PREDICTION-BASED SIMILARITY IDENTIFICATION FOR AUTOREGRESSIVE PROCESSES

Hanwei Wu, Qiwen Wang and Markus Flierl

School of Electrical Engineering and Computer Science
KTH Royal Institute of Technology, Stockholm
{hanwei, qiwenw, mflierl}@kth.se

ABSTRACT

The task of similarity identification is to identify items in a database which are similar to a given query item for a given metric. The *identification rate* of a compression scheme characterizes the minimum rate that can be achieved which guarantees reliable answers with respect to a given similarity threshold [1]. In this paper, we study a prediction-based quadratic similarity identification for autoregressive processes. We use an ideal linear predictor to remove linear dependencies in autoregressive processes. The similarity identification is conducted on the residuals. We show that the relation between the distortion of query and database processes and the distortion of their residuals is characterized by a sequence of eigenvalues. We derive the identification rate of our prediction-based approach for autoregressive Gaussian processes. We characterize the identification rate for the special case where only the smallest value in the sequence of eigenvalues is required to be known and derive its analytical upper bound by approximating a sequence of matrices with a sequence of Toeplitz matrices.

I. INTRODUCTION

The problem of similarity identification of compressed data with zero false negative errors is studied in [1] [2]. This problem is investigated from an information-theoretic viewpoint and introduces the term *identification rate* of the source. It characterizes the minimal compression rate that allows query answers with a vanishing false positive probability. [1] [2] also provide identification rates for i.i.d. Gaussian sources with quadratic distortion and for binary sources with Hamming distance. For similarity identification of correlated sources, [3] proposes tree-structured vector quantizers that hierarchically cluster the data using k-center clustering. In [4], the authors compare two transform-based similarity identification schemes to cope with exponentially growing codebooks for high-dimensional data. The component-based approach shows both good performance and low search complexity. In [5], the authors present an analysis of the identification rate of the component-based approach for correlated Gaussian sources.

For large time series databases, [6] proposes the Piecewise Aggregate Approximation (PAA) to perform dimensionality reduction on the data and to index the reduced data with a spatial access method. In this paper, we study the similarity identification of autoregressive processes using a prediction-based approach.

The paper is organized as follows: In Section 2, we give a brief review of the key concepts of the problem[1]. In Section 3, we discuss the similarity identification rate of our prediction-based approach for autoregressive Gaussian processes. In Section 4, we characterize the identification rate for a special case and derive its analytical upper bound. Numerical examples for an $AR(1)$ process are shown in Section 5, and the conclusion is given in Section 6.

II. QUADRATIC SIMILARITY QUERIES

Let $\mathbf{y} = (y_1, y_2, .., y_n)^T$ denote the query sequence and $\mathbf{x} = (x_1, x_2, .., x_n)^T$ the data sequence. A rate-R_{ID} identification system (T, g) consists of a *signature assignment* function $T : \mathbb{R}^n \to \{1, 2, \cdots, 2^{nR_{\text{ID}}}\}$ and a *query function* g:$\{1, 2, \cdots, 2^{nR_{\text{ID}}}\} \times \mathbb{R}^n \to \{\text{no}, \text{maybe}\}$.

The sequences $\mathbf{x}$ and $\mathbf{y}$ are called D_{ID}-*similar* if $d(\mathbf{x}, \mathbf{y}) \leq D_{\text{ID}}$, specifically, the quadratic similarity is

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

where $\|\cdot\|$ is the standard Euclidean norm. A similarity query retrieves all data items that are D_{ID}-similar. The database keeps only a short signature $T(\mathbf{x})$ for each $\mathbf{x}$. That is, given a query and the signatures, the identification system indicates whether the given query is D_{ID}-similar to an item in the database. A system is called D_{ID}-$admissible$ if we obtain the output $g\left(T(\mathbf{x}), \mathbf{y}\right) = \text{maybe}$ for any pair of data item and query $(\mathbf{x}, \mathbf{y})$ which is D_{ID}-*similar*.

Such D_{ID}-admissible systems do not allow false negative errors. Hence, $\Pr\{\text{maybe}\}$ represents only the false positive

[1]We follow the problem setup and adopt most notations in [1] and [2]. Therefore, we refer to [1] and [2] for more detailed background and problem description.

errors of the identification system. Further, for given distributions P_X and P_Y and a similarity threshold D_{ID}, a rate R is said to be D_{ID}-*achievable* if there exits a sequence of D_{ID}-admissible systems $\left(T^{(n)}, g^{(n)}\right)$ that can achieve a vanishing $\Pr\{\text{maybe}\}$ as n approaches infinity, i.e.

$$\lim_{n\to\infty} \Pr\left\{g^{(n)}\left(T^{(n)}(\mathbf{X}), \mathbf{Y}\right) = \text{maybe}\right\} = 0. \quad (2)$$

The *identification rate* R_{ID} of the source is defined as the infimum of all D_{ID}-*achievable* rates.

III. IDENTIFICATION RATE OF AUTOREGRESSIVE GAUSSIAN PROCESSES

A. Prediction-based Approach

We use a prediction-based approach to handle autoregressive processes for D_{ID}-similarity identification. The prediction-based approach exploits the linear dependencies among random variables by embedding a $\hat{D}_{\text{ID}}$-admissible scheme in an open-loop prediction scheme. The optimal linear predictors remove the linear dependencies in the stationary process such that its prediction residual process is a white noise process. If the data source is an autoregressive process, its prediction residual process is an i.i.d. process. The query is also handled by a linear predictor and the prediction residual is fed into the $\hat{D}_{\text{ID}}$-admissible scheme. Fig. 1 illustrates the described prediction-based scheme.

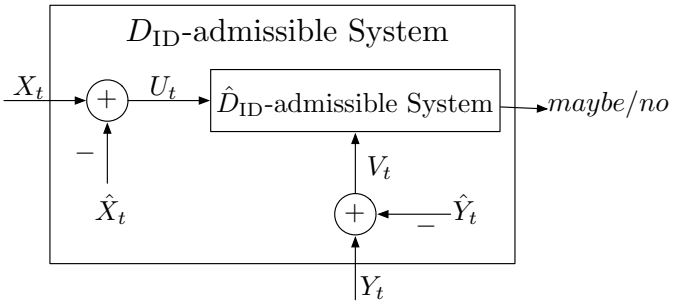

Fig. 1. Prediction-based approach for autoregressive sources. $\{X_t\}$ and $\{Y_t\}$ are the database and query processes, respectively. $\{\hat{X}_t\}$ and $\{\hat{Y}_t\}$ are the corresponding optimal predictors. $\{U_t\}$ and $\{V_t\}$ are the prediction residuals.

We first describe the linear prediction part of the scheme. Consider an autoregressive process of order m (AR(m)) and mean μ_S as

$$S_t = Z_t + \mu_S(1 - \mathbf{a}_m^T \mathbf{e}_m) + \mathbf{a}_m^T \mathbf{S}_{t-1}^{(m)}, \quad (3)$$

where $\{Z_t\}$ is a zero-mean i.i.d. process, $\mathbf{a}_m = (a_1, \ldots, a_m)^T$, $\mathbf{S}_{t-1}^{(m)} = (S_{t-1}, \ldots, S_{t-m})^T$, and $\mathbf{e}_m$ is an m-length vector with all entries being one.

For linear one-step prediction, we use N_p preceding samples $\mathbf{S}_{t-1}^{(N_p)}$ to predict the current random variable S_t, where N_p is greater than or equal to the order m. The predictor is then given by

$$\hat{S}_t = \mathbf{h}_{N_p}^T \mathbf{S}_{t-1}^{(N_p)}, \quad (4)$$

where the constant vector $\mathbf{h}_{N_p} = (h_1, \ldots, h_{N_p})^T$ characterizes the predictor.

The optimal linear predictor for the autoregressive process is given by

$$\mathbf{h}_{N_p}^* = \mathbf{a}_{N_p}, \quad (5)$$

with $\mathbf{a}_{N_p} = (a_1, \ldots, a_m, 0, \ldots, 0)^T$, where the first m elements are given by the process parameter vector $\mathbf{a}_m$ and the remaining $N_p - m$ elements are all zeros [7]. Then, the prediction residual for optimal prediction is

$$U_t = S_t - \hat{S}_t = Z_t + \mu_U, \quad (6)$$

where $\{U_t\}$ is an i.i.d. process (white noise) with mean $\mu_U = \mu_S(1 - \mathbf{a}_m^T \mathbf{e}_m)$ and variance $\sigma_U^2 = \mathbb{E}[Z_t^2]$.

B. Identification rate $\mathbf{R}_{\text{ID}}^{\mathbf{P}}$

We define the *identification rate of the prediction-based scheme* R_{ID}^P as the infimum over all D_{ID}-achievable rates of the prediction-based approach. Theorem 1 gives R_{ID}^P for autoregressive Gaussian processes.

Theorem 1. *Let both the database and the query be generated by zero-mean autoregressive processes* $X_t = U_t + \mathbf{a}_m^T \mathbf{X}_{t-1}^{(m)}$ *and* $Y_t = V_t + \mathbf{a}_m^T \mathbf{Y}_{t-1}^{(m)}$ *according to (3) respectively, where* $\{U_t\}$ *and* $\{V_t\}$ *are two independent zero-mean Gaussian i.i.d. processes with unit variance. The identification rate of the autoregressive Gaussian source is*

$$R_{ID}^P = \frac{1}{t}\sum_{i=1}^{t} R_{ID}^{(i)} \quad (7)$$

$$R_{ID}^{(i)} = \max\left\{\log_2\left(\frac{2\ln(2)\lambda_{t,i}}{v}\right), 0\right\} \quad (8)$$

with $v \in \{0, 2\lambda_{t,1}\ln(2)\}$, *where* $\lambda_{t,1} \geq \cdots \geq \lambda_{t,t}$ *are the eigenvalues of* $\mathbf{M}_t^T\mathbf{M}_t$ *with the model* $\mathbf{x} = \mathbf{M}_t\mathbf{u}$. *The corresponding similarity threshold is determined by*

$$D_{ID}^P = \frac{1}{t}\sum_{i=1}^{t} \lambda_{t,i} 2\left(1 - 2^{-R_{ID}^{(i)}}\right). \quad (9)$$

Proof. From the modeling of the database process, the database vector can be represented by a matrix-vector multiplication, $\mathbf{x} = \mathbf{M}_t\mathbf{u}$, i.e.

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_t \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ m_1 & 1 & & & \\ & m_1 & 1 & & \\ \vdots & & \ddots & & \\ m_{t-1} & & \cdots & m_1 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_t \end{bmatrix},$$

where $(x_1, x_2, \cdots, x_t)$ is the database process and $(u_1, u_2, \cdots, u_t)$ the residual process. The parameters of the

process form the matrix $\mathbf{M}_t$ and its diagonal elements are $m_{i,i} = 1$ for $i = 1, 2, \ldots, t$. The other entries below the diagonal depend on the coefficients of $\mathbf{a}_m$ which are assumed to be estimated by an optimal predictor. The entries above the diagonal are equal to zero, *i.e.,* $\mathbf{M}_t$ is a lower triangular matrix. Similarly, we can write the query process as

$$\mathbf{y} = \mathbf{M}_t \mathbf{v}, \tag{10}$$

where $\mathbf{y}$ is the input query and $\mathbf{v}$ the residual of the query.

Now, we analyze the similarity between query and database processes $d(\mathbf{x}, \mathbf{y})$.

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &= \frac{1}{t} \sum_{i=1}^{t} \|x_i - y_i\|^2 && (11)\\ &= \frac{1}{t} (\mathbf{x} - \mathbf{y})^T (\mathbf{x} - \mathbf{y}) && (12)\\ &= \frac{1}{t} (\mathbf{M}_t(\mathbf{u} - \mathbf{v}))^T (\mathbf{M}_t(\mathbf{u} - \mathbf{v})) && (13)\\ &= \frac{1}{t} (\mathbf{u} - \mathbf{v})^T \mathbf{M}_t^T \mathbf{M}_t (\mathbf{u} - \mathbf{v}) && (14)\\ &= \frac{1}{t} (\mathbf{u} - \mathbf{v})^T \mathbf{Q} \mathbf{\Lambda}_t \mathbf{Q}^T (\mathbf{u} - \mathbf{v}) && (15)\\ &= \frac{1}{t} (\tilde{\mathbf{u}} - \tilde{\mathbf{v}})^T \mathbf{\Lambda}_t (\tilde{\mathbf{u}} - \tilde{\mathbf{v}}) && (16)\\ &= \frac{1}{t} \sum_{i=1}^{t} \lambda_{t,i} (\tilde{u}_i - \tilde{v}_i)^2 && (17)\\ &= \frac{1}{t} \sum_{i=1}^{t} \lambda_{t,i} d(\tilde{u}_i, \tilde{v}_i) && (18) \end{aligned}$$

with $d(\tilde{u}_i, \tilde{v}_i) := (\tilde{u}_i - \tilde{v}_i)^2$. Equation (15) comes from the eigendecomposition of a symmetric matrix, where $\mathbf{Q}$ is a $t \times t$ square matrix whose i-th column is the eigenvector $\mathbf{q}_i$ of $\mathbf{M}_t^T \mathbf{M}_t$, and $\mathbf{q}_i^T \mathbf{q}_i = 1$. $\mathbf{\Lambda}_t$ is the diagonal matrix whose diagonal elements Λ_{ii} are the corresponding eigenvalues $\lambda_{t,i}$. Note the eigenvalues $\lambda_{t,i}$ are positive for all finite i and t since $\mathbf{M}_t^T \mathbf{M}_t$ is positive definite. Therefore, the Gaussian variables $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{v}}$ are the projections of $\mathbf{u}$ and $\mathbf{v}$ on the eigenvectors of $\mathbf{M}_t^T \mathbf{M}_t$, $\tilde{\mathbf{u}} = \mathbf{Q}^T \mathbf{u}$ and $\tilde{\mathbf{v}} = \mathbf{Q}^T \mathbf{v}$. The projected variables are still i.i.d.

$$\mathbb{E}[(\mathbf{Q}^T \mathbf{u})(\mathbf{Q}^T \mathbf{u})^T] = \mathbb{E}[\mathbf{Q}^T \mathbf{u} \mathbf{u}^T \mathbf{Q}] = \mathbf{Q}^T \mathbf{I} \mathbf{Q} = \mathbf{I}. \tag{19}$$

We assume for each time step an ideal identification system for unit variance Gaussian data, i.e., $d(\tilde{u}_i, \tilde{v}_i) \leq \tilde{D}_{\text{ID}}^{(i)}$ with identification rate $\tilde{R}_{\text{ID}}^{(i)}$ such that $\tilde{D}_{\text{ID}}^{(i)} = 2\left(1 - 2^{-\tilde{R}_{\text{ID}}^{(i)}}\right)$ [1]. With above analysis, we have $\mathbf{x}$ and $\mathbf{y}$ D_{ID}-similar with

$$d(\mathbf{x}, \mathbf{y}) \leq \frac{1}{t} \sum_{i=1}^{t} \lambda_{t,i} \tilde{D}_{\text{ID}}^{(i)} := D_{\text{ID}}. \tag{20}$$

Then, the optimal rate allocation for each time step is obtained by

$$\begin{aligned} \max_{\tilde{D}_{\text{ID}}^{(1)}, \ldots, \tilde{D}_{\text{ID}}^{(t)}} \; & D_{\text{ID}} = \frac{1}{t} \sum_{i=1}^{t} \lambda_{t,i} \tilde{D}_{\text{ID}}^{(i)} \\ \text{s.t.} \quad & \frac{1}{t} \sum_{i=1}^{t} \tilde{R}_{\text{ID}}^{(i)} \leq R_{\text{ID}}, \\ \text{s.t.} \quad & \tilde{R}_{\text{ID}}^{(i)} \geq 0. \end{aligned} \tag{21}$$

As all similarity-rate functions $\tilde{D}_{\text{ID}}^{(i)}\left(\tilde{R}_{\text{ID}}^{(i)}\right)$ are concave and strictly increasing, we consider the equivalent problem

$$\begin{aligned} \max \quad & J = D_{\text{ID}} - v R_{\text{ID}} \\ \text{s.t.} \quad & \tilde{R}_{\text{ID}}^{(i)} \geq 0, \end{aligned} \tag{22}$$

where v is a positive Lagrange multiplier. The derivative of the cost function J with respect to $\tilde{R}_{\text{ID}}^{(i)}$ is

$$\frac{\partial J}{\partial \tilde{R}_{\text{ID}}^{(i)}} = \frac{2}{t} \lambda_{t,i} 2^{-\tilde{R}_{\text{ID}}^{(i)}} \ln(2) - \frac{1}{t} v. \tag{23}$$

Setting (23) to zero, we obtain

$$\tilde{R}_{\text{ID}}^{(i)} = -\log_2 \left(\frac{v}{2 \ln(2) \lambda_{t,i}} \right). \tag{24}$$

In order to satisfy the nonnegative constraint for the rate, each time step is only assigned a rate if the multiplier v is smaller than $2 \ln(2) \lambda_{t,i}$, with the largest multiplier being

$$v_{\max} = 2 \lambda_{t,1} \ln(2). \tag{25}$$

Then we sweep over permitted values of the Lagrange multiplier $v \in \{0, 2\lambda_1 \ln(2)\}$ to obtain the $\left(R_{\text{ID}}^P, D_{\text{ID}}^P\right)$ curve. □

IV. SPECIAL CASE

A. $R_{\text{ID}}^{\text{PS}}$ for Autoregressive Processes

We denote the symmetric matrix $\mathbf{M}_t^T \mathbf{M}_t$ from (14) as $\mathbf{P}_t$. $\mathbf{P}_t$ is the product of triangular Toeplitz matrices. Note that such a product can also be used to represent the inverse of the autocorrelation matrix of an autoregressive process as defined in [8]. We discuss a special case of the prediction-based approach if only the smallest eigenvalue of $\mathbf{P}_t$ is known. As we will show in the next section, the smallest eigenvalue of $\mathbf{P}_t$ is bounded by the *essential infimum* of the power spectral density of the stochastic process, which is usually available in practical scenarios.

In this section, we give an approximation $R_{\text{ID}}^{\text{PS}}$ for the identification rate. For that, we consider the same database and query autoregressive processes as in Theorem 1. Proposition 1 gives our result for $R_{\text{ID}}^{\text{PS}}$.

Proposition 1. *Given a prediction-based identification system with the same setup as in Theorem 1, its identification rate is approximated by*

$$R_{ID}^{PS}(D_{ID}) = \log_2 \left(\frac{2\lambda_{\min}}{2\lambda_{\min} - D_{ID}} \right), \tag{26}$$

where λ_{min} is the smallest eigenvalue of $\mathbf{M}_t^T\mathbf{M}_t$ for $t \to \infty$.

Proof. Continuing from (17), we have

$$d(\mathbf{x},\mathbf{y}) = \frac{1}{t}\sum_{i=1}^{t}\lambda_{t,i}(\tilde{u}_i - \tilde{v}_i)^2 \tag{27}$$

$$\geq \lambda_{\min}\frac{1}{t}\sum_{i=1}^{t}(\tilde{u}_i - \tilde{v}_i)^2 \tag{28}$$

$$= \lambda_{\min}\frac{1}{t}\|\tilde{\mathbf{u}} - \tilde{\mathbf{v}}\|^2 \tag{29}$$

$$= \lambda_{\min}\frac{1}{t}\|\mathbf{u} - \mathbf{v}\|^2 \tag{30}$$

$$= \lambda_{min} d(\mathbf{u},\mathbf{v}), \tag{31}$$

where (30) is due to the orthonormality of $\mathbf{Q}$.

Now, let the residual processes $\mathbf{u}$ and $\mathbf{v}$ be $\hat{D}_{\text{ID}}$-similar, i.e., $d(\mathbf{u},\mathbf{v}) \leq \hat{D}_{\text{ID}}$. With above result, we have

$$d(\mathbf{u},\mathbf{v}) \leq \frac{d(\mathbf{x},\mathbf{y})}{\lambda_{min}} \leq \frac{D_{\text{ID}}}{\lambda_{min}} := \hat{D}_{\text{ID}}. \tag{32}$$

Further, the residual process is unit variance Gaussian, and we have the identification rate [1]

$$R_{\text{ID}}^{\text{PS}} = \log_2\left(\frac{2}{2-\hat{D}_{\text{ID}}}\right). \tag{33}$$

□

B. Asymptotic Upper Bound of $R_{\text{ID}}^{\text{PS}}$

We first derive an asymptotic lower bound for the $\lambda_{\min}$ of $\mathbf{P}_t$. Similar to the derivation in [8], the inverse covariance matrix $\mathbf{P}_t$

$$\mathbf{P}_t = \left\{\sum_{k=0}^{t-\max(i,j)} m_k m_{k+|i-j|}\right\} \tag{34}$$

is asymptotically equivalent to a Toeplitz matrix, as defined by a nonnegative function $g(\omega)$ with its domain $[-\pi,\pi]$,

$$T_t(g) = \{r_{i,j}\} = \left\{\sum_{k=0}^{\infty} m_k m_{k+|i-j|}\right\} \tag{35}$$

where $g(\omega)$ is

$$g(\omega) = \sum_{k=-\infty}^{\infty} e^{-\mathbf{j}k\omega}\left\{\sum_{i=0}^{\infty} m_i m_{i+k}\right\} = \left|\sum_{k=0}^{\infty} m_k e^{-\mathbf{j}k\omega}\right|^2. \tag{36}$$

The entries of the Toeplitz matrix are given by $g(\omega)$:

$$r_{i,j} = r_{i-j} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-\mathbf{j}|i-j|}g(\omega)d(\omega) \tag{37}$$

Let $\tau_{t,i}$ be the eigenvalues of $T_t(g)$. According to the Theorems 2.1 and 6.2 of [9], if two sequences $\{\mathbf{P}_t\}$ and $\{T_t(g)\}$ are asymptotically equivalent, then there are constants N_l and N_u such that

$$N_l \leq \lambda_{t,k}, \tau_{t,k} \leq N_u, t = 1,2,\cdots, k = 0,1,\cdots,t-1, \tag{38}$$

where the *essential infimum* $N_l = \operatorname{ess\,inf} g$ is defined as the largest value of a for which $g(\omega) \geq a$ except on a set of total length or measure 0, and the *essential supremum* $N_u = \operatorname{ess\,sup} g$. According to Corollary 4.2 of [9], the minimum eigenvalue of a Toeplitz matrix converges to the essential infimum of g

$$\lim_{t\to\infty}\min_i \tau_{t,i} = N_l. \tag{39}$$

Hence, we have N_l as the lower bound of $\lambda_{\min}$. Since the identification rate is an increasing function of the similarity threshold, $R_{\text{ID}}^{\text{PS}}$ is upper bounded by using N_l.

V. NUMERICAL RESULTS

We consider a zero-mean Gauss-Markov source with unit variance and correlation coefficient $\rho = 0.3$. $\mathbf{P}_t$ approaches a Toeplitz matrix which is defined by using

$$g(\omega) = \frac{1}{1 - 2\rho\cos(\omega) + \rho^2}. \tag{40}$$

The $\operatorname{ess\,inf} g(\omega) = 0.5917$ when $\cos(\omega) = -1$. We simulate sequences of length $t = 20000$ and observe that the minimum eigenvalue $\lambda_{\min}$ of $\mathbf{M}_t^T\mathbf{M}_t$ converges to its lower bound 0.5917. We compare R_{ID}^{P} and $R_{\text{ID}}^{\text{PS}}$ in Fig. 2.

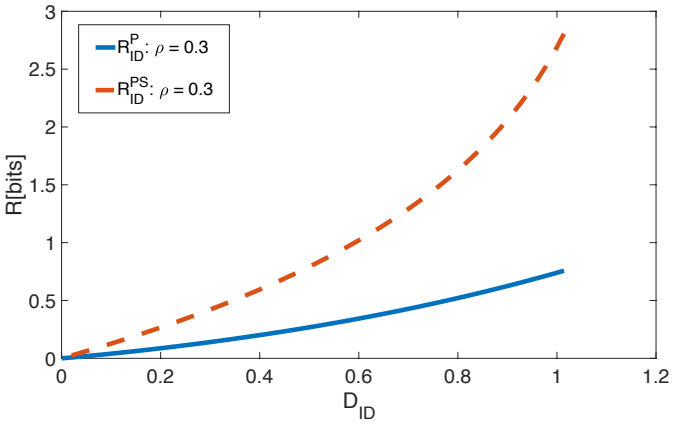

Fig. 2. R_{ID}^{P} and $R_{\text{ID}}^{\text{PS}}$ for AR(1) sequences with $\rho = 0.3$.

VI. CONCLUSIONS

In this work, we derive the identification rate of a prediction-based approach for autoregressive Gaussian processes. We show that it depends on a sequence of eigenvalues that we derive from our prediction model. Further, we approximate the identification rate for a special case and derive its upper bound by using the minimum eigenvalue of Toeplitz matrices.

VII. REFERENCES

[1] A. Ingber, T. Courtade, and T. Weissman, "Compression for quadratic similarity queries," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2729 –2747, May 2015.

[2] A. Ingber and T. Weissman, "The minimal compression rate for similarity identification," 2013, [Online]. Available: http://arxiv.org/abs/1312.2063.

[3] H. Wu, Q. Wang, and M. Flierl, "Tree-structured vector quantization for similarity queries," in *2017 Data Compression Conference (DCC)*, April 2017.

[4] H. Wu and M. Flierl, "Transform-based compression for quadratic similarity queries," in *Conference on Signals, Systems, and Computers*, Oct. 2017, pp. 377–381.

[5] ——, "Component-based quadratic similarity identification for multivariate gaussian sources," in *2018 Data Compression Conference (DCC)*, Mar. 2018.

[6] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra, "Dimensionality reduction for fast similarity search in large time series databases," *Knowledge and Information Systems*, vol. 9, pp. 263–286, 2001.

[7] J. R. Ohm, *Multimedia Signal Coding and Transmission*. Berlin-Heidelberg: Spring-Verlag, 2015, pp. 193 –195.

[8] R. M. Gray, "Information rates of autoregressive processes," *IEEE Trans. Inf. Theory*, vol. 16, no. 4, pp. 412 – 421, Jul. 1970.

[9] ——, *Toeplitz and Circulant Matrices: A Review*. Delft, The Netherlands: Now publishers Inc, 2006.