mu, \sigma^2), N(\mu, \sigma^2))$ and rate distortion function $R(D)$ for a Gaussian source are plotted in Fig. 3, and as seen in the figure, $R(D)$ is monotonically *decreasing* in D, while (15) is monotone *increasing*. This can be intuitively explained by thinking of the compression scheme as a quantizer, where all the $\mathbf{x}$ sequences mapped to the same $i \in \{1, 2, \ldots, 2^{nR}\}$ define a quantization cell. Since the scheme must answer `maybe` for all sequences $\mathbf{y}$ similar to $\mathbf{x}$, it therefore has to answer `maybe` for all $\mathbf{y}$ in the D-expansion of the quantization cell (all sequences that are at distance D from *any* point in the cell). The probability of `maybe` is, therefore, the probability that $\mathbf{Y}$ falls in the expanded cell, and this probability increases as either D grows, or as the size of the quantization cell itself grows (i.e. the rate decreases).

B. The Identification Exponent for Gaussian Sources

Having established the identification rate for Gaussian sources, we now turn our attention to the identification exponent. In order to simplify the notation for the identification exponents, we define the following functions

$$\mathbf{E}_Z(\rho) \triangleq \frac{1}{2\ln 2}(\rho - 1 - \ln\rho) \tag{16}$$

$$\wp(R, D, z_1, z_2) \triangleq -\log\sin\min\left[\frac{\pi}{2}, \left(\arcsin\left(2^{-R}\right) + \arccos\frac{z_1 + z_2 - D}{2\sqrt{z_1 z_2}}\right)\right]. \tag{17}$$

Theorem 2: Let $P_X = N(\mu, \sigma_X^2)$ and $P_Y = N(\mu, \sigma_Y^2)$. For any fixed rate $R > R_{\mathrm{ID}}(D, P_X, P_Y)$,

$$\mathbf{E}_{\mathrm{ID}}(R, D, P_X, P_Y) = \min_{\rho_X, \rho_Y} \mathbf{E}_Z(\rho_X) + \mathbf{E}_Z(\rho_Y) + \wp(R, D, \rho_X\sigma_X^2, \rho_Y\sigma_Y^2), \tag{18}$$

where the minimization is over all $\rho_X, \rho_Y > 0$ satisfying

$$\left|\sqrt{\rho_X\sigma_X^2} - \sqrt{\rho_Y\sigma_Y^2}\right| < \sqrt{D}, \quad \rho_X\sigma_X^2 + \rho_Y\sigma_Y^2 > D. \tag{19}$$

Remark 1: We note that, for $P_X = N(\mu, \sigma_X^2)$ and $P_Y = N(\mu, \sigma_Y^2)$, the exponent $\mathbf{E}_{\mathrm{ID}}(R, D, P_X, P_Y)$ is strictly positive for $R > R_{\mathrm{ID}}(D, P_X, P_Y)$, and is equal to zero at $R = R_{\mathrm{ID}}(D, P_X, P_Y)$. Therefore, the direct part of Theorem 1 is implied by Theorem 2. However, the converse part of Theorem 1 is not *implied by Theorem 2, as the latter does not exclude the possibility that the probability of* `maybe` *can be made to vanish with a sub-exponential decay rate when the exponent is equal to zero.*

In light of Theorem 2, it is instructive to revisit the relationship between false-positive and `maybe` probabilities specified in (13). To this end, consider the setting where $P_X = N(\mu, \sigma_X^2)$, $P_Y = N(\mu, \sigma_Y^2)$, and $D \le \sigma_X^2 + \sigma_Y^2$. In this case, the random variable $\frac{1}{n(\sigma_X^2 + \sigma_Y^2)}\|\mathbf{X} - \mathbf{Y}\|^2$ has a chi-squared distribution with n degrees of freedom. Therefore, it follows by Cramer's Theorem (see [22, Th. 2.2.3]) that

$$\lim_{n\to\infty} -\frac{1}{n}\log\Pr\{d(\mathbf{X}, \mathbf{Y}) \le D\} = \mathbf{E}_Z\left(\frac{D}{\sigma_X^2 + \sigma_Y^2}\right). \tag{20}$$

In this setting, it is a straightforward algebraic exercise to see that

$$\mathbf{E}_{\mathrm{ID}}(R, D, P_X, P_Y) < \mathbf{E}_Z\left(\frac{D}{\sigma_X^2 + \sigma_Y^2}\right) \tag{21}$$

for $R < \infty$ by putting

$$\rho_X = \frac{\sigma_X^2 D + \sigma_Y^2(\sigma_X^2 + \sigma_Y^2)}{(\sigma_X^2 + \sigma_Y^2)^2},$$

$$\rho_Y = \frac{\sigma_Y^2 D + \sigma_X^2(\sigma_X^2 + \sigma_Y^2)}{(\sigma_X^2 + \sigma_Y^2)^2} \tag{22}$$

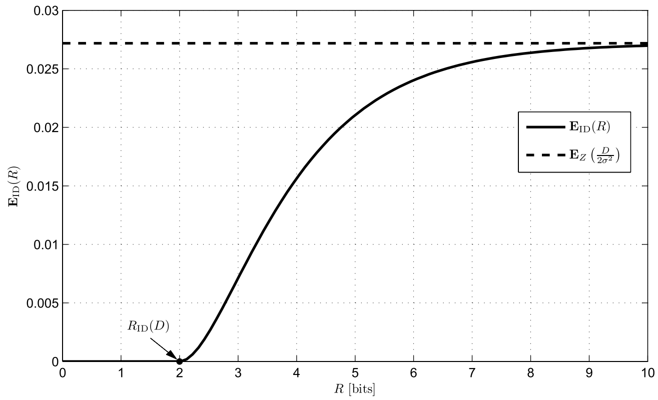

Fig. 4. Plot of $\mathbf{E}_{\rm ID}(R) \triangleq \mathbf{E}_{\rm ID}(R, D, N(\mu, \sigma^2), N(\mu, \sigma^2))$ for $D/\sigma^2 = 1.5$. In this case, $R_{\rm ID}(D, N(\mu, \sigma^2), N(\mu, \sigma^2)) = 2$ bits per symbol.

in (18). Therefore, $\mathbf{E}_{\rm ID}(R, D, P_X, P_Y)$ also precisely characterizes the best-possible exponent corresponding to the probability of a false positive event in this setting due to the relation (13).

In the case where $P_X = P_Y = N(\mu, \sigma^2)$, the symmetry in (18) can be exploited to yield the following corollary.

Corollary 2: Let $P_X = P_Y = N(\mu, \sigma^2)$. For any fixed rate $R > R_{\rm ID}(D, P_X, P_Y)$,

$$\mathbf{E}_{\rm ID}(R, D, P_X, P_Y) = \min_{\rho}\ 2\mathbf{E}_Z(\rho) + \wp(R, D, \rho\sigma^2, \rho\sigma^2), \tag{23}$$

where the minimization is over all ρ satisfying

$$2\sigma^2 \geq 2\rho\sigma^2 \geq D. \tag{24}$$

A formal proof is given in Section IV. The identification exponent (23) for the case of $D/\sigma^2 = 1.5$ is illustrated in Fig. 4.

Before proceeding, we briefly note that the identification exponent $\mathbf{E}_{\rm ID}(R, D, P_X, P_Y)$ can sometimes be strictly positive at $R \to 0$.[3] For instance, if

$$\left| \frac{1}{\sqrt{n}}\mathbb{E}\|\mathbf{X}\| - \frac{1}{\sqrt{n}}\mathbb{E}\|\mathbf{Y}\| \right| > \sqrt{D} + \epsilon \tag{25}$$

for some $\epsilon > 0$, then the signature $T(\mathbf{X})$ can simply indicate whether or not $\left| \frac{1}{\sqrt{n}}\|\mathbf{X}\| - \frac{1}{\sqrt{n}}\mathbb{E}\|\mathbf{X}\| \right| > \epsilon/2$ holds, requiring rate $R = 1/n$ (a single bit per sequence). Then, the query function g returns `maybe` only if

$$\left| \frac{1}{\sqrt{n}}\|\mathbf{X}\| - \frac{1}{\sqrt{n}}\mathbb{E}\|\mathbf{X}\| \right| > \epsilon/2, \text{ or} \tag{26}$$

$$\left| \frac{1}{\sqrt{n}}\|\mathbf{Y}\| - \frac{1}{\sqrt{n}}\mathbb{E}\|\mathbf{Y}\| \right| > \epsilon/2. \tag{27}$$

If neither (26) nor (27) occur, then it is readily verified that $d(\mathbf{X}, \mathbf{Y}) > D$ using the triangle inequality. Whenever the random variables X^2 and Y^2 satisfy a large deviations principle (as in the Gaussian case, and for many other distributions, see [22]), we see that g returns `maybe` with probability exponentially decaying in n, and we can conclude that $\lim_{R\to 0^+} \mathbf{E}_{\rm ID}(R, D, P_X, P_Y) > 0$. If this is indeed the case, then it also follows that $R_{\rm ID}(D, P_X, P_Y) = 0$ by definition. Though this discussion applies for arbitrary distributions P_X, P_Y, this latter point is concretely reflected in Theorems 1 and 2 for the case where $D \leq (\sigma_X - \sigma_Y)^2$.

C. Upper Bounds on the Identification Rate

In the previous two subsections, we focused our attention primarily to the case where P_X and P_Y were Gaussian distributions. In the sequel, we consider more general distributions and show that Gaussian P_X, P_Y constitute an extremal case in terms of the identification rate.

Theorem 3: Suppose P_X and P_Y are distributions with finite second moments σ_X^2 and σ_Y^2, respectively. Then

$$R_{\rm ID}(D, P_X, P_Y) \leq \overline{R}_{\rm ID}(D, P_X, P_Y) \triangleq \inf_{P_{\hat{X}|X}} I(X; \hat{X}), \tag{28}$$

where the infimum is taken over all conditional distributions $P_{\hat{X}|X}$ satisfying

$$\sqrt{\mathbb{E}\left[\left(\sqrt{\frac{\sigma_X}{\sigma_Y}}Y - \hat{X}\right)^2\right]} \geq \sqrt{\mathbb{E}\left[\left(\sqrt{\frac{\sigma_Y}{\sigma_X}}X - \hat{X}\right)^2\right]} + \sqrt{D - (\sigma_X - \sigma_Y)^2} \tag{29}$$

for $(Y, X, \hat{X}) \sim P_Y(y)P_X(x)P_{\hat{X}|X}(\hat{x}|x)$. Moreover,

$$\mathbf{E}_{\rm ID}(R, D, P_X, P_Y) > 0 \tag{30}$$

for any $R > \overline{R}_{\rm ID}(D, P_X, P_Y)$.

Remark 2: Note that Theorem 3 does not require P_X and P_Y to have identical means.

Remark 3: Also note, that the achievability result and the proof technique carry over to general distortion criteria satisfying the triangle inequality. We omit the details as the focus of this paper is on the quadratic similarity criterion.

For general source distributions P_X, P_Y, we lack a matching lower bound on $R_{\rm ID}(D, P_X, P_Y)$. However, such a converse was proved in the Gaussian setting (see Theorem 1). The key ingredient in the proof of Theorem 1 is the isoperimetric inequality on the surface of a hypersphere – the set on which the probability of a high dimensional Gaussian random vector concentrates (see Section IV for details). In general, precise isoperimetric inequalities are unknown and therefore establishing a similar general converse appears to be extremely difficult.[4]

In spite of this, an application of Theorem 3 reveals the interesting fact that Gaussian P_X and P_Y correspond to sources which are "most difficult" to compress for queries. This is analogous to the setting of classical lossy compression, where the Gaussian source requires the maximum rate for compression subject to a quadratic distortion constraint. Formally,

Theorem 4: Suppose P_X and P_Y have identical means and finite variances σ_X^2 and σ_Y^2, respectively. Then

$$R_{\rm ID}(D, P_X, P_Y) \leq R_{\rm ID}(D, N(0, \sigma_X^2), N(0, \sigma_Y^2)). \tag{31}$$

[3] Note that whenever R is *equal* to zero, the probability of `maybe` is equal to 1 (unless the supports of P_X and P_Y are disjoint in a way making *any* two sequences $\mathbf{x}$ and $\mathbf{y}$ dissimilar, making the problem degenerate).

[4] For discrete memoryless sources and general distortion measure, see the efforts made by Ahlswede et al. [1] and also recently in [23].

In particular, Gaussian P_X and P_Y demand the largest identification rate for given variances.

D. Robust Identification Schemes

In addition to the extremal property of Gaussian sources described in Theorem 4, there exists a sequence of rate-R identification schemes $\{T^{(n)}, g^{(n)}\}_{n\to\infty}$, where $(T^{(n)}, g^{(n)})$ denotes a blocklength-n identification scheme, designed for Gaussian sources which are *robust* in the following sense. Using the construction described in the achievability proof of Theorem 1, we can construct a sequence of D-admissible, rate-R schemes $\{T^{(n)}, g^{(n)}\}_{n\to\infty}$ which satisfy

$$\lim_{n\to\infty} \Pr\left\{g^{(n)}\left(T^{(n)}(\mathbf{X}), \mathbf{Y}\right) = \texttt{maybe}\right\} = 0 \tag{32}$$

when $\mathbf{X}$, $\mathbf{Y} \sim \prod_{i=1}^n P_X(x_i)P_Y(y_i)$, $P_X = N(0,\sigma_X^2)$, $P_Y = N(0,\sigma_Y^2)$ and

$$R > R_{\text{ID}}\left(D, N(0,\sigma_X^2), N(0,\sigma_Y^2)\right). \tag{33}$$

It turns out that this particular sequence $\{T^{(n)}, g^{(n)}\}_{n\to\infty}$ is robust to the source distributions in the sense that we also have

$$\lim_{n\to\infty} \Pr\left\{g^{(n)}\left(T^{(n)}(\tilde{\mathbf{X}}), \tilde{\mathbf{Y}}\right) = \texttt{maybe}\right\} = 0 \tag{34}$$

when $\tilde{\mathbf{X}}, \tilde{\mathbf{Y}} \sim \prod_{i=1}^n P_{\tilde{X}}(\tilde{x}_i)P_{\tilde{Y}}(\tilde{y}_i)$, and $P_{\tilde{X}}$, $P_{\tilde{Y}}$ are zero-mean distributions with variances σ_X^2 and σ_Y^2, respectively. Moreover, the sequence $\{T^{(n)}, g^{(n)}\}_{n\to\infty}$ continues to be D-admissible for the sources $\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}$. Thus, roughly speaking, a scheme (T, g) which is "good" for Gaussian sources $\mathbf{X}, \mathbf{Y}$ can be expected to perform well for arbitrary sources $\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}$, provided the respective variances match their Gaussian counterparts and the blocklength n is large. The proof of this robustness property is given in Section IV-F.

IV. Proofs

In this section, we prove each of the main results. Proofs are organized by subsection. We begin with a primer on the key geometric ideas that are used throughout the proofs.

A. Geometric Preliminaries

For the proofs we require the following definitions related to n-dimensional Euclidean geometry.

For $r > 0, \mathbf{u} \in \mathbb{R}^n$, let $\text{BALL}_r(\mathbf{u}) \subseteq \mathbb{R}^n$ denote the ball with radius r centered at $\mathbf{u}$:

$$\text{BALL}_r(\mathbf{u}) \triangleq \left\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{u}\| \le r\right\}. \tag{35}$$

$\text{BALL}_r(\mathbf{0})$ will be denoted BALL_r.

Denote by $S_r \subseteq \mathbb{R}^n$ the spherical shell with radius r centered at the origin:

$$S_r \triangleq \left\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = r\right\}. \tag{36}$$

For any two vectors $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, the angle between them shall be denoted by

$$\angle(\mathbf{x}_1, \mathbf{x}_2) \triangleq \arccos\left(\frac{\mathbf{x}_1^T\mathbf{x}_2}{\|\mathbf{x}_1\|\|\mathbf{x}_2\|}\right) \in [0, \pi]. \tag{37}$$

For $\theta \in [0, \pi]$ and a point $\mathbf{u} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, define the cone with half angle θ and axis going through $\mathbf{u}$:

$$\text{CONE}(\mathbf{u}, \theta) \triangleq \left\{\mathbf{x} \in \mathbb{R}^n : \angle(\mathbf{u}, \mathbf{x}) \le \theta\right\}. \tag{38}$$

Note that $\text{CONE}(\mathbf{u}, 0)$ is the half-infinite line $\{\alpha\mathbf{u} : \alpha > 0\}$, that $\text{CONE}(\mathbf{u}, \pi/2)$ is the half-space containing $\mathbf{u}$ that is bordered by the hyperplane orthogonal to $\mathbf{u}$ which passes through the origin, and that $\text{CONE}(\mathbf{u}, \pi)$ is the entire space $\mathbb{R}^n$. Also, note that $\text{CONE}(\mathbf{u}_1, \theta) = \text{CONE}(\mathbf{u}_2, \theta)$ for any $\mathbf{u}_1 = \lambda\mathbf{u}_2$, $\lambda > 0$.

For $r > 0$, $\mathbf{u} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ and $\theta \in [0, \pi]$, denote by $\text{CAP}_r(\mathbf{u}, \theta)$ the spherical cap:

$$\text{CAP}_r(\mathbf{u}, \theta) \triangleq S_r \cap \text{CONE}(\mathbf{u}, \theta). \tag{39}$$

Let $\boldsymbol{\Omega}(\theta)$ denote the fraction of the (hyper-)surface area of S_r that is occupied by $\text{CAP}_r(\mathbf{u}, \theta)$:

$$\boldsymbol{\Omega}(\theta) \triangleq \frac{|\text{CAP}_r(\mathbf{u}, \theta)|}{|S_r|}. \tag{40}$$

Note that the value of $\boldsymbol{\Omega}(\theta)$ depends neither on r nor on $\mathbf{u}$. Also note that $\boldsymbol{\Omega}(\theta) = 1$ for all $\theta \ge \pi$. The following bounds on $\boldsymbol{\Omega}(\theta)$ will be useful:

Lemma 1 [24, Cor. 3.2]: For $0 < \theta < \arccos(1/\sqrt{n}) < \frac{\pi}{2}$, *we have*

$$\boldsymbol{\Omega}(\theta) < \frac{1}{\sqrt{2\pi(n-1)}} \cdot \frac{1}{\cos\theta} \cdot \sin^{n-1}\theta, \tag{41}$$

$$\boldsymbol{\Omega}(\theta) > \frac{1}{3\sqrt{2\pi n}} \cdot \frac{1}{\cos\theta} \cdot \sin^{n-1}\theta. \tag{42}$$

For positive $r_1 \le r_2 \in \mathbb{R}$, let $S_{r_1,r_2} \subseteq \mathbb{R}^n$ be a spherical shell of inner radius r_1 and outer radius r_2:

$$S_{r_1,r_2} \triangleq \left\{\mathbf{x} \in \mathbb{R}^n : r_1 \le \|\mathbf{x}\| \le r_2\right\}. \tag{43}$$

For a given half-angle $\theta \in [0, \pi]$, define the (r_1, r_2)-spherical cap with half-angle θ and axis going through $\mathbf{u}$ as

$$\text{CAP}_{r_1,r_2}(\mathbf{u}, \theta) \triangleq \text{CONE}(\mathbf{u}, \theta) \cap S_{r_1,r_2}. \tag{44}$$

For a set $A \subseteq \mathbb{R}^n$ and $D > 0$, the D-expansion of A, denoted $\Gamma^D(A)$ is defined as

$$\Gamma^D(A) \triangleq \{\mathbf{y} \in \mathbb{R}^n : \exists_{\mathbf{x}\in A} d(\mathbf{x}, \mathbf{y}) \le D\} \tag{45}$$

$$= A + \text{BALL}_{\sqrt{nD}}, \tag{46}$$

where we have used $+$ to denote the Minkowski sum.

We will require the following lemma, which says that among all sets on the surface of a sphere with equal surface area, the spherical cap has the smallest expanded set. This is known as the isoperimetric inequality on the surface of a sphere, and also as Levy's lemma (see, [25, Th. 1.1]).

Lemma 2 (Levy's Lemma): Let $A \subseteq S_r$, where S_r is the surface of a sphere of radius $r > 0$. Let $\mu(A)$ be the surface area occupied by A, and let $C \subseteq S_r$ be a spherical cap with surface area equal to $\mu(A)$. Then,

$$\mu(S_r \cap \Gamma^D(C)) \le \mu(S_r \cap \Gamma^D(A)).$$

We note that the lemma holds also in the case where the distance is measured by the arc length (or, equivalently, the angle between two points on the sphere).

B. Codes That Cover a Spherical Shell

Definition 4: Let $S_r \subseteq \mathbb{R}^n$ be the spherical shell with radius r. We say that a set of points $\mathcal{C} = \{\mathbf{u}_1, \ldots, \mathbf{u}_m : \mathbf{u}_i \in \mathbb{R}^n\}$ is a code that D-covers S_r if

$$S_r \subseteq \bigcup_{\mathbf{u}\in\mathcal{C}} \mathrm{BALL}_{\sqrt{nD}}(\mathbf{u}). \tag{47}$$

The rate *of $\mathcal{C}$ is defined as $\frac{1}{n}\log m$.*

When not explicitly stated, the ambient dimension n of the code $\mathcal{C}$ will be clear from context.

Lemma 3 (Following [26]): Fix $\sigma^2 > 0$ and the dimension n. For any $0 < D_0 < \sigma^2$, there exists a code $\mathcal{C}$ that D_0-covers $S_{\sqrt{n\sigma^2}}$ with rate

$$R_0 = \frac{1}{n}\log|\mathcal{C}| \le \frac{1}{2}\log\frac{\sigma^2}{D_0} + O\left(\frac{\log n}{n}\right). \tag{48}$$

Moreover, for all $\mathbf{u}\in\mathcal{C}$, we have $\|\mathbf{u}\| = \sqrt{n(\sigma^2 - D_0)}$, and

$$\mathrm{CAP}_{\sqrt{n\sigma^2}}(\mathbf{u}, \theta_0) = S_{\sqrt{n\sigma^2}} \cap \mathrm{BALL}_{\sqrt{nD_0}}(\mathbf{u}), \tag{49}$$

where

$$\theta_0 \triangleq \arcsin(\sqrt{D_0/\sigma^2}) < \frac{\pi}{2}. \tag{50}$$

Proof: Appendix A. □

It is no surprise that the term $\frac{1}{2}\log\frac{\sigma^2}{D_0}$ appearing in (48) is identical to the rate-distortion function for the Gaussian source with variance σ^2 evaluated at distortion-level D_0. We could have therefore used any standard (random code-like) construction. However, using Lemma 3 will be more convenient for our purposes since each point in S_r is guaranteed to be covered, and hence we do not need to account for another error event. This fact will make the subsequent proofs more straightforward.

We briefly note that the idea of covering the shell of a hypersphere can be thought of as the Gaussian counterpart of the type-covering lemma for discrete sources [27]. A similar usage of "Gaussian types" of this sort can also be found in [28, Sec. V-A].

C. Identification Rate

The proof of Theorem 1 is somewhat lengthy, so we first give the key ideas here before moving onto the formal details.

The proof of the theorem relies on the fact that a high-dimensional Gaussian random vector – with independent entries having zero mean and variance σ_X^2 – concentrates near a thin hyper-spherical shell of radius $r_0 \triangleq \sqrt{n\sigma_X^2}$, which we call the *typical sphere*. The signature assignment constructed in the direct part of the proof quantizes the surface of the typical sphere into regions roughly described by spherical caps. The query function g, knowing which cap $\mathbf{X}$ lies in from the received signature, returns `maybe` only if $\mathbf{Y}$ lies within Euclidean distance $\sqrt{nD}$ of the cap in which $\mathbf{X}$ lies. Thus, the goal in the direct part is to show that, for sufficiently large rate R, the probability that $\mathbf{Y}$ falls into the Γ^D-expansion of any given cap is vanishing.

The key ingredient in proving the converse is the isoperimetric inequality on the surface of the hypersphere, known as Levy's lemma (see Lemma 2 above). In a nutshell, we apply Levy's lemma to prove that any given identification system (T, g) requires a rate that is essentially as large as an identification system that uniquely assigns caps on the typical sphere to signatures (as is done by the achievability scheme). The apparent need for a refined isoperimetric inequality to prove the converse distinguishes our problem from the class of standard rate-distortion problems.

Proof of Theorem 1: Before beginning the proof, we first note that it is sufficient to consider D in the interval $(\sigma_X - \sigma_Y)^2 < D < \sigma_X^2 + \sigma_Y^2$. The claims that $R_{\mathrm{ID}}(D, P_X, P_Y) = 0$ for $D \le (\sigma_X - \sigma_Y)^2$, and $R_{\mathrm{ID}}(D, P_X, P_Y) = \infty$ for $D \ge \sigma_X^2 + \sigma_Y^2$ then follow from monotonicity of $R_{\mathrm{ID}}(D, P_X, P_Y)$ in D.

Direct Part: Fix a small $\epsilon > 0$, and define $r_X \triangleq \sqrt{n\sigma_X^2}$ (i.e., the radius of the typical sphere). Let D be a desired similarity threshold in the interval $(\sigma_X - \sigma_Y)^2 < D < \sigma_X^2 + \sigma_Y^2$, and let $\eta > 0$ be sufficiently small so that

$$(1-\epsilon)\left[\frac{\sigma_X^2 + \sigma_Y^2 - D}{2\sigma_X\sigma_Y}\right]^2 < \left[\frac{\sigma_X^2 + \sigma_Y^2 - 2\eta - D}{2\sqrt{(\sigma_X^2+\eta)(\sigma_Y^2+\eta)}}\right]^2. \tag{51}$$

Next, define a constant $D_0 > 0$ satisfying

$$\begin{aligned}(1-\epsilon)\sigma_X^2 &\left[\frac{\sigma_X^2 + \sigma_Y^2 - 2\eta - D}{2\sqrt{(\sigma_X^2+\eta)(\sigma_Y^2+\eta)}}\right]^2 \\ &< D_0 \\ &< \sigma_X^2\left[\frac{\sigma_X^2 + \sigma_Y^2 - 2\eta - D}{2\sqrt{(\sigma_X^2+\eta)(\sigma_Y^2+\eta)}}\right]^2.\end{aligned} \tag{52}$$

The motivation behind the choices of η and D_0 satisfying (51) and (52) will become clear as the proof proceeds.

By our assumption that $D > (\sigma_X - \sigma_Y)^2$, it follows that $0 < D_0 < \sigma_X^2$. By Lemma 3, there exists a code $\mathcal{C}$ which D_0-covers S_{r_X} with rate R_0 bounded by

$$R_0 \le \frac{1}{2}\log\frac{\sigma_X^2}{D_0} + O\left(\frac{\log n}{n}\right). \tag{53}$$

Let $T_0 : S_{r_X} \to \mathcal{C}$ be the quantization operation defined by

$$T_0(\mathbf{x}) = \arg\min_{\mathbf{u}\in\mathcal{C}} \|\mathbf{x} - \mathbf{u}\| \quad \text{for } \mathbf{x} \in S_{r_X}. \tag{54}$$

That is, the function $T_0(\mathbf{x})$ maps $\mathbf{x} \in S_{r_X}$ to the closest reconstruction point $\mathbf{u} \in \mathcal{C}$. Since $\mathcal{C}$ is a code that D_0-covers S_{r_X}, it follows that

$$\|T_0(\mathbf{x}) - \mathbf{x}\| \le \rho_0 \triangleq \sqrt{nD_0} \quad \text{for all } \mathbf{x} \in S_{r_X}. \tag{55}$$

Denote the points in S_{r_X} that are mapped to $\mathbf{u}$ by $T_0^{-1}(\mathbf{u})$. With this notation, it follows by construction that

$$T_0^{-1}(\mathbf{u}) \subseteq \mathrm{CAP}_{r_X}(\mathbf{u}, \theta_0), \tag{56}$$

where $\theta_0 \triangleq \arcsin(\sqrt{D_0/\sigma_X^2})$ courtesy of Lemma 3. The set $\mathrm{CAP}_{r_X}(\mathbf{u}, \theta_0)$ is illustrated in Fig. 5.

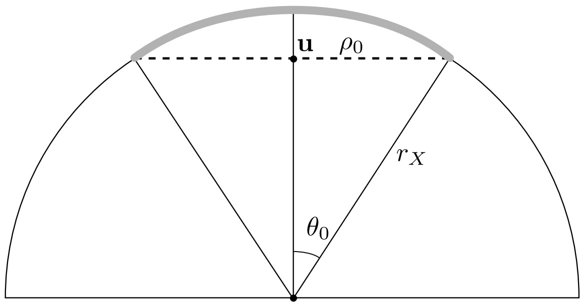

Fig. 5. Illustration of a single cap $\mathrm{CAP}_{r_X}(\mathbf{u}, \theta_0)$ (denoted in grey).

Define S_X^{typ} to be the set of all vectors $\mathbf{x} \in \mathbb{R}^n$ s.t. $\sigma_X^2 - \eta \leq \frac{1}{n}\|\mathbf{x}\|^2 \leq \sigma_X^2 + \eta$. In other words,

$$S_X^{\mathrm{typ}} \triangleq S_{r^-,r^+}, \tag{57}$$

where $r^{\pm} \triangleq \sqrt{n(\sigma_X^2 \pm \eta)}$. Note that $\Pr\{\mathbf{X} \notin S_X^{\mathrm{typ}}\}$ vanishes[5] with n, which motivates the notation S_X^{typ}.

Next, we construct a mapping $T : S_X^{\mathrm{typ}} \to \mathcal{C}$ defined as follows:

$$T(\mathbf{x}) = T_0\left(\mathbf{x} \cdot \frac{\sqrt{n\sigma_X^2}}{\|\mathbf{x}\|}\right). \tag{58}$$

Since $T_0^{-1}(\mathbf{u})$ is contained in $\mathrm{CAP}_{r_X}(\mathbf{u}, \theta_0)$, we similarly have that the inverse map T^{-1} satisfies

$$T^{-1}(\mathbf{u}) \subseteq \mathrm{CAP}_{r^-,r^+}(\mathbf{u}, \theta_0). \tag{59}$$

The signature assignment for our identification scheme for $\mathbf{x} \in S_X^{\mathrm{typ}}$ shall be given by the function $T(\cdot)$ defined above. For $\mathbf{x} \notin S_X^{\mathrm{typ}}$ we define $T(\mathbf{x}) = \mathtt{e}$, where $\mathtt{e}$ is an additional "erasure" symbol, denoting the fact that the signature does not convey any information about $\mathbf{x}$ in this case (and the decision function $g(\cdot,\cdot)$ must output maybe). Note that the additional rate incurred by the erasure symbol is negligible and we still have that the signature assignment's rate R is bounded by

$$\begin{aligned} R &= \frac{1}{n}\log(|\mathcal{C}| + 1) && (60)\\ &\leq \frac{1}{2}\log\frac{\sigma_X^2}{D_0} + O\left(\frac{\log n}{n}\right) && (61)\\ &\leq \log\frac{2\sigma_X\sigma_Y}{\sigma_X^2 + \sigma_Y^2 - D} + \log\frac{1}{1-\epsilon} + O\left(\frac{\log n}{n}\right), && (62) \end{aligned}$$

where the final inequality follows from (51) and (52).

The query function $g(\cdot,\cdot)$ is defined to be the optimal one given the signature mapping $T(\cdot)$:

$$g(t, \mathbf{y}) = \begin{cases} \mathtt{maybe} & \text{If } t = \mathtt{e}, \\ & \text{or if } \exists \mathbf{x}' \in T^{-1}(t) \text{ s.t. } d(\mathbf{x}', \mathbf{y}) \leq D \\ \mathtt{no} & \text{otherwise.} \end{cases} \tag{63}$$

Using the shorthand notation

$$\Pr\{\mathtt{maybe}\} \triangleq \Pr\{g(T(\mathbf{X}), \mathbf{Y}) = \mathtt{maybe}\}, \tag{64}$$

we analyze $\Pr\{\mathtt{maybe}\}$ as follows. First, define a typical set for the $\mathbf{Y}$-sequences:

$$S_Y^{\mathrm{typ}} \triangleq S_{r_Y^-, r_Y^+}, \tag{65}$$

where $r_Y^{\pm} \triangleq \sqrt{n(\sigma_Y^2 \pm \eta)}$, and write

$$\begin{aligned} \Pr\{\mathtt{maybe}\} \leq \Pr\{\mathtt{maybe}|\mathbf{X} \in S_X^{\mathrm{typ}}, \mathbf{Y} \in S_Y^{\mathrm{typ}}\} \\ + \Pr\{\mathbf{X} \notin S_X^{\mathrm{typ}}\} + \Pr\{\mathbf{Y} \notin S_Y^{\mathrm{typ}}\}. \end{aligned} \tag{66}$$

Note that the latter two terms in (66) vanish as n grows large, thus we focus on bounding the first term. To this end, we require the following lemma.

Lemma 4: Let $\mathcal{C}$ and η be as defined above. For any $\mathbf{u} \in \mathcal{C}$, we have

$$\Gamma^D\left(T^{-1}(\mathbf{u})\right) \cap S_Y^{\mathrm{typ}} \subseteq \mathrm{CONE}(\mathbf{u}, \theta'), \tag{67}$$

where

$$\theta' \triangleq \theta_0 + \theta_1 < \frac{\pi}{2}, \tag{68}$$

and the angles θ_0 and θ_1 are given by

$$\theta_0 \triangleq \arcsin\left(\sqrt{\frac{D_0}{\sigma_X^2}}\right) \tag{69}$$

$$\theta_1 \triangleq \arccos\left(\frac{\sigma_X^2 + \sigma_Y^2 - 2\eta - D}{2\sqrt{(\sigma_X^2 + \eta)(\sigma_Y^2 + \eta)}}\right). \tag{70}$$

Proof: Appendix B. □

Fig. 6 illustrates the claim in the lemma.

Let θ' be as defined in Lemma 4 above. We continue with

$$\begin{aligned} &\Pr\{\mathtt{maybe}|\mathbf{X} \in S_X^{\mathrm{typ}}, \mathbf{Y} \in S_Y^{\mathrm{typ}}\} \\ &\overset{(a)}{=} \Pr\left\{\mathbf{Y} \in \Gamma^D\left(T^{-1}(T(\mathbf{X}))\right) | \mathbf{X} \in S_X^{\mathrm{typ}}, \mathbf{Y} \in S_Y^{\mathrm{typ}}\right\} \\ &\overset{(b)}{\leq} \Pr\left\{\mathbf{Y} \in \mathrm{CONE}(T(\mathbf{X}), \theta') | \mathbf{X} \in S_X^{\mathrm{typ}}, \mathbf{Y} \in S_Y^{\mathrm{typ}}\right\} \\ &\overset{(c)}{=} \Omega(\theta') \\ &\overset{(d)}{\leq} \frac{1}{\sqrt{2\pi(n-1)}} \cdot \frac{1}{\cos\theta'} \cdot \sin^{n-1}\theta'. \end{aligned} \tag{71}$$

Identity (a) follows by definition of the query function $g(\cdot,\cdot)$. Inequality (b) follows from Lemma 4. Equality (c) follows since $\mathbf{Y}$ is uniformly distributed within each shell S_r of radius $r > 0$ (due to the spherical symmetry of the Gaussian distribution), and the probability of falling in a cap of a given half-angle θ' is precisely the fraction of the surface that is occupied by the cap, $\Omega(\theta')$. Inequality (d) follows since $\theta' \leq \arccos(1/\sqrt{n})$ for sufficiently large n, and therefore (41) applies.

Since $\theta' < \pi/2$, we have $\sin\theta' < 1$, and it therefore follows from (71) that the probability $\Pr\{\mathtt{maybe}|\mathbf{X} \in S^{\mathrm{typ}}, \mathbf{Y} \in S_Y^{\mathrm{typ}}\}$ vanishes with n. Thus, since ϵ was arbitrary, recalling (62) completes the direct part of the proof.

Remark 4: The alert reader will observe that the direct part also follows from the direct part of Theorem 2. However, we have chosen to include an explicit proof here to introduce the notations and ideas crucial for proving Theorem 2.

[5]This can easily be shown by the weak law of large numbers for the average $\frac{1}{n}\|\mathbf{X}\|^2 = \frac{1}{n}\sum_{i=1}^n X_i^2$. In fact, the vanishing is *exponential* with n, a fact that be shown, e.g. using the Chernoff bound.

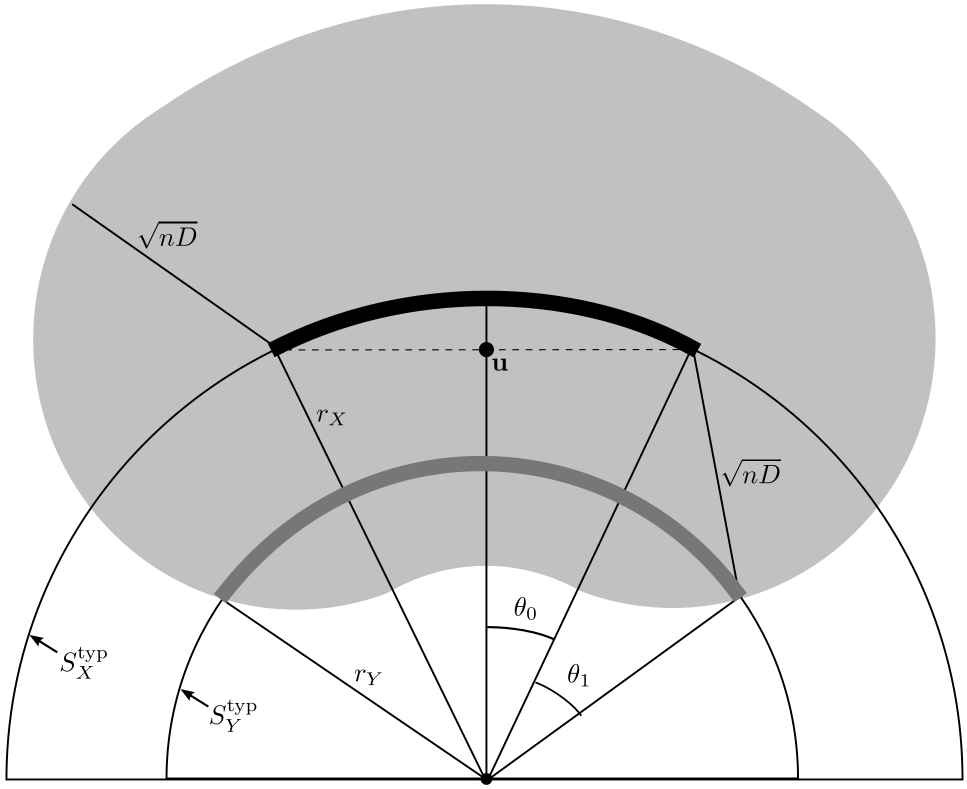

Fig. 6. Illustration for Lemma 4. The black region marks $\text{CAP}_{r^-,r^+}(\mathbf{u},\theta_0)$. The grey area denotes $\Gamma^D\left(\text{CAP}_{r^-,r^+}(\mathbf{u},\theta_0)\right)$, and the dark grey region marks the intersection $\Gamma^D\left(\text{CAP}_{r^-,r^+}(\mathbf{u},\theta_0)\right)\cap S_Y^{\text{typ}}$.

Converse Part: Let $\eta > 0$ and define S_X^{typ} as in (57). Let $T : \mathbb{R}^n \to \{1,\dots,2^{nR}\}$ be a given signature function corresponding to a D-admissible system (T, g), and assume that

$$\Pr\{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}\} \le \frac{1}{4} \tag{72}$$

since we are only interested in D-achievable rates R. As before, we will use the shorthand notation $\Pr\{\texttt{maybe}\} \triangleq \Pr\{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}\}$ to simplify the presentation.

We shall restrict our attention to the typical sphere. To this end, define the mapping $\tilde{T} : S_X^{\text{typ}} \to \{1,\dots,2^{nR}\}$, where $\tilde{T}(\mathbf{x}) = T(\mathbf{x})$ for $\mathbf{x} \in S_X^{\text{typ}}$. Let $\tilde{T}^{-1}(\cdot)$ denote the inverse mapping of $\tilde{T}(\cdot)$, i.e.

$$\tilde{T}^{-1}(i) \triangleq \{\mathbf{x} \in S_X^{\text{typ}} : T(\mathbf{x}) = i\} \tag{73}$$

$$= T^{-1}(i) \cap S_X^{\text{typ}}. \tag{74}$$

Let $p_i \triangleq \Pr\{\mathbf{X} \in \tilde{T}^{-1}(i) | \mathbf{X} \in S_X^{\text{typ}}\}$. Clearly, we have $\sum_{i=1}^{2^{nR}} p_i = 1$. Define the set $A_i \subseteq S_{r_X}$ to be projection of $\tilde{T}^{-1}(i)$ onto the sphere S_{r_X}:

$$A_i = \left\{ r_X \frac{\mathbf{x}}{\|\mathbf{x}\|} : \mathbf{x} \in \tilde{T}^{-1}(i) \right\}. \tag{75}$$

Let α_i denote the fraction of the surface area of S_{r_X} that is occupied by A_i. By the spherical symmetry of the pdf of $\mathbf{X}$, α_i is also equal to the probability that the projection of $\mathbf{X}$ onto S_{r_0} lies in A_i. Therefore $\alpha_i \ge p_i$, with equality if and only if $\tilde{T}^{-1}(i)$ is a thick cap with inner and outer radii $r^\pm \triangleq \sqrt{n(\sigma_X^2 \pm \eta)}$.

Let $D' \triangleq (\sqrt{D} + \sqrt{\sigma_X^2 - \eta} - \sqrt{\sigma_X^2})^2 < D$. It can easily be verified that

$$\Gamma^{D'}(A_i) \subseteq \Gamma^D\left(\tilde{T}^{-1}(i)\right). \tag{76}$$

Now let $D'' \triangleq (\sqrt{D'} + \sqrt{\sigma_Y^2 - \eta} - \sqrt{\sigma_Y^2})^2$, and let the set B_i denote the D''-expansion of A_i, restricted to the sphere S_{r_Y}, i.e.

$$B_i \triangleq \Gamma^{D''}(A_i) \cap S_{r_Y}. \tag{77}$$

The set B_i can also be thought of an expansion of a set $\tilde{A}_i \triangleq \frac{\sigma_Y}{\sigma_X}\cdot A_i$, with the alternative distance measure $\check{d}(\cdot,\cdot)$ defined over the sphere S_{r_Y} that measures the arc-length between the two points (i.e., the geodesic distance). Also note that $\alpha_i = \frac{|A_i|}{|S_{r_X}|} = \frac{|\tilde{A}_i|}{|S_{r_Y}|}$ where $|\cdot|$ is used to denote the (hyper-) surface area. Let $\beta_i = \frac{|B_i|}{|S_{r_Y}|}$ denote the fraction of S_{r_Y} that is occupied by B_i.

Let the set C_i denote the r_Y^-, r_Y^+ thickening of B_i as follows:

$$C_i = \left\{ \mathbf{y} \in S_Y^{\text{typ}} : r_Y \frac{\mathbf{y}}{\|\mathbf{y}\|} \in B_i \right\}. \tag{78}$$

Next, it can also be verified that

$$C_i \subseteq \Gamma^{D'}(A_i). \tag{79}$$

Suppose that $\mathbf{x} \in S_X^{\text{typ}}$ and that $T(\mathbf{x}) = i$. Then we have:

$$\begin{aligned}\Pr\{\texttt{maybe}|\mathbf{X} = \mathbf{x} \in S_X^{\text{typ}}\} &\geq \Pr\left\{\mathbf{Y} \in \Gamma^D\left(T^{-1}(i)\right)\right\}\\ &\overset{(a)}{\geq} \Pr\left\{\mathbf{Y} \in \Gamma^D\left(\tilde{T}^{-1}(i)\right)\right\}\\ &\overset{(b)}{\geq} \Pr\left\{\mathbf{Y} \in \Gamma^{D'}(A_i)\right\}\\ &\overset{(c)}{\geq} \Pr\{\mathbf{Y} \in C_i\},\end{aligned}$$

where (a) follows since $\tilde{T}^{-1}(i) \subseteq T^{-1}(i)$, and (b) and (c) follow from (76) and (79) respectively.

Let $f_{\mathbf{Y}}$ be the density of $\mathbf{Y}$. Then, we continue with

$$\Pr\{\mathbf{Y} \in C_i\} = \int_{C_i} f_{\mathbf{Y}}(\mathbf{y})d\mathbf{y} = \beta_i \cdot \Pr\{\mathbf{Y} \in S_Y^{\text{typ}}\},$$

where the second equality follows from the spherical symmetry of $f_{\mathbf{Y}}(\mathbf{y})$.

We now arrive at the main step in proving the converse. The key ingredient we require is the isoperimetric inequality on the surface of a hypersphere (Lemma 2, see also [25, Th. 1.1]) which states that, among all subsets of the hypersphere with a given surface area, spherical caps have minimum D-expansion measured under geodesic distance. As noted before, the set $B_i \subseteq S_{r_Y}$ is an expansion of the set $\tilde{A}_i \subseteq S_{r_Y}$ with the arclength (i.e., geodesic) distance measure. Therefore, it follows from the isoperimetric inequality that

$$\begin{aligned}|B_i| = \left|\Gamma^{D''}(A_i) \cap S_{r_Y}\right| &\geq \left|\Gamma^{D''}\left(\text{CAP}_{r_X}(\mathbf{u}, \theta_i)\right) \cap S_{r_Y}\right|\\ &= \left|\text{CAP}_{r_Y}(\mathbf{u}, \theta_i + \theta_{D''})\right|, \qquad (80)\end{aligned}$$

where $\mathbf{u}$ is an arbitrary point and

$$\theta_i \triangleq \Omega^{-1}(\alpha_i) \qquad (81)$$

$$\theta_{D''} \triangleq \arccos\left(\frac{\sigma_X^2 + \sigma_Y^2 - D''}{2\sigma_X\sigma_Y}\right). \qquad (82)$$

Therefore, we can conclude that if $\mathbf{x} \in S_X^{\text{typ}}$ and $T(\mathbf{x}) = i$, then

$$\Pr\{\texttt{maybe}|\mathbf{X} = \mathbf{x}\} \geq \Pr\{\mathbf{Y} \in S_Y^{\text{typ}}\} \cdot \Omega\left(\theta_{D''} + \Omega^{-1}(\alpha_i)\right).$$

Now, the average quantity $\Pr\{\texttt{maybe}|\mathbf{X} \in S_X^{\text{typ}}\}$ is bounded as follows

$$\begin{aligned}&\Pr\{\texttt{maybe}|\mathbf{X} \in S_X^{\text{typ}}\}\\ &= \sum_{i=1}^{2^{nR}} \Pr\{T(\mathbf{X}) = i|\mathbf{X} \in S_X^{\text{typ}}\}\\ &\quad \times \Pr\{\texttt{maybe}|T(\mathbf{X}) = i, \mathbf{X} \in S_X^{\text{typ}}\}\\ &\geq \sum_{i=1}^{2^{nR}} p_i \cdot \Pr\{\mathbf{Y} \in S_Y^{\text{typ}}\} \cdot \Omega\left(\theta_{D''} + \Omega^{-1}(\alpha_i)\right) \qquad (83)\\ &\geq \Pr\{\mathbf{Y} \in S_Y^{\text{typ}}\} \cdot \sum_{i=1}^{2^{nR}} p_i \cdot \Omega\left(\theta_{D''} + \Omega^{-1}(p_i)\right), \qquad (84)\end{aligned}$$

where the last inequality follows since $\alpha_i \geq p_i$ and the function $\Omega(\theta_{D''} + \Omega^{-1}(\cdot))$ is monotone increasing.

If the scheme at hand were to satisfy $p_i = 2^{-nR}$ for all i, then we could simply continue with analyzing $\Omega\left(\theta_{D''} + \Omega^{-1}(2^{-nR})\right)$. However, in general this might not be the case. We therefore require the following lemma:

Lemma 5: Let $0 < \Omega^* < 1$ *and* $0 < c < 1$ *be given constants. Define* p^* *to be the solution to* $\Omega(\theta_{D''} + \Omega^{-1}(p)) = \Omega^*$. *Then if*

$$\sum_{i=1}^{2^{nR}} p_i \cdot \Omega\left(\theta_{D''} + \Omega^{-1}(p_i)\right) \leq c \cdot \Omega^*, \qquad (85)$$

then

$$R \geq \frac{1}{n}\log\frac{1-c}{p^*}. \qquad (86)$$

Proof: Appendix C. □

For our purposes[6] we set $\Omega^* = \frac{1}{2}$ so that $\Omega(\theta_{D''} + \Omega^{-1}(p^*)) = \frac{1}{2}$. Now use (41) to upper bound $\Omega(\cdot)$ and evaluate p^*:

$$\begin{aligned}p^* &= \Omega\left(\frac{\pi}{2} - \theta_{D''}\right)\\ &\leq \frac{1}{\sqrt{2\pi(n-1)}} \cdot \frac{1}{\cos\left(\frac{\pi}{2} - \theta_{D''}\right)} \cdot \sin^{n-1}\left(\frac{\pi}{2} - \theta_{D''}\right)\\ &\leq \frac{1}{\sqrt{2\pi(n-1)}} \cdot \cos^{n-1}(\theta_{D''}).\end{aligned}$$

Recalling the definition of $\theta_{D''}$, we have

$$\cos(\theta_{D''}) = \frac{\sigma_X^2 + \sigma_Y^2 - D''}{2\sigma_X\sigma_Y},$$

therefore

$$\frac{1}{n}\log\frac{1}{p^*} = \log\frac{2\sigma_X\sigma_Y}{\sigma_X^2 + \sigma_Y^2 - D''} + O\left(\frac{\log n}{n}\right). \qquad (87)$$

Our goal, now, is to show that the rate R must be *lower* bounded by the identification rate from (14). Recalling (72), it follows that

$$\begin{aligned}\frac{1}{4} &\geq \Pr\{\texttt{maybe}\} \qquad (88)\\ &= \Pr\{\mathbf{X} \in S_X^{\text{typ}}\} \cdot \Pr\{\texttt{maybe}|\mathbf{X} \in S_X^{\text{typ}}\} \qquad (89)\\ &\quad + \Pr\{\mathbf{X} \notin S_X^{\text{typ}}\} \cdot \Pr\{\texttt{maybe}|\mathbf{X} \notin S_X^{\text{typ}}\} \qquad (90)\\ &\geq \Pr\{\mathbf{X} \in S_X^{\text{typ}}\} \cdot \Pr\{\texttt{maybe}|\mathbf{X} \in S_X^{\text{typ}}\} \qquad (91)\\ &\geq \Pr\{\mathbf{X} \in S_X^{\text{typ}}\} \cdot \Pr\{\mathbf{Y} \in S_Y^{\text{typ}}\}\\ &\qquad \cdot \sum_{i=1}^{2^{nR}} p_i \cdot \Omega\left(\theta_{D''} + \Omega^{-1}(p_i)\right), \qquad (92)\end{aligned}$$

where the final inequality is simply (84).

Since $\Pr\{\mathbf{Y} \in S_Y^{\text{typ}}\}$ and $\Pr\{\mathbf{X} \in S_X^{\text{typ}}\}$ both approach 1 as n grows, we may assume that both probabilities are above $\frac{3}{4}$ (for large enough n). Then, we can now invoke Lemma 5 with $c = 8/9$ and $\Omega^* = 1/2$, combined with (87), to conclude that

$$R \geq \log\frac{2\sigma_X\sigma_Y}{\sigma_X^2 + \sigma_Y^2 - D''} + O\left(\frac{\log n}{n}\right). \qquad (93)$$

As η can be taken to be arbitrarily small, D'' can be arbitrarily close to D, completing the proof of the converse. □

[6] We shall use Lemma 5 again for proving the identification exponent results, but with a different Ω^*.

D. Identification Exponent

As with Theorem 1, the proof of Theorem 2 is rather involved, so we first sketch the main ideas before moving on to the formal proof. Characterizing the optimal exponent requires a slightly more sophisticated scheme than characterizing the identification rate, but the proofs are very similar in spirit.

The achievability proof builds upon that of Theorem 1 in the sense that we refine the signature assignment to quantize $\mathbf{x}/\|\mathbf{x}\|$ and $\|\mathbf{x}\|$ separately. Intuitively, we can think of our scheme as quantizing the *direction* and *amplitude* of the vector $\mathbf{x}$ (similarly to 'shape-gain' quantizers [29, Ch. 12]). Similar to the achievability proof of Theorem 1, the set of vectors $\mathbf{x}/\|\mathbf{x}\|$ are quantized by covering the unit sphere with regions roughly described by caps. It will turn out that the achievable identification exponent emerges through the analysis of quantizing the amplitudes $\mathbf{x}$.

For the converse proof, we take the ρ_X^*, ρ_Y^* to minimize (18), and focus on the case where $\mathbf{X}$ lies in a spherical shell with radius $\sqrt{n\rho_X^*\sigma_X^2}$ and small, nonzero thickness. Then, the converse proceeds similar to that of Theorem 1, in the sense that the "typical shell" is replaced by the new shell that depends on ρ_X^*.

Proof of Theorem 2:

Direct Part: We will rely on the code construction given in the achievability proof of Theorem 1, and hence we adopt the notation previously defined there. To this end, let (T, g) be the rate-R, D-admissible identification system defined in the achievability proof of Theorem 1. Recall that

$$\angle\left(T\left(r_X\frac{\mathbf{x}}{\|\mathbf{x}\|}\right), \frac{\mathbf{x}}{\|\mathbf{x}\|}\right) \le \theta_0, \tag{94}$$

where θ_0 was defined as (69).

In a variation on the scheme used previously, we describe the amplitude $\|\mathbf{x}\|$ by quantization as follows. Let $\sigma_{\max}^2(n) \triangleq n\cdot\sigma_X^2$, and recall that η was chosen to be a small positive constant. Define the spherical shells $S^{(i)}$ as follows:

$$S^{(i)} \triangleq S_{r^{(i)},r^{(i+1)}}, \tag{95}$$

where $r^{(i)} \triangleq \sqrt{n\cdot i\cdot\eta}$.

The modified signature assignment T' then describes the "direction" and "amplitude" of $\mathbf{x}$ as follows:

- If $\frac{1}{n}\|\mathbf{x}\|^2 \le \sigma_{\max}^2(n)$, then $T'(\mathbf{x}) = \left(T\left(r_X\frac{\mathbf{x}}{\|\mathbf{x}\|}\right), i\right)$, where i is chosen to satisfy $\mathbf{x}\in S^{(i)}$.
- If $\frac{1}{n}\|\mathbf{x}\|^2 > \sigma_{\max}^2(n)$, then the signature $T'(\mathbf{x})$ is defined to be the erasure symbol e.

The overall rate of the modified signature assignment T' described above is R (i.e., the rate of $T(\cdot)$), plus an additional $\frac{1}{n}\log\frac{\sigma_{\max}^2(n)}{\eta} = O(\frac{\log n}{n})$ (required for the quantization of $\|\mathbf{x}\|$), and therefore remains essentially unchanged. Therefore, the upper bound (62) also upper bounds the rate of the modified signature assignment function. Let g' be the optimal query function corresponding to T' (defined in an analogous manner to (63)).

Thus, we only need to analyze the exponent attained by the proposed scheme. To this end, let Z be a Chi-square random variable with n degrees of freedom. The pdf of Z is given by

$$f_Z(z) = \frac{z^{\frac{n}{2}-1}e^{-\frac{z}{2}}}{2^{n/2}\Gamma(\frac{n}{2})}, \tag{96}$$

where Γ in (96) is the usual Gamma function, and should not be confused with the set-expansion operator Γ^D defined previously. Now, define the random variables $Z_X \triangleq \frac{1}{\sigma_X^2}\|\mathbf{X}\|^2$ and $Z_Y \triangleq \frac{1}{\sigma_Y^2}\|\mathbf{Y}\|^2$. Note that both Z_X and Z_Y are distributed according to (96). In order to proceed, we require the following lemma.

Lemma 6: The probability $\Pr\left\{\frac{1}{n}\|\mathbf{X}\|^2 > \sigma_{\max}^2(n)\right\}$ *vanishes super-exponentially with* n.

Proof: Appendix D. □

Now, we are in a position to analyze $\Pr\{\mathtt{maybe}\}$, where we again employ the shorthand notation $\Pr\{\mathtt{maybe}\} \triangleq \Pr\{g'(T'(\mathbf{X}),\mathbf{Y}) = \mathtt{maybe}\}$ to simplify the presentation.

$$\begin{aligned}&\Pr\{\mathtt{maybe}\}\\ &\le \Pr\left\{\mathtt{maybe}, \frac{1}{n}\|\mathbf{X}\|^2 \le \sigma_{\max}^2(n), \frac{1}{n}\|\mathbf{Y}\|^2 \le \sigma_{\max}^2(n)\right\}\\ &\quad + \Pr\left\{\frac{1}{n}\|\mathbf{X}\|^2 > \sigma_{\max}^2(n)\right\} + \Pr\left\{\frac{1}{n}\|\mathbf{Y}\|^2 > \sigma_{\max}^2(n)\right\}.\end{aligned}$$

By Lemma 6, and a similar argument for $\Pr\left\{\frac{1}{n}\|\mathbf{Y}\|^2 > \sigma_{\max}^2(n)\right\}$, the last two terms of the above expression vanish super-exponentially and do not affect the exponent of $\Pr\{\mathtt{maybe}\}$. We therefore concentrate on the first term.

We can now write

$$\Pr\left\{\mathtt{maybe}, \frac{1}{n}\|\mathbf{X}\|^2 \le \sigma_{\max}^2(n), \frac{1}{n}\|\mathbf{Y}\|^2 \le \sigma_{\max}^2(n)\right\} \tag{97}$$

$$= \Pr\left\{\mathtt{maybe}, Z_X \le n^2, Z_Y \le \frac{\sigma_Y^2}{\sigma_X^2}n^2\right\} \tag{98}$$

$$= \int_0^{n^2}\int_0^{\frac{\sigma_Y^2}{\sigma_X^2}n^2} \Pr\{\mathtt{maybe} \mid Z_X = z_X, Z_Y = z_Y\} \cdot f_Z(z_X)f_Z(z_Y)dz_Y dz_X \tag{99}$$

$$\le \frac{\sigma_Y^2}{\sigma_X^2}n^4 \max_{\substack{0\le z_X\le n^2\\ 0\le z_Y\le n^2\sigma_Y^2/\sigma_X^2}} \Pr\{\mathtt{maybe} \mid Z_X = z_X, Z_Y = z_Y\} \cdot f_Z(z_X)f_Z(z_Y) \tag{100}$$

$$\le \frac{\sigma_Y^2}{\sigma_X^2}n^4 \max_{0\le\rho_X,\rho_Y} \Pr\{\mathtt{maybe} \mid Z_X = n\rho_X, Z_Y = n\rho_Y\} \cdot f_Z(n\rho_X)f_Z(n\rho_Y), \tag{101}$$

where $\rho_X \triangleq z_X/n$ and $\rho_Y \triangleq z_Y/n$.

The event $\{\mathtt{maybe}\}$ coincides with the event $\{\mathbf{Y} \in \Gamma^D(T'^{-1}(T'(\mathbf{X})))\}$. Let $\mathbf{U} = T(r_X\mathbf{X}/\|\mathbf{X}\|)$, and observe that if $\frac{1}{n}\|\mathbf{X}\|^2 \le \sigma_{\max}^2(n)$, then

$$\Gamma^D(T'^{-1}(T'(\mathbf{X}))) \subseteq \Gamma^D(\mathrm{CAP}_{r^{(i)},r^{(i+1)}}(\mathbf{U},\theta_0)) \tag{102}$$

$$\subseteq \Gamma^{D'}(\mathrm{CAP}_{\|\mathbf{X}\|}(\mathbf{U},\theta_0))), \tag{103}$$

where (102) follows from similar arguments leading to (59), and (103) follows with $D' \triangleq (\sqrt{D}+\sqrt{\eta})^2$. We therefore continue with

$$\Pr\{\texttt{maybe} \mid Z_X = n\rho_X, Z_Y = n\rho_Y\}$$
$$\leq \Pr\left\{\mathbf{Y} \in \Gamma^{D'}(\mathrm{CAP}_{\|\mathbf{X}\|}(\mathbf{U}, \theta_0)) \;\middle|\; \frac{1}{n}\|\mathbf{X}\|^2 = \rho_X\sigma_X^2, \frac{1}{n}\|\mathbf{Y}\|^2 = \rho_Y\sigma_Y^2\right\}$$
$$= \Pr\left\{\mathbf{Y} \in \Gamma^{D'}(\mathrm{CAP}_{\sqrt{n\rho_X\sigma_X^2}}(\mathbf{U}, \theta_0)) \mid \frac{1}{n}\|\mathbf{Y}\|^2 = \rho_Y\sigma_Y^2\right\} \tag{104}$$
$$= \begin{cases} 0 & \text{if } |\sqrt{\rho_X\sigma_X^2} - \sqrt{\rho_Y\sigma_Y^2}| \geq \sqrt{D'} \\ 1 & \text{if } \sqrt{\rho_X\sigma_X^2} + \sqrt{\rho_Y\sigma_Y^2} \leq \sqrt{D'} \\ \Omega(\theta_0 + \theta_1') & \text{otherwise,} \end{cases} \tag{105}$$

where (104) follows by spherical symmetry of the Gaussian distribution, and

$$\theta_1' \triangleq \arccos \frac{\rho_x\sigma_X^2 + \rho_Y\sigma_Y^2 - D'}{2\sqrt{\rho_X\sigma_X^2 \cdot \rho_Y\sigma_Y^2}} \in [0, \pi]. \tag{106}$$

Note that the condition $\sqrt{\rho_X\sigma_X^2} + \sqrt{\rho_Y\sigma_Y^2} \leq \sqrt{D'}$ is not the only case in which the probability is equal to one. The other case is when $\theta_0 + \theta_1' \geq \pi$, so $\Omega(\theta_0 + \theta_1') = 1$.

The identity (106) follows from the law of cosines. The geometric image now is similar to that depicted in Fig. 6, where here $r_X \triangleq \sqrt{n\sigma^2\rho_X}$ and $r_Y \triangleq \sqrt{n\sigma^2\rho_Y}$ denote the actual radii of the vectors $\mathbf{X}$ and $\mathbf{Y}$ (as opposed to their average value in the proof of Theorem 1).

Next, using the bound (41) we have

$$\frac{1}{n}\log\frac{1}{\Omega(\theta)} \geq \begin{cases} -\log\sin\theta + \frac{c}{n}\log n, & 0 < \theta < \arccos(\frac{1}{\sqrt{n}}); \\ 0, & \text{otherwise.} \end{cases} \tag{107}$$

where c is a universal constant.

Combined with (107), we compactly write the exponent corresponding to expression (105) as

$$\mathbf{E}_\Omega(\theta_0, D', \sigma_X^2, \sigma_Y^2, \rho_X, \rho_Y) \triangleq$$
$$= \begin{cases} \infty, & \text{if } |\sqrt{\rho_X\sigma_X^2} - \sqrt{\rho_Y\sigma_Y^2}| \geq \sqrt{D'} \\ 0, & \text{if } \rho_X\sigma_X^2 + \rho_Y\sigma_Y^2 \leq D' \\ -\log\sin\min\left[\frac{\pi}{2}, \theta_0 + \theta_1'\right], & \text{otherwise.} \end{cases} \tag{108}$$

with θ_1' given in (106). Note that the case of $D' \in [\rho_X\sigma_X^2 + \rho_Y\sigma_Y^2, (\sqrt{\rho_X\sigma_X^2} + \sqrt{\rho_Y\sigma_Y^2})^2]$ corresponds to the case of $\theta_1' \in [\pi/2, \pi]$, so $\Omega(\theta_0 + \theta_1') \to 1$ as n grows. In the case of $D' > (\sqrt{\rho_X\sigma_X^2} + \sqrt{\rho_Y\sigma_Y^2})^2$ it follows from (105) that the probability is equal to zero (for all n). Hence the simplified condition $\rho_X\sigma_X^2 + \rho_Y\sigma_Y^2 \leq D'$ in (108).

Before we plug the above result into (101), we note that by Stirling's approximation we may write, for any fixed $\rho > 0$:

$$f_Z(n\rho) = \frac{1}{n\rho}\left(\frac{n\rho}{2}\right)^{n/2}\exp(-n\rho/2)\frac{1}{\Gamma(n/2)}$$
$$= \frac{1}{n\rho}\left(\frac{n\rho}{2}\right)^{n/2}\exp(-n\rho/2)\frac{1}{\sqrt{\frac{4\pi}{n}}\left(\frac{n}{2e}\right)^{n/2}}\left(1+O\left(\frac{1}{n}\right)\right)$$
$$= \exp\left[-n\left(\frac{\rho}{2} - \frac{1}{2} - \frac{1}{2}\log\rho\right)\right]\frac{1}{\rho\sqrt{4\pi n}}\left(1+O\left(\frac{1}{n}\right)\right)$$
$$\leq 2^{-n\mathbf{E}_Z(\rho)} \cdot n^c, \tag{109}$$

where $\mathbf{E}_Z(\cdot)$ was defined in (16) and c is a universal constant.

Finally, we plug (108) and (109) into the upper bound (101) on the (conditional) probability for maybe and conclude that the following exponent is achievable:

$$\min_{\rho_X,\rho_Y\geq 0} \mathbf{E}_Z(\rho_X) + \mathbf{E}_Z(\rho_Y) + \mathbf{E}_\Omega(\theta_0, D', \sigma_X^2, \sigma_Y^2, \rho_X, \rho_Y). \tag{110}$$

Since η is arbitrarily small we may replace D' with D in the above. We may therefore rewrite the achievable exponent as

$$\min_{\rho_X,\rho_Y\geq 0} \mathbf{E}_Z(\rho_X) + \mathbf{E}_Z(\rho_Y) + \mathbf{E}_\Omega\left(\arcsin(2^{-R}), D, \sigma_X^2, \sigma_Y^2, \rho_X, \rho_Y\right). \tag{111}$$

In order to simplify matters further, note that in (111), the minimizing (ρ_X, ρ_Y) must satisfy:

$$\left|\sqrt{\rho_X\sigma_X^2} - \sqrt{\rho_Y\sigma_Y^2}\right| < \sqrt{D} \tag{112}$$
$$\rho_X\sigma_X^2 + \rho_Y\sigma_Y^2 > D. \tag{113}$$

The condition (112) must hold because otherwise the term $\mathbf{E}_\Omega$ is infinite [see (108)].

To prove that (113) must hold, assume, for contradiction, that (111) is minimized for (ρ_X^*, ρ_Y^*) that satisfy

$$\rho_X\sigma_X^2 + \rho_Y\sigma_Y^2 \leq D' \tag{114}$$

In this case, the value of (111) at the minimizing point is $\mathbf{E}_Z(\rho_X^*)+\mathbf{E}_Z(\rho_Y^*)$. If, say $\rho_X^* > 1$, then we may replace it with another value $0 < \rho_X^{**} < 1$ that satisfies $\mathbf{E}_Z(\rho_X^{**}) = \mathbf{E}_Z(\rho_X^*)$ that is guaranteed to exist (see the definition of $\mathbf{E}_Z(\cdot)$). The same argument holds for ρ_Y^*, and therefore we may assume that in this case both $\rho_X^*, \rho_Y^* \in (0, 1]$. Next, since $\mathbf{E}_Z(\rho)$ is monotone decreasing for $\rho \in (0, 1)$, we may increase ρ_X^* and ρ_Y^*, while still in $(0, 1]^2$, until (113) is met with an equality. Since the value of the objective function decreases, we arrive at a contradiction, meaning that (113) must hold for any minimizing ρ_X, ρ_Y.

Therefore the achievable exponent can be simplified to the expression (18) and the proof of the direct part is concluded.

Converse Part: Let ρ_X^*, ρ_Y^* denote the minimizers of (18) (in light of the discussion above, we can assume without loss of generality that ρ_X^*, ρ_Y^* satisfy (19)). The proof of the converse proceeds by focusing on values of $\mathbf{X}$ and $\mathbf{Y}$ that satisfy $\frac{1}{n}\|\mathbf{X}\|^2 \cong \rho_X^*\sigma_X^2$ and $\frac{1}{n}\|\mathbf{Y}\|^2 \cong \rho_Y^*\sigma_Y^2$. The details are as follows:

Let $0 < \eta < \min(\rho_X^*, \rho_Y^*)$ be a small but fixed value. Define the following spherical caps:

$$S_X^* \triangleq S_{r_X^-, r_X^+}, \quad S_Y^* \triangleq S_{r_Y^-, r_Y^+}, \tag{115}$$

where $r_X^\pm \triangleq \sqrt{n\sigma_X^2(\rho_X^* \pm \eta)}$ and $r_Y^\pm \triangleq \sqrt{n\sigma_Y^2(\rho_Y^* \pm \eta)}$.

We then write the following:

$$\begin{aligned}\Pr\{\texttt{maybe}\} &\geq \Pr\left\{\texttt{maybe}, \mathbf{X} \in S_X^*, \mathbf{Y} \in S_Y^*\right\} \\ &= \Pr\left\{\texttt{maybe} | \mathbf{X} \in S_X^*, \mathbf{Y} \in S_Y^*\right\} \\ &\quad \cdot \Pr\{\mathbf{X} \in S_X^*\} \cdot \Pr\{\mathbf{Y} \in S_Y^*\}. \end{aligned} \tag{116}$$

Consider the term $\Pr\{\mathbf{X} \in S_X^*\}$:

$$\Pr\left\{\mathbf{X} \in S_X^*\right\} = \Pr\left\{\frac{1}{n\sigma_X^2}\|\mathbf{X}\|^2 \in (\rho_X^* - \eta, \rho_X^* + \eta)\right\} \tag{117}$$

$$= \int_{n(\rho_X^*-\eta)}^{n(\rho_X^*+\eta)} f_Z(z)dz \tag{118}$$

$$\geq 2n\eta \min_{z\in[n(\rho_X^*-\eta), n(\rho_X^*+\eta)]} f_Z(z) \tag{119}$$

$$\geq 2n\eta \cdot n^c \cdot 2^{-n \max_{\rho_X \in [\rho_X^*-\eta, \rho_X^*+\eta]} \mathbf{E}_Z(\rho_X)}, \tag{120}$$

where (120) follows from Stirling's approximation similar to (109). A similar derivation applies for $\Pr\{\mathbf{Y} \in S_Y^*\}$. Thus, it follows from (116) and continuity of $\mathbf{E}_Z(\cdot)$ that

$$\begin{aligned}-\frac{1}{n}\log \Pr\{\texttt{maybe}\} \leq& -\frac{1}{n}\log\left[\Pr\left\{\texttt{maybe}|\mathbf{X} \in S_X^*, \mathbf{Y} \in S_Y^*\right\}\right] \\ &+ \mathbf{E}_Z(\rho_X^*) + \mathbf{E}_Z(\rho_Y^*) + \eta' \\ &+ O\left(\frac{\log n}{n}\right),\end{aligned} \tag{121}$$

where η' is a quantity tending to zero as $\eta \to 0$.

We now concentrate on the term $\Pr\left\{\texttt{maybe}|\mathbf{X} \in S_X^*, \mathbf{Y} \in S_Y^*\right\}$, and proceed in a manner similar to the converse proof of Theorem 1. To this end, let $T : \mathbb{R}^n \to \{1, \ldots, 2^{nR}\}$ denote the signature assignment for the scheme at hand. Define the mapping $\tilde{T} : S_X^* \to \{1, \ldots, 2^{nR}\}$ as $\tilde{T}(\mathbf{x}) = T(\mathbf{x})$ for all $\mathbf{x} \in S_X^*$. That is, $\tilde{T}(\cdot)$ is the restriction of $T(\cdot)$ to S_X^*. Let $\tilde{T}^{-1}(\cdot)$ denote the inverse mapping of $\tilde{T}(\cdot)$:

$$\begin{aligned}\tilde{T}^{-1}(i) &\triangleq \{\mathbf{x} \in S_X^* : T(\mathbf{x}) = i\} &(122)\\ &= T^{-1}(i) \cap S_X^*. &(123)\end{aligned}$$

Let $p_i \triangleq \Pr\{\mathbf{X} \in \tilde{T}^{-1}(i) | \mathbf{X} \in S_X^*\}$, so $\sum_{i=1}^{2^{nR}} p_i = 1$. Define $r_X \triangleq \sqrt{n\sigma_X^2\rho_X^*}$, and let the set $A_i \subseteq S_{r_X}$ denote the projection of $\tilde{T}^{-1}(i)$ onto the sphere S_{r_X}. In other words,

$$A_i = \left\{r_X \frac{\mathbf{x}}{\|\mathbf{x}\|} : \mathbf{x} \in \tilde{T}^{-1}(i)\right\}. \tag{124}$$

Let α_i denote the fraction of the surface area of S_{r_X} that is occupied by A_i. By the spherical symmetry of the distribution of $\mathbf{X}$, α_i is also equal to the probability that the projection of $\mathbf{X}$ onto S_{r_X} lies in A_i. Therefore $\alpha_i \geq p_i$, with equality if and only if $\tilde{T}^{-1}(i)$ is a thick cap with inner and outer radii r_X^- and r_X^+ respectively.

Let $D' \triangleq (\sqrt{D} + \sqrt{\sigma_X^2(\rho_X^* - \eta)} - \sqrt{\sigma_X^2\rho_X^*})^2$. As in (76) we have

$$\Gamma^{D'}(A_i) \subseteq \Gamma^D\left(\tilde{T}^{-1}(i)\right). \tag{125}$$

Now let $D'' \triangleq (\sqrt{D'} + \sqrt{\sigma_Y^2(\rho_Y^* - \eta)} - \sqrt{\sigma_Y^2\rho_Y^*})^2$, and let the set $B_i \subseteq S_{r_Y}$ denote the D''-expansion of A_i, restricted to the sphere S_{r_Y}, where $r_Y \triangleq \sqrt{n\sigma_Y^2\rho_Y^*}$, i.e.

$$B_i \triangleq \Gamma^{D''}(A_i) \cap S_{r_Y}. \tag{126}$$

Let β_i denote the fraction of S_{r_Y} that is occupied by B_i. Let the set C_i denote the r_Y^-, r_Y^+ thickening of B_i as follows:

$$C_i = \left\{\mathbf{y} \in S_Y^* : r_Y \frac{\mathbf{y}}{\|\mathbf{y}\|} \in B_i\right\}. \tag{127}$$

As in (79) we have

$$C_i \subseteq \Gamma^{D'}(A_i). \tag{128}$$

Suppose that $\mathbf{X} = \mathbf{x} \in S_X^*$ and that $T(\mathbf{x}) = i$. Then we have, with the aid of (125) and (128):

$$\begin{aligned}&\Pr\{\texttt{maybe}|\mathbf{X} = \mathbf{x} \in S_X^*, \mathbf{Y} \in S_Y^*\} \\ &\quad\geq \Pr\left\{\mathbf{Y} \in \Gamma^D\left(T^{-1}(i)\right) | \mathbf{Y} \in S_Y^*\right\} &(129)\\ &\quad\geq \Pr\left\{\mathbf{Y} \in \Gamma^D\left(\tilde{T}^{-1}(i)\right) | \mathbf{Y} \in S_Y^*\right\} &(130)\\ &\quad\geq \Pr\left\{\mathbf{Y} \in \Gamma^{D'}(A_i) | \mathbf{Y} \in S_Y^*\right\} &(131)\\ &\quad\geq \Pr\left\{\mathbf{Y} \in C_i | \mathbf{Y} \in S_Y^*\right\} &(132)\\ &\quad= \beta_i, &(133)\end{aligned}$$

where the last equality follows from the spherical symmetry of the pdf of $\mathbf{Y}$.

As in the proof of the converse of Theorem 1, we apply the isoperimetric inequality on the sphere (Lemma 2) for the sets A_i and B_i. We get that the set A_i^* that minimizes β_i for given α_i is the set $\text{CAP}_{r_X}(\mathbf{u}, \theta_i)$, where $\mathbf{u}$ is an arbitrary point, $\theta_i \triangleq \mathbf{\Omega}^{-1}(\alpha_i)$, and B_i^* is the set $\text{CAP}_{r_Y}(\mathbf{u}, \theta_i')$, defined by

$$\theta_i' \triangleq \theta_i + \theta_{D''} \tag{134}$$

where[7]

$$\theta_{D''} \triangleq \arccos \frac{\rho_X^*\sigma_X^2 + \rho_Y^*\sigma_Y^2 - D''}{2\sqrt{\rho_X^*\sigma_X^2 \cdot \rho_Y^*\sigma_Y^2}}. \tag{135}$$

Therefore the (normalized) surface area of B_i^* is given by $\beta_i^* = \mathbf{\Omega}(\theta_i')$. It follows that

$$\Pr\{\texttt{maybe}|\mathbf{X} = \mathbf{x} \in S_X^*, \mathbf{Y} \in S_Y^*\} \geq \mathbf{\Omega}(\theta_{D''} + \mathbf{\Omega}^{-1}(p_i)), \tag{136}$$

[7]Note that since ρ_X^* and ρ_Y^* satisfy (19), cannot be outside the region $[-1, 1]$.

and the average (conditional) probability $\Pr\{\texttt{maybe}|\mathbf{X} \in S_X^*, \mathbf{Y} \in S_Y^*\}$ is bounded by

$$\Pr\{\texttt{maybe}|\mathbf{X} \in S_X^*, \mathbf{Y} \in S_Y^*\} \tag{137}$$

$$= \sum_{i=1}^{2^{nR}} \Pr\{T(\mathbf{X}) = i|\mathbf{X} \in S_X^*\} \times \Pr\{\texttt{maybe}|T(\mathbf{X}) = i, \mathbf{X} \in S_X^*, \mathbf{Y} \in S_Y^*\} \tag{138}$$

$$\geq \sum_{i=1}^{2^{nR}} p_i \cdot \Omega\left(\theta_{D''} + \Omega^{-1}(p_i)\right). \tag{139}$$

Now, let $0 < c < 1$, and invoke Lemma 5 to conclude that

$$R \geq \frac{1}{n} \log \frac{1-c}{p^*}, \tag{140}$$

where p^* is the solution to

$$\Omega(\theta_{D''} + \Omega^{-1}(p)) = c^{-1} \Pr\{\texttt{maybe}|\mathbf{X} \in S_X^*, \mathbf{Y} \in S_Y^*\}. \tag{141}$$

Since $\Omega(\cdot)$ is monotone increasing, so is $\Omega^{-1}(\cdot)$. Therefore, (140) and (141) imply the inequality

$$\Pr\{\texttt{maybe}|\mathbf{X} \in S_X^*, \mathbf{Y} \in S_Y^*\} \geq c \cdot \Omega\left(\theta_{D''} + \Omega^{-1}\left((1-c)2^{-nR}\right)\right). \tag{142}$$

It is a straightforward exercise to verify (e.g., by Taylor series expansion) that

$$\Omega^{-1}\left((1-c)2^{-nR}\right) = \arcsin\left(2^{-R}\right) + O\left(\tfrac{\log n}{n}\right). \tag{143}$$

If $\theta_{D''} + \arcsin\left(2^{-R}\right) \geq \pi/2$, then (142) and the definition of $\Omega(\cdot)$ yield

$$\Pr\{\texttt{maybe}|\mathbf{X} \in S_X^*, \mathbf{Y} \in S_Y^*\} \geq c/2, \tag{144}$$

which, combined with (121), yields the desired upper bound

$$-\frac{1}{n} \log \Pr\{\texttt{maybe}\} \leq -\log\sin\left(\frac{\pi}{2}\right) + \mathbf{E}_Z(\rho_X^*) + \mathbf{E}_Z(\rho_Y^*) + \eta' + O\left(\frac{\log n}{n}\right). \tag{145}$$

On the other hand, if $\theta_{D''} + \arcsin\left(2^{-R}\right) < \pi/2$, then the hypothesis of Lemma 1 is satisfied for n sufficiently large, and the estimate (42) gives

$$-\frac{1}{n} \log \Pr\{\texttt{maybe}|\mathbf{X} \in S_X^*, \mathbf{Y} \in S_Y^*\} \leq -\log\sin\left(\theta_{D''} + \Omega^{-1}\left((1-c)2^{-nR}\right)\right) + O\left(\frac{\log n}{n}\right). \tag{146}$$

By letting η be arbitrarily small we can infer from (145) and (146) that any sequence of identification schemes $\{g^{(n)}, T^{(n)}\}_{n\to\infty}$ must satisfy

$$\limsup_{n\to\infty} -\frac{1}{n} \log \Pr\{g^{(n)}(T^{(n)}(\mathbf{X}), \mathbf{Y}) = \texttt{maybe}\} \leq \mathbf{E}_Z(\rho_X^*) + \mathbf{E}_Z(\rho_Y^*) - \log\sin\min\left[\frac{\pi}{2}, \arcsin\left(2^{-R}\right) + \arccos\frac{\rho_X^*\sigma_X^2 + \rho_Y^*\sigma_Y^2 - D}{2\sqrt{\rho_X^*\sigma_X^2 \cdot \rho_Y^*\sigma_Y^2}}\right],$$

as desired. □

Proof of Corollary 2: Let ρ_X, ρ_Y satisfy (19). We claim that the quantity

$$\mathbf{E}_Z(\rho_X) + \mathbf{E}_Z(\rho_Y) + \wp(R, D, \rho_X\sigma^2, \rho_Y\sigma^2) \tag{147}$$

can not increase if ρ_X and ρ_Y are both replaced by their average $\overline{\rho} \triangleq (\rho_X + \rho_Y)/2$, which continues to satisfy (19). To see that this is indeed the case, note that $\mathbf{E}_Z(\cdot)$ is convex, and therefore Jensen's inequality implies

$$\mathbf{E}_Z(\rho_X) + \mathbf{E}_Z(\rho_Y) \geq 2\mathbf{E}_Z(\overline{\rho}). \tag{148}$$

Next, the inequality of arithmetic and geometric means implies

$$\frac{\rho_X\sigma^2 + \rho_Y\sigma_Y^2 - D}{2\sigma^2\sqrt{\rho_X\rho_Y}} \geq \frac{2\overline{\rho}\sigma^2 - D}{2\overline{\rho}\sigma^2}, \tag{149}$$

and therefore, since $\arccos(x)$ is monotone decreasing on $x \in [0, 1]$,

$$\arccos\frac{\rho_X\sigma^2 + \rho_Y\sigma_Y^2 - D}{2\sigma^2\sqrt{\rho_X\rho_Y}} \leq \arccos\frac{2\overline{\rho}\sigma^2 - D}{2\overline{\rho}\sigma^2}. \tag{150}$$

Since $-\log\sin(x)$ is decreasing on $x \in [0, \pi/2]$, (150) implies

$$\wp(R, D, \rho_X\sigma^2, \rho_Y\sigma^2) \geq \wp(R, D, \overline{\rho}\sigma^2, \overline{\rho}\sigma^2), \tag{151}$$

which proves that (147) can not increase if ρ_X and ρ_Y are both replaced by their average $\overline{\rho}$. The observation that

$$2\mathbf{E}_Z(\rho) + \wp(R, D, \rho\sigma^2, \rho\sigma^2) \tag{152}$$

is monotone increasing for $\rho > 1$ completes the proof. □

E. General Sources and the Extremal Property of the Gaussian

The proof of Theorem 3 can be accomplished by restricting our attention to the setting where X and Y are discrete random variables. Therefore, the usual typicality machinery will be useful to us, and we review a few facts before beginning the proof of Theorem 3. We should also note that the method of types is used in the proofs in [1], but the proof here, which is similar in spirit, is significantly simpler and shorter, partially because we are only interested in the achievable rate (and not in the exponent). To this end, let $\mathcal{T}_\epsilon^{(n)}$ denote the usual ϵ-typical set (see [30, Ch. 2]). That is, we define the empirical pmf of $\mathbf{w} \in \mathcal{W}^n$ as

$$\pi(w|\mathbf{w}) = \frac{|i : w_i = w|}{n} \quad \text{for } w \in \mathcal{W}, \tag{153}$$

and, for $W \sim P_W$, the set of ϵ-typical n-sequences is defined by

$$\mathcal{T}_\epsilon^{(n)}(W) = \{\mathbf{w} : |\pi(w|\mathbf{w}) - P_W(w)| \leq \epsilon P_W(w) \text{ for all } w \in \mathcal{W}\}. \tag{154}$$

Observe that if $\mathbf{W} \sim \prod_{i=1}^n P_W(w_i)$, then the union of events bound and Hoeffding's inequality imply

$$\Pr\left\{\mathbf{W} \notin \mathcal{T}_\epsilon^{(n)}(W)\right\} \leq \sum_{w\in\mathcal{W}} \Pr\{|\pi(w|\mathbf{W}) - P_W(w)| > \epsilon P_W(w)\} \tag{155}$$

$$\leq \sum_{\substack{w\in\mathcal{W}:\\ P_W(w)>0}} 2\exp\left(-n(\epsilon P_W(w))^2\right). \tag{156}$$

Therefore, if $|\mathcal{W}| < \infty$,

$$\Pr\left\{\mathbf{W} \notin \mathcal{T}_\epsilon^{(n)}(W)\right\} \le \exp\left(-n\delta(\epsilon)\right), \tag{157}$$

where $\delta(\epsilon)$ denotes a positive quantity satisfying $\lim_{\epsilon\to 0}\delta(\epsilon) = 0$.

One useful fact is the so-called *Typical Average Lemma* [30, Sec. 2.4]:

Lemma 7 (Typical Average Lemma): If $\mathbf{w} \in \mathcal{T}_\epsilon^{(n)}(W)$*, then*

$$(1-\epsilon)\mathbb{E}[f(W)] \le \frac{1}{n}\sum_{i=1}^n g(w_i) \le (1+\epsilon)\mathbb{E}[f(W)]$$

for any nonnegative function $f(w)$ *on* $\mathcal{W}$.

Now, we state a simple variant of the Covering Lemma [30, Lemma 3.3]:

Lemma 8: Let P_{WV} *be a joint probability distribution on the finite alphabet* $\mathcal{W}\times\mathcal{V}$*, with corresponding marginals* P_W *and* P_V*. Let* $\mathbf{W} \sim \prod_{i=1}^n P_W(w_i)$ *and let* $\mathbf{V}(m)$*,* $m \in \{1, 2, \ldots, 2^{nR}\}$*, be random sequences, independent of each other and of* $\mathbf{W}$*, each distributed according to* $\prod_{i=1}^n P_V(v_i)$*. Then, for* n *sufficiently large, there exists positive functions* $\delta(\epsilon), \tilde{\delta}(\epsilon)$ *satisfying* $\lim_{\epsilon\to 0}\delta(\epsilon) = \lim_{\epsilon\to 0}\tilde{\delta}(\epsilon) = 0$ *and*

$$\Pr\left\{(\mathbf{W}, \mathbf{V}(m)) \notin \mathcal{T}_\epsilon^{(n)}(W, V) \text{ for all } m\right\} \le \exp\left(-n\delta(\epsilon)\right) + \exp\left(-2^{n(R - I(W;V) - \tilde{\delta}(\epsilon))}\right).$$

Proof: The proof follows that of [30, Lemma 3.3] verbatim, invoking (157) where appropriate. □

We require one more result before moving on to the proof of Theorem 3.

Lemma 9: Let P_W *and* P_V *be probability distributions with finite second moments* σ_W^2 *and* σ_V^2*, respectively. If* $\mathbf{w} \in \mathcal{T}_\epsilon^{(n)}(W)$*,* $\mathbf{v} \in \mathcal{T}_\epsilon^{(n)}(V)$*, and* $\frac{1}{n}\|\mathbf{w}-\mathbf{v}\|^2 \le D$*, then*

$$\frac{1}{n}\left\|\sqrt{\frac{\sigma_V}{\sigma_W}}\mathbf{w} - \sqrt{\frac{\sigma_W}{\sigma_V}}\mathbf{v}\right\|^2 \le D - (\sigma_W - \sigma_V)^2 + \epsilon|\sigma_W^2 - \sigma_V^2|. \tag{158}$$

Proof: Without loss of generality, assume $\sigma_V \ge \sigma_W$. Note that the assumption $\frac{1}{n}\|\mathbf{w}-\mathbf{v}\|^2 \le D$ implies

$$-\frac{2}{n}\mathbf{w}^T\mathbf{v} \le D - \frac{1}{n}\|\mathbf{w}\|^2 - \frac{1}{n}\|\mathbf{v}\|^2. \tag{159}$$

Moreover, Lemma 7 implies the following inequalities

$$\frac{1}{n}\|\mathbf{w}\|^2 \le (1+\epsilon)\sigma_W^2 \tag{160}$$

$$\frac{1}{n}\|\mathbf{v}\|^2 \ge (1-\epsilon)\sigma_V^2. \tag{161}$$

Therefore, it follows that

$$\begin{aligned}
&\frac{1}{n}\left\|\sqrt{\frac{\sigma_V}{\sigma_W}}\mathbf{w} - \sqrt{\frac{\sigma_W}{\sigma_V}}\mathbf{v}\right\|^2 \\
&= \frac{1}{n}\left(\frac{\sigma_V}{\sigma_W}\|\mathbf{w}\|^2 + \frac{\sigma_W}{\sigma_V}\|\mathbf{v}\|^2 - 2\mathbf{w}^T\mathbf{v}\right) \\
&\le D + \frac{1}{n}\left(\left(\frac{\sigma_V}{\sigma_W} - 1\right)\|\mathbf{w}\|^2 + \left(\frac{\sigma_W}{\sigma_V} - 1\right)\|\mathbf{v}\|^2\right) \\
&\le D + (1+\epsilon)\sigma_W^2\left(\frac{\sigma_V}{\sigma_W} - 1\right) + (1-\epsilon)\sigma_V^2\left(\frac{\sigma_W}{\sigma_V} - 1\right) \\
&= D - (\sigma_W - \sigma_V)^2 + \epsilon(\sigma_V^2 - \sigma_W^2).
\end{aligned}$$

Considering the symmetric case where $\sigma_V \le \sigma_W$ gives

$$\frac{1}{n}\left\|\sqrt{\frac{\sigma_V}{\sigma_W}}\mathbf{w} - \sqrt{\frac{\sigma_W}{\sigma_V}}\mathbf{v}\right\|^2 \le D - (\sigma_W - \sigma_V)^2 + \epsilon(\sigma_W^2 - \sigma_V^2),$$

completing the proof. □

Proof of Theorem 3: We can assume that X and Y are discrete random variables with finite alphabet $\mathcal{X} \subset \mathbb{R}$. The extension to continuous distributions with finite second moments follows by the usual quantization arguments and continuity of $\|\cdot\|$. Fix $\epsilon > 0$ and a conditional pmf $P_{\hat{X}|X}(\hat{x}|x)$, where the alphabet $\hat{\mathcal{X}}$ is an arbitrary subset of $\mathbb{R}$ with finite support. Throughout, the random variables $(Y, X, \hat{X})$ are drawn according to the joint distribution

$$\begin{aligned}
P_{YX\hat{X}}(y, x, \hat{x}) &= P_Y(y)P_{X\hat{X}}(x, \hat{x}) \\
&= P_Y(y)P_X(x)P_{\hat{X}|X}(\hat{x}|x).
\end{aligned}$$

Random Signature Assignment: Randomly and independently generate 2^{nR} sequences $\hat{\mathbf{x}}(t), t \in \{1, 2, \ldots, 2^{nR}\}$, each according to $\prod_{i=1}^n P_{\hat{X}}(\hat{x}_i)$. Given a sequence $\mathbf{x}$, find an index t such that $(\mathbf{x}, \hat{\mathbf{x}}(t)) \in \mathcal{T}_\epsilon^{(n)}(X, \hat{X})$ and put $T(\mathbf{x}) = t$. If there is more than one such index, break ties arbitrarily. If there is no such index, put $T(\mathbf{x}) = \mathtt{e}$. Observe that the rate R is negligibly affected by the addition of the additional "erasure" signature $\mathtt{e}$ (as in the proofs of Theorems 1 and 2).

Definition of the Query Function: In order to simplify notation, define the quantity

$$\Psi \triangleq \sqrt{(1+\epsilon)\mathbb{E}\left[\left(\sqrt{\frac{\sigma_Y}{\sigma_X}}X - \hat{X}\right)^2\right]} + \sqrt{D - (\sigma_X - \sigma_Y)^2 + \epsilon|\sigma_X^2 - \sigma_Y^2|}. \tag{162}$$

For a signature $t \in \{1, 2, \ldots, 2^{nR}\} \cup \{\mathtt{e}\}$ and a sequence $\mathbf{y}$, define

$$g(t, \mathbf{y}) \triangleq \begin{cases} \mathtt{maybe} & \text{if } \begin{cases} \mathbf{y} \notin \mathcal{T}_\epsilon^{(n)}(Y), \text{ or} \\ t = \mathtt{e}, \text{ or} \\ \frac{1}{\sqrt{n}}\left\|\sqrt{\frac{\sigma_X}{\sigma_Y}}\mathbf{y} - \hat{\mathbf{x}}(t)\right\| \le \Psi \text{ and } t \ne \mathtt{e} \end{cases} \\ \mathtt{no} & \text{otherwise.} \end{cases}$$

Scheme Analysis: First, we check to ensure that $g(\cdot,\cdot)$ does not produce any false negatives; that is, we need to verify that (T, g) is D-admissible. Note that $g(T(\mathbf{x}), \mathbf{y})$ returns `maybe` if $\mathbf{y} \notin \mathcal{T}_\epsilon^{(n)}(Y)$ or $T(\mathbf{x}) = \mathtt{e}$. Therefore, we only need to show that $g(T(\mathbf{x}), \mathbf{y})$ returns `maybe` if $\mathbf{y} \in \mathcal{T}_\epsilon^{(n)}(Y)$, $(\mathbf{x}, \hat{\mathbf{x}}(T(\mathbf{x}))) \in \mathcal{T}_\epsilon^{(n)}(X, \hat{X})$, and $\frac{1}{n}\|\mathbf{x}-\mathbf{y}\|^2 \le D$.

Under these assumptions, note that Lemma 7 implies

$$\frac{1}{n}\left\|\sqrt{\frac{\sigma_Y}{\sigma_X}}\mathbf{x} - \hat{\mathbf{x}}(t)\right\|^2 \le (1+\epsilon)\mathbb{E}\left[\left(\sqrt{\frac{\sigma_Y}{\sigma_X}}X - \hat{X}\right)^2\right]. \tag{163}$$

Next, recall that $(\mathbf{x}, \hat{\mathbf{x}}(T(\mathbf{x}))) \in \mathcal{T}_\epsilon^{(n)}(X, \hat{X})$ implies $\mathbf{x} \in \mathcal{T}_\epsilon^{(n)}(X)$. Hence, under the assumption that

$\frac{1}{n}\|\mathbf{x}-\mathbf{y}\|^2 \le D$, Lemma 9 implies

$$\frac{1}{n}\left\|\sqrt{\frac{\sigma_Y}{\sigma_X}}\mathbf{x}-\sqrt{\frac{\sigma_X}{\sigma_Y}}\mathbf{y}\right\|^2 \le D-(\sigma_X-\sigma_Y)^2+\epsilon|\sigma_X^2-\sigma_Y^2|. \tag{164}$$

Combining the triangle inequality, (163), (164), and (162), we have

$$\begin{aligned}\frac{1}{\sqrt{n}}&\left\|\sqrt{\frac{\sigma_X}{\sigma_Y}}\mathbf{y}-\hat{\mathbf{x}}(t)\right\| \\ &\le \frac{1}{\sqrt{n}}\left\|\sqrt{\frac{\sigma_Y}{\sigma_X}}\mathbf{x}-\hat{\mathbf{x}}(t)\right\|+\frac{1}{\sqrt{n}}\left\|\sqrt{\frac{\sigma_Y}{\sigma_X}}\mathbf{x}-\sqrt{\frac{\sigma_X}{\sigma_Y}}\mathbf{y}\right\| \\ &\le \Psi.\end{aligned}$$

Hence, $g(T(\mathbf{x}),\mathbf{y})$ returns `maybe` if $\mathbf{y} \in \mathcal{T}_\epsilon^{(n)}(Y)$, $(\mathbf{x},\hat{\mathbf{x}}(T(\mathbf{x}))) \in \mathcal{T}_\epsilon^{(n)}(X,\hat{X})$, and $\frac{1}{n}\|\mathbf{x}-\mathbf{y}\|^2 \le D$. Therefore, (T, g) is D-admissible as desired.

Next, we check to ensure that $\Pr\{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}\}$ is small. To this end, consider the events

$$\begin{aligned}\mathcal{E}_0 &= \left\{\mathbf{y} \notin \mathcal{T}_\epsilon^{(n)}(Y)\right\}, \\ \mathcal{E}_1 &= \{T(\mathbf{X}) = \mathrm{e}\}, \\ \mathcal{E}_2 &= \left\{\frac{1}{\sqrt{n}}\left\|\sqrt{\frac{\sigma_X}{\sigma_Y}}\mathbf{Y}-\hat{\mathbf{X}}(T(\mathbf{X}))\right\| \le \Psi\right\},\end{aligned}$$

and observe that $\Pr\{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}\} \le \Pr\{\mathcal{E}_0\} + \Pr\{\mathcal{E}_1\} + \Pr\{\mathcal{E}_2\}$ by the union of events bound.

We have already seen in (157) that

$$\Pr\{\mathcal{E}_0\} \le \exp(-n\delta(\epsilon)) \tag{165}$$

for some positive $\delta(\epsilon)$ satisfying $\lim_{\epsilon\to 0}\delta(\epsilon) = 0$.

Next, Lemma 8 implies that, for n sufficiently large,

$$\mathbb{E}_T[\Pr\{\mathcal{E}_1\}] \le \exp(-n\delta(\epsilon)) + \exp\left(-2^{n(R-I(X;\hat{X})-\tilde{\delta}(\epsilon))}\right), \tag{166}$$

where $\mathbb{E}_T[\Pr\{\mathcal{E}_1\}]$ denotes the value of $\Pr(\mathcal{E}_1)$ averaged over the random choice of the signature assignment $T(\cdot)$.

Let $\hat{\mathbf{X}}$ be distributed according to $\prod_{i=1}^n P_{\hat{X}}(\hat{x}_i)$, independent of $\mathbf{Y} \sim \prod_{i=1}^n P_Y(y_i)$. An application of Hoeffding's inequality implies

$$\Pr\left(\frac{1}{\sqrt{n}}\left\|\sqrt{\frac{\sigma_X}{\sigma_Y}}\mathbf{Y}-\hat{\mathbf{X}}\right\| \le \sqrt{\mathbb{E}\left[\left(\sqrt{\frac{\sigma_X}{\sigma_Y}}Y-\hat{X}\right)^2\right]}-\epsilon\right) \le \exp(-n\delta(\epsilon)). \tag{167}$$

Since the sequence $\mathbf{Y}$ is independent of $\mathbf{X}$, and is therefore also independent of $\hat{\mathbf{X}}(T(\mathbf{X}))$, (167) implies that

$$\mathbb{E}_T[\Pr\{\mathcal{E}_2\}] \le \exp(-n\delta(\epsilon)) \tag{168}$$

if

$$\Psi \le \sqrt{\mathbb{E}\left[\left(\sqrt{\frac{\sigma_X}{\sigma_Y}}Y-\hat{X}\right)^2\right]}-\epsilon. \tag{169}$$

Therefore, if (169) holds, we have

$$\begin{aligned}&\mathbb{E}_T[\Pr\{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}\}] \\ &\quad\le 3\exp(-n\delta(\epsilon)) + \exp\left(-2^{n(R-I(X;\hat{X})-\tilde{\delta}(\epsilon))}\right),\end{aligned} \tag{170}$$

implying the existence of a sequence of D-admissible, rate $R > I(X;\hat{X})$ schemes for which $\Pr\{g(T(\mathbf{X}),\mathbf{Y}) = \texttt{maybe}\}$ is exponentially small in n. Since ϵ was arbitrary, the proof is complete. □

Proof of Theorem 4: Since $d(\cdot,\cdot)$ is translation invariant, we can assume without loss of generality that P_X and P_Y have mean zero. Also, note that it is sufficient to consider D in the interval $(\sigma_X-\sigma_Y)^2 \le D \le \sigma_X^2+\sigma_Y^2$. Indeed, if $D > \sigma_X^2+\sigma_Y^2$, then the theorem asserts that $R_{\mathrm{ID}}(D, P_X, P_Y) \le \infty$, which is trivially true. The other case of $D < (\sigma_X-\sigma_Y)^2$ will follow from the monotonicity of $R_{\mathrm{ID}}(D, P_X, P_Y)$ in D and from the fact that $R_{\mathrm{ID}}((\sigma_X-\sigma_Y)^2, P_X, P_Y) = 0$ (which is proved in the main case).

Therefore, assume $(\sigma_X-\sigma_Y)^2 < D \le \sigma_X^2+\sigma_Y^2$ and consider the conditional distribution $P_{\hat{X}|X}$ defined by $\hat{X} = \rho\sqrt{\frac{\sigma_Y}{\sigma_X}}X+Z$, where $Z \sim N(0,\sigma_Z^2)$ is independent of X and ρ, σ_Z^2 are given by

$$\begin{aligned}\rho &= \frac{(\sigma_X+\sigma_Y)^2-D}{(2\sigma_X\sigma_Y)} \\ \sigma_Z^2 &= \frac{((\sigma_X+\sigma_Y)^2-D)(\sigma_X^2+\sigma_Y^2-D)^2}{4\sigma_X\sigma_Y(D-(\sigma_X-\sigma_Y)^2)}.\end{aligned}$$

With $P_{\hat{X}|X}$ defined in this way, the following identities are readily verified

$$\begin{aligned}\sqrt{\mathbb{E}\left[\left(\sqrt{\frac{\sigma_X}{\sigma_Y}}Y-\hat{X}\right)^2\right]} &= \sqrt{\sigma_X\sigma_Y(1+\rho^2)+\sigma_Z^2} \\ &= \frac{2\sigma_X\sigma_Y}{\sqrt{D-(\sigma_X-\sigma_Y)^2}}\end{aligned} \tag{171}$$

$$\begin{aligned}\sqrt{\mathbb{E}\left[\left(\sqrt{\frac{\sigma_Y}{\sigma_X}}X-\hat{X}\right)^2\right]} &= \sqrt{\sigma_X\sigma_Y(1-\rho)^2+\sigma_Z^2} \\ &= \frac{\sigma_X^2+\sigma_Y^2-D}{\sqrt{D-(\sigma_X-\sigma_Y)^2}}.\end{aligned} \tag{172}$$

Therefore, (171) and (172) yield the identity

$$\sqrt{\mathbb{E}\left[\left(\sqrt{\frac{\sigma_X}{\sigma_Y}}Y-\hat{X}\right)^2\right]} = \sqrt{\mathbb{E}\left[\left(\sqrt{\frac{\sigma_Y}{\sigma_X}}X-\hat{X}\right)^2\right]} + \sqrt{D-(\sigma_X-\sigma_Y)^2}.$$

Since $\hat{X}$ has density and the Gaussian distribution maximizes differential entropy for a given variance (see [17]), we have the inequality $h(\hat{X}) \le \frac{1}{2}\log\left(2\pi e(\rho^2\sigma_X\sigma_Y+\sigma_Z^2)\right)$. It follows that

$$\begin{aligned}I(X;\hat{X}) &\le \frac{1}{2}\log\left(\frac{\rho^2\sigma_X\sigma_Y+\sigma_Z^2}{\sigma_Z^2}\right) \\ &= \log\left(\frac{2\sigma_X\sigma_Y}{\sigma_X^2+\sigma_Y^2-D}\right) \\ &= R_{\mathrm{ID}}(D, N(0,\sigma_X^2), N(0,\sigma_Y^2)).\end{aligned}$$

Thus, for $D \ne (\sigma_X-\sigma_Y)^2$, an application of Theorem 3 implies that

$$R_{\mathrm{ID}}(D, P_X, P_Y) \le R_{\mathrm{ID}}(D, N(0,\sigma_X^2), N(0,\sigma_Y^2)). \tag{173}$$

To handle the case where $D = (\sigma_X - \sigma_Y)^2$, we note that $R_{\text{ID}}(D, P_X, P_Y)$ is nondecreasing in D. Since

$$\lim_{D\downarrow(\sigma_X-\sigma_Y)^2} R_{\text{ID}}(D, N(0,\sigma_X^2), N(0,\sigma_Y^2)) = 0, \quad (174)$$

inequality (173) implies that we must have $R_{\text{ID}}(D, P_X, P_Y) = 0$ at $D = (\sigma_X - \sigma_Y)^2$. This completes the proof. □

F. Robust Identification Schemes

Fix $R > R_{\text{ID}}\left(D, N(0,\sigma_X^2), N(0,\sigma_Y^2)\right)$ and consider the setup described in section III-D. Specifically, let $P_{\tilde{X}}$, $P_{\tilde{Y}}$ be zero-mean distributions with variances σ_X^2 and σ_Y^2, respectively. Recall that, for a given blocklength n, the argument in the achievability proof of Theorem 1 constructs a signature assignment function $T^{(n)} : \tilde{\mathbf{x}} \to \mathbb{R}^n$ for which the query $g^{(n)}\left(T^{(n)}(\tilde{\mathbf{x}}), \tilde{\mathbf{y}}\right)$ returns "maybe" only if

1) The angle $\angle(\tilde{\mathbf{y}}, T^{(n)}(\tilde{\mathbf{x}}))$ is at most θ', where $\theta' < \pi/2$ is a fixed constant defined in (68), and
2) We have $\tilde{\mathbf{x}} \in S_X^{\text{typ}}$, where S_X^{typ} is the "typical shell" of $\tilde{\mathbf{x}}$ vectors defined in (57).

We remark that the gap between $\pi/2 - \theta'$ and the thickness of the shell S_X^{typ} depend on the parameter $\eta > 0$, which is a function of the gap between R and $R_{\text{ID}}\left(D, N(0,\sigma_X^2), N(0,\sigma_Y^2)\right)$.

In light of the conditions under which $g^{(n)}\left(T^{(n)}(\tilde{\mathbf{x}}), \tilde{\mathbf{y}}\right)$ returns "maybe", the probability of the event $\left\{g^{(n)}\left(T^{(n)}(\tilde{\mathbf{X}}), \tilde{\mathbf{Y}}\right) = \texttt{maybe}\right\}$ is bounded by

$$\Pr\left\{g^{(n)}\left(T^{(n)}(\tilde{\mathbf{X}}), \tilde{\mathbf{Y}}\right) = \texttt{maybe}\right\} \le \Pr\left\{\angle(\tilde{\mathbf{Y}}, T^{(n)}(\tilde{\mathbf{X}})) \le \theta'\right\} + \Pr\left\{\tilde{\mathbf{X}} \notin S_X^{\text{typ}}\right\}.$$

The term $\Pr\left\{\tilde{\mathbf{X}} \notin S_X^{\text{typ}}\right\}$ vanishes by the weak law of large numbers as $n \to \infty$. Therefore, since $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{Y}}$ are independent, it is sufficient to show that $\Pr\left\{\angle(\tilde{\mathbf{Y}}, \boldsymbol{\alpha}) \le \theta'\right\}$ vanishes for any given unit vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ and constant $\theta' \in (0, \pi/2)$. To this end, define $\beta_n \triangleq \frac{\sigma_Y}{2}\sqrt{n}$, and observe that

$$\begin{aligned}\Pr\left\{\angle(\tilde{\mathbf{Y}}, \boldsymbol{\alpha}) \le \theta'\right\} &= \Pr\left\{\sum_{i=1}^n \alpha_i \tilde{Y}_i \ge \|\tilde{\mathbf{Y}}\| \cos\theta'\right\} \\ &\le \Pr\left\{\sum_{i=1}^n \alpha_i \tilde{Y}_i \ge \beta_n \cos\theta'\right\} \\ &\quad + \Pr\left\{\|\tilde{\mathbf{Y}}\| \le \beta_n\right\}.\end{aligned}$$

First, note $\lim_{n\to\infty} \Pr\left\{\|\tilde{\mathbf{Y}}\| \le \beta_n\right\} = 0$ by the weak law of large numbers. Next, since $\boldsymbol{\alpha}$ is a unit vector, we have $\sum_{i=1}^n \alpha_i^2 = 1$, and it follows that

$$VAR\left(\sum_{i=1}^n \alpha_i \tilde{Y}_i\right) = \sigma_Y^2. \quad (175)$$

Since $\mathbb{E}[\tilde{Y}_i] = 0$, Chebyshev's inequality implies

$$\Pr\left\{\sum_{i=1}^n \alpha_i \tilde{Y}_i \ge \beta_n \cos\theta'\right\} \le \frac{\sigma_Y^2}{\beta_n^2 \cos^2\theta'} = \frac{4}{n\cos^2\theta'}, \quad (176)$$

proving that $\Pr\left\{g^{(n)}\left(T^{(n)}(\tilde{\mathbf{X}}), \tilde{\mathbf{Y}}\right) = \texttt{maybe}\right\} \to 0$ as desired. Since the D-admissibility of the scheme $(T^{(n)}, g^{(n)})$ did not depend on the Gaussianity assumption in the proof of Theorem 1, the scheme $(T^{(n)}, g^{(n)})$ continues to be D-admissible for the sources $\tilde{\mathbf{X}}, \tilde{\mathbf{Y}}$.

Therefore, we can conclude that a sequence of rate-R, D-admissible schemes $\{T^{(n)}, g^{(n)}\}_{n\to\infty}$ constructed as described in the proof of Theorem 1 exhibit the robustness property explained in Section III-D.

V. Concluding Remarks

We studied the problem of answering similarity queries from compressed data from an information-theoretic perspective. We focused on the setting where the similarity criterion is the (normalized) quadratic distance. For the case of i.i.d. Gaussian data, we gave an explicit characterization of the minimal compression rate which permits reliable queries (i.e., the identification rate). Furthermore, we characterized the best exponential rate at which the probability for false positives can be made to vanish.

For general sources, we derived an upper bound on the identification rate, and proved that it is at most that of the Gaussian source of the same variance. Finally, we presented a single, robust, scheme that compresses *any* source at the Gaussian identification rate, while permitting reliable responses to queries.

The identification rate and exponent, studied in this paper, are asymptotic in nature, i.e. correspond to the case of infinite (or growing) dimension n. Extensions of this work to the non-asymptotic case, as well as constructive achievability schemes, can be found in [31].

Appendix A
Covering a Shell With Spheres

Proof of Lemma 3: According to [26, Th. 1], for any $r > \rho > 0$ there exists a covering of S_r with balls of radius ρ, with density ϑ upper bounded by

$$\vartheta \le (n-1)\log(n-1)\left(\frac{1}{2} + \frac{2\log\log(n-1)+5}{\log(n-1)}\right) \quad (177)$$
$$\le n\log(n), \quad (178)$$

where (177) holds for all $n \ge 4$, and (178) holds for n large enough so that $\frac{2\log\log(n-1)+5}{\log(n-1)} \le \frac{1}{2}$. This translates to k balls of radius ρ that cover S_r, where

$$k \le \frac{n\log(n)}{\Omega(\theta)}, \quad (179)$$

and $\theta \triangleq \arcsin(\rho/r)$.

We choose $r = r_0 = \sqrt{n\sigma^2}$ and $\rho = \rho_0 = \sqrt{nD_0}$, so $\theta = \theta_0 = \arcsin(\sqrt{D_0/\sigma^2}) < \pi/2$ and is independent of n. When n is large enough s.t. $\theta \le \arccos(1/\sqrt{n})$, we may use (42) and get an upper bound on the covering size:

$$k \leq \frac{n \log(n)}{\Omega(\theta_0)} \tag{180}$$
$$\leq \frac{n \log(n)}{\frac{1}{3\sqrt{2\pi n}\cos\theta_0} \sin^{n-1}\theta_0} \tag{181}$$
$$\leq 3\sqrt{2\pi} n^{3/2} \log(n)(\rho_0/r_0)^{n-1}, \tag{182}$$

which proves (48).

Note that for a code that covers a spherical shell, the biggest covering by any single point $\mathbf{u} \in \mathcal{C}$ is obtained when the point $\mathbf{u}$ is located at distance $\sqrt{r_0^2 - \rho_0^2}$ from the origin. We therefore can assume, without altering the covering property of $\mathcal{C}$, that $\|\mathbf{u}\| = \sqrt{r_0^2 - \rho_0^2}$ for all $\mathbf{u} \in \mathcal{C}$ (see also [26, eq. (1)] and the discussion that follows). The intersection of $\text{BALL}_{\rho_0}(\mathbf{u})$ and S_{r_0} is precisely given by $\text{CAP}_{r_0}(\mathbf{u}, \theta_0)$. □

Appendix B

Proof of Lemma 4: Let $\mathbf{y} \in \Gamma^D\left(T^{-1}(\mathbf{u})\right) \cap S_Y^{\text{typ}}$. Our goal is to show that $\mathbf{y} \in \text{CONE}(\mathbf{u}, \theta')$. In other words, we need to show that

$$\angle(\mathbf{u}, \mathbf{y}) \leq \theta'. \tag{183}$$

Since $\mathbf{y} \in \Gamma^D T^{-1}(\mathbf{u})$, there exists $\mathbf{x} \in T^{-1}(\mathbf{u})$ s.t. $d(\mathbf{x}, \mathbf{y}) \leq D$. By the triangle inequality for the angle operator (which is proportional to the geodesic metric in spherical geometry) we can write

$$\angle(\mathbf{u}, \mathbf{y}) \leq \angle(\mathbf{u}, \mathbf{x}) + \angle(\mathbf{x}, \mathbf{y}). \tag{184}$$

Since $T^{-1}(\mathbf{u}) \subseteq \text{CAP}_{r^-,r^+}(\mathbf{u}, \theta_0)$, we know that $\angle(\mathbf{u}, \mathbf{x}) \leq \theta_0$. Further, by the law of cosines for the triangle $(\mathbf{x}, \mathbf{y}, \mathbf{0})$ we can write

$$\angle(\mathbf{x}, \mathbf{y}) = \arccos\left[\frac{\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - \|\mathbf{x}-\mathbf{y}\|^2}{2\|\mathbf{x}\|\|\mathbf{y}\|}\right] \tag{185}$$
$$\overset{(a)}{\leq} \arccos\left[\frac{\sigma_X^2 + \sigma_Y^2 - 2\eta - nD}{2\sqrt{(\sigma_X^2 + \eta)(\sigma_Y^2 + \eta)}}\right] \tag{186}$$
$$= \theta_1, \tag{187}$$

where (a) follows since $\mathbf{x} \in S_X^{\text{typ}}$, $\mathbf{y} \in S_Y^{\text{typ}}$ and $d(\mathbf{x}, \mathbf{y}) \leq D$. Therefore by definition we have $\mathbf{y} \in \text{CONE}(\mathbf{u}, \theta')$. All there's left to show is that $\theta' < \frac{\pi}{2}$. This follows immediately since D_0 satisfies (52) by definition, and from the fact that $\arcsin(\phi) + \arccos(\phi) = \frac{\pi}{2}$. □

Appendix C

Proof of Lemma 5: Define $\mathcal{I}$ as the set of indices i for which $p_i \geq p^*$:

$$\mathcal{I} \triangleq \{i : p_i \geq p^*\}. \tag{188}$$

Clearly $\Omega\left(\theta_{D''} + \Omega^{-1}(p_i)\right) \geq \Omega^*$ if and only if $i \in \mathcal{I}$, so $\mathcal{I}$ can be thought of as the set of 'bad' values for i, i.e. those that contribute a lot to the sum in (85).

Consider the following sequence of inequalities:

$$\begin{aligned} c \cdot \Omega^* &\geq \sum_{i=1}^{2^{nR}} p_i \cdot \Omega\left(\theta_{D''} + \Omega^{-1}(p_i)\right) \\ &\geq \sum_{i \in \mathcal{I}} p_i \cdot \Omega\left(\theta_{D''} + \Omega^{-1}(p_i)\right) \\ &\geq \Omega^* \sum_{i \in \mathcal{I}} p_i. \end{aligned} \tag{189}$$

On the other hand,

$$\begin{aligned} 1 &= \sum_i p_i \\ &= \sum_{i \in \mathcal{I}} p_i + \sum_{i \notin \mathcal{I}} p_i \\ &\overset{(a)}{\leq} c + \sum_{i \notin \mathcal{I}} p_i \\ &\overset{(b)}{\leq} c + \sum_{i \notin \mathcal{I}} p^* \\ &\leq c + p^* 2^{nR}. \end{aligned}$$

(a) follows from (189). (b) follows from the definition of $\mathcal{I}$. Eq. (86) follows immediately. □

Appendix D

Proof of Lemma 6: For any $t > 0$ and $a > 0$ we have

$$\begin{aligned} \Pr\{\|\mathbf{X}\|^2 > a\} &\overset{(a)}{\leq} e^{-t\cdot a}\mathbb{E}\left[\exp\left(t \cdot \|\mathbf{X}\|^2\right)\right] \\ &\overset{(b)}{=} e^{-t\cdot a}(1 - 2t\sigma_X^2)^{-n/2}. \end{aligned}$$

(a) is the Chernoff bound. (b) follows since the moment generating function of Z is given by

$$\mathbb{E}\left[e^{tZ}\right] = (1-2t)^{-n/2}, \quad \text{for } t < \frac{1}{2}. \tag{190}$$

Here it holds for any $t < \frac{1}{2\sigma_X^2}$. We choose $t = \frac{1}{4\sigma_X^2}$ and write:

$$\Pr\{\|\mathbf{X}\|^2 > a\} \leq e^{-\frac{a}{4\sigma_X^2}} \cdot 2^{n/2}. \tag{191}$$

Choosing $a = n\sigma_{\max}^2(n) = n^2\sigma_X^2$ results in

$$\Pr\{\|\mathbf{X}\|^2 > n\sigma_{\max}^2(n)\} \leq e^{-\frac{1}{4}n^2 + o(n^2)}. \tag{192}$$

□

Acknowledgement

The authors would like to thank Golan Yona for stimulating discussions that motivated this work, and to Fabian Steiner, Steffen Dempfle and the anonymous reviewers for helpful comments that improved the quality of this paper.

References

[1] R. Ahlswede, E.-H. Yang, and Z. Zhang, "Identification via compressed data," *IEEE Trans. Inf. Theory*, vol. 43, no. 1, pp. 48–70, Jan. 1997.

[2] E. Tuncel, P. Koulgi, and K. Rose, "Rate-distortion approach to databases: Storage and content-based retrieval," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 953–967, Jun. 2004.

[3] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.
[4] J. A. O'Sullivan and N. Schmid, "Large deviations performance analysis for biometrics recognition," in *Proc. Allerton Conf. Commun., Control, Comput.*, Oct. 2002, pp. 1–19.
[5] F. Willems, T. Kalker, J. Goseling, and J.-P. Linnartz, "On the capacity of a biometrical identification system," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2003, p. 82.
[6] M. B. Westover and J. A. O'Sullivan, "Achievable rates for pattern recognition," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 299–320, Jan. 2008.
[7] E. Tuncel, "Capacity/storage tradeoff in high-dimensional identification systems," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2097–2106, May 2009.
[8] E. Tuncel and D. Gündüz, "Identification and lossy reconstruction in noisy databases," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 822–831, Feb. 2014.
[9] B. H. Bloom, "Space/time trade-offs in hash coding with allowable errors," *Commun. ACM*, vol. 13, no. 7, pp. 422–426, Jul. 1970.
[10] E. Porat, "An optimal Bloom filter replacement based on matrix solving," in *Computer Science–Theory and Applications* (Lecture Notes in Computer Science), vol. 5675, A. Frid, A. Morozov, A. Rybalchenko, and K. W. Wagner, Eds. Berlin, Germany: Springer-Verlag, 2009, pp. 263–273.
[11] A. Broder and M. Mitzenmacher, "Network applications of Bloom filters: A survey," *Internet Math.*, vol. 1, no. 4, pp. 485–509, 2003.
[12] A. Andoni and P. Indyk, "Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions," *Commun. ACM*, vol. 51, no. 1, pp. 117–122, 2008.
[13] W. B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," in *Proc. Conf. Modern Anal. Probab.*, vol. 26. 1984, pp. 189–206.
[14] P. Indyk, *Sketching, Streaming and Sublinear-Space Algorithms*. Cambridge, MA, USA: MIT, 2007. [Online]. Available: http://stellar.mit.edu/S/course/6/fa07/6.895/
[15] P. T. Boufounos and S. Rane, "Efficient coding of signal distances using universal quantized embeddings," in *Proc. Data Compress. Conf. (DCC)*, Snowbird, UT, USA, Mar. 2013, pp. 251–260.
[16] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.
[17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York, NY, USA: Wiley, 2006.
[18] K. Marton, "Error exponent for source coding with a fidelity criterion," *IEEE Trans. Inf. Theory*, vol. 20, no. 2, pp. 197–199, Mar. 1974.
[19] S. Ihara and M. Kubo, "Error exponent for coding of memoryless Gaussian sources with a fidelity criterion," *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. 83-A, no. 10, pp. 1891–1897, 2000.
[20] S. Ihara and M. Kubo, "Error exponent of coding for stationary memoryless sources with a fidelity criterion," *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. E88-A, no. 5, pp. 1339–1345, May 2005.
[21] Y. Zhong, F. Alajaji, and L. L. Campbell, "A type covering lemma and the excess distortion exponent for coding memoryless Laplacian sources," in *Proc. 23rd Biennial Symp. Commun.*, 2006, pp. 100–103.
[22] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications* (Stochastic Modelling and Applied Probability), vol. 38, 2nd ed. New York, NY, USA: Springer-Verlag, 1998.
[23] A. Ingber and T. Weissman. (2013). "The minimal compression rate for similarity identification," Submitted to IEEE Trans. Inf. Theory. [Online]. Available: http://arxiv.org/abs/1312.2063
[24] K. Böröczky, Jr., and G. Wintsche, "Covering the sphere by equal spherical balls," in *Discrete and Computational Geometry: The Goodman-Pollack Festschrift*. Berlin, Germany: Springer-Verlag, 2003, pp. 235–251.
[25] M. Ledoux and M. Talagrand, *Probability in Banach Spaces: Isoperimetry and Processes* (Ergebnisse der Mathematik und Ihrer Grenzgebiete). New York, NY, USA: Springer-Verlag, 2011.
[26] I. Dumer, "Covering spheres with spheres," *Discrete Comput. Geometry*, vol. 38, no. 4, pp. 665–679, 2007.
[27] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[28] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 1041–1056, May 1998.
[29] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA, USA: Kluwer, 1992.
[30] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[31] F. Steiner, S. Dempfle, A. Ingber, and T. Weissman. (2014). "Compression for quadratic similarity queries: Finite blocklength and practical schemes," Submitted to IEEE Trans. Inf. Theory. [Online]. Available: http://arxiv.org/abs/1404.5173

Amir Ingber (S'08–M'12) received a B.Sc. in electrical engineering and computer science (summa cum laude), an M.Sc. in Electrical Engineering (summa cum laude), and a Ph.D. in electrical engineering, all from Tel-Aviv University, Israel, in 2005, 2006 and 2012 respectively. He is now a Senior Research Scientist at Yahoo! Labs, Sunnyvale, CA. Between 2011-2013 he was a postdoctoral scholar at the Electrical Engineering Department at Stanford University, Stanford, CA. Between 2002-2011 he served as a DSP and Algorithms Engineer and consultant at several companies, including Intel Corporation, Comsys Communication and Signal Processing, and Amimon Ltd.

Dr. Ingber was a recipient of the Best Student Paper Award at the 2011 IEEE International Symposium on Information Theory (ISIT), St. Petersburg, Russia. Among his other recent awards are the Rothschild Fellowship for postdoctoral studies (2011-2013), the Intel award for excellence in academic studies and research (2011), the Adams Fellowship (2008-2011) awarded by the Israel Academy of Sciences and Humanities, a Motorola scholarship in the area of Advanced Communication (2007), and the Weinstein Prize (2006, 2009) for research in the area of signal processing.

Thomas Courtade (S'06–M'13) is an Assistant Professor in the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley. Prior to joining UC Berkeley in 2014, he was a postdoctoral fellow supported by the NSF Center for Science of Information. He received his Ph.D. and M.S. degrees from UCLA in 2012 and 2008, respectively, and he graduated summa cum laude with a B.Sc. in Electrical Engineering from Michigan Technological University in 2007.

His honors include a Distinguished Ph.D. Dissertation award and an Excellence in Teaching award from the UCLA Department of Electrical Engineering, and a Jack Keil Wolf Student Paper Award for the 2012 International Symposium on Information Theory.

Tsachy Weissman (S'99–M'02–SM'07–F'13) is on the faculty of the Department of Electrical Engineering at Stanford University, where he holds the STMicroelectronics Chair in the school of Engineering. He received his B.Sc. and Ph.D. from the Technion in 1997 and 2001. He has published extensively on Information Theory, Statistical Signal Processing, the interplay between them and their applications. He is an inventor of several patents and is involved in a number of hi-tech companies as a member of the technical board. Much of his recent research has been dedicated to the theory and practice of genomic data compression. His research has been recognized with numerous awards, including best paper awards, a fellowship for Leaders in Science and Technology, and prizes for excellence in research. He serves on the editorial boards of the IEEE TRANSACTIONS ON INFORMATION THEORY, and *Foundations and Trends in Communications and Information Theory*.