THE EFFECTIVE RANK: A MEASURE OF EFFECTIVE DIMENSIONALITY

Olivier Roy † *and Martin Vetterli* †⋄

† School of Computer and Communication Sciences
Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
⋄ Department of Electrical Engineering and Computer Sciences
University of California at Berkeley (UCB), Berkeley, CA 94720, USA
Email: {olivier.roy, martin.vetterli}@epfl.ch

ABSTRACT

Many signal processing algorithms include numerical problems where the solution is obtained by adjusting the value of parameters such that a specific matrix exhibits rank deficiency. Since rank minimization is generally not practicable owing to its integer nature, we propose a real-valued extension that we term effective rank. After proving some of its properties, the effective rank is provided with an operational meaning using a result on the coefficient rate of a stationary random process. Finally, the proposed measure is assessed in a practical scenario and other potential applications are suggested.

1. INTRODUCTION

Various signal processing tasks involve optimization problems that consist in finding the values of a set of parameters which make a matrix rank deficient. Examples include subspace based signal analysis techniques [1], the registration of multiple sets of samples shifted by unknown offsets [2] or the estimation of parametric diffusive sources using tomographic methods [3]. Since the rank is an integer quantity by definition, minimizing its value with respect to the involved parameters is unfeasible with standard numerical optimization methods (e.g. gradient descent, Newton's method). If the minimum rank is known not to be larger than N, the alternative usually envisioned simply amounts to minimize its $N+1$-th largest singular value. While being simple, this strategy has three major drawbacks: it is very sensitive to noise, it requires an a-priori knowledge of the minimum rank and it does not take into account the full singular value spectrum.

To overcome these limitations, we introduce the concept of *effective rank* which can be considered as a real-valued extension of the rank. We prove some of its properties and compute it for a simple example. We then provide its operational meaning using the notion of coefficient rate introduced by Campbell in [4]. In particular, our effective rank can be thought as a coefficient rate in discrete form. Our main contribution in this paper is thus to endow this quantity with a rank interpretation. It should be noted that, while the use of spectral entropy based measures have attracted attention in different contexts (see e.g. [5, 6]), Campbell's result has largely gone unnoticed. Some interesting work relating coefficient rate to source coding concepts can be found in [7].

The work presented in this paper was supported in part by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322.

Finally, the practicality of our effective rank is demonstrated in a specific scenario and compared to the largest singular value minimization described previously. Other possible applications are briefly outlined.

The paper is organized as follows: in Section 2, we introduce the effective rank, prove some of its properties and provide a simple illustrative example. Its operational meaning is then investigated in Section 3. Finally, Section 4 applies the effective rank to a practical scenario and suggests other potential applications.

2. THE EFFECTIVE RANK

In this section, we first define the notion of effective rank and comment on its intuitive meaning. We then prove some of its properties as a means to relate it to the rank of a matrix. A simple illustrative example is finally provided.

2.1 Definition

Let us consider a complex-valued non-all-zero matrix A of size $M \times N$ whose singular value decomposition (SVD) is given by $A = UDV$ where U and V are unitary matrices of size $M \times M$ and $N \times N$, respectively, and D is an $M \times N$ diagonal matrix containing the (real positive) singular values

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_Q \geq 0\,,$$

with $Q = \min\{M, N\}$. For notational simplicity, we further define $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_Q)^T$ and the singular value distribution

$$p_k = \frac{\sigma_k}{\|\sigma\|_1} \quad \text{for } k = 1, 2, \ldots, Q\,, \tag{1}$$

where the superscript ^{T} denotes the transpose and $\|\cdot\|_1$ the ℓ_1-norm defined as

$$\|\sigma\|_1 = \Sigma_{k=1}^{Q} |\sigma_k|\,.$$

In the sequel, all logarithms are to the base e and we adopt the convention that $0 \log 0 = 0$. The effective rank is defined as follows.

Definition 1 (Effective Rank) *The effective rank of the matrix A, denoted* $\text{erank}(A)$, *is defined as*

$$\text{erank}(A) = \exp\{H(p_1, p_2, \ldots, p_Q)\}\,,$$

where $H(p_1, p_2, \ldots, p_Q)$ *is the (Shannon) entropy given by*

$$H(p_1, p_2, \ldots, p_Q) = -\sum_{k=1}^{Q} p_k \log p_k\,.$$

Note that the above entropy is often referred to as spectral entropy [4, 7].

While the effective rank can be given a precise operational meaning using Campbell's result [4] (see Section 3), the above definition is intuitively motivated by the following observation. The matrix A is a linear mapping from the vector space $\mathbb{C}^N$ to the vector space $\mathbb{C}^M$. A possible orthonormal basis for $\mathbb{C}^M$ is given by the columns of U, denoted u_k ($k = 1,2,\ldots,M$). Similarly, the N columns of V, denoted v_l ($l = 1,2,\ldots,N$), form an orthonormal basis of $\mathbb{C}^N$. They satisfy the following relation

$$w_k \triangleq Av_k = \begin{cases} \sigma_k u_k & \text{for } k = 1,2,\ldots,Q\,, \\ 0 & \text{otherwise.} \end{cases}$$

The space spanned by the vectors w_k's is commonly referred to as the range of A [8, Section 0.2.3]. We observe that each basis vector u_k is multiplied by a factor σ_k which hence provides the transformation A with a geometrical shaping interpretation. In this context, the rank of A corresponds to the number of dimensions retained by the transformation (i.e. the dimension of its range) but says nothing about the induced shaping. The effective rank, however, quantifies such geometrical transformation by means of the spectral entropy. It thus provides the range of A with an "effective dimension". Note that, unlike the matrices U and V, the singular value distribution is unique and so is the effective rank.

To intuitively understand the difference between the rank and the effective rank, a typical example is that of a bi-dimensional Gaussian random vector with highly correlated components. Its covariance matrix is of rank two, but the corresponding Gaussian distribution exhibits most of its energy along the direction of one singular vector. In this case, the spectral entropy approaches zero, hence resulting in an effective rank slightly greater than one.

2.2 Properties

This section provides a few properties of the effective rank along with their proofs. It should be noted that, while some properties of the rank naturally extend to the effective rank, this is not true in general owing to the strong dependance on the singular value distribution.

Property 1 *It holds that*

$$1 \leq \mathrm{erank}(A) \leq \mathrm{rank}(A) \leq Q$$

where the first inequality holds with equality if and only if

$$\sigma = (\|\sigma\|_1, 0, \ldots, 0)^T\,,$$

and the second one if and only if

$$\sigma = (\|\sigma\|_1/k, \ldots, \|\sigma\|_1/k, 0, \ldots, 0)^T$$

for some $k \in \{1,2,\ldots,Q\}$.

Proof: The entropy $H(p_1, p_2, \ldots, p_Q)$ satisfies [9, Section D.1]

$$\begin{aligned} 0 &= H(1,0,\ldots,0) \\ &\leq H(p_1,p_2,\ldots,p_Q) \\ &\leq H(1/Q,1/Q,\ldots,1/Q) \\ &= \log Q\,. \end{aligned}$$

The effective rank thus satisfies $1 \leq \mathrm{erank}(A)$ with equality if and only if $(p_1, p_2, \ldots, p_Q) = (1, 0, \ldots, 0)$, i.e. $\sigma = (\|\sigma\|_1, 0, \ldots, 0)^T$. Suppose now that only k singular values of A are non-zero for some $k \in \{1,2,\ldots,Q\}$. In this case, $\mathrm{rank}(A) = k$ and $H(p_1,p_2,\ldots,p_Q) = H(p_1,p_2,\ldots,p_k) \leq \log k$. Hence $\mathrm{erank}(A) \leq \mathrm{rank}(A) \leq Q$ with $\mathrm{erank}(A) = \mathrm{rank}(A)$ if and only if $(p_1 \ldots, p_k, p_{k+1}, \ldots, p_Q) = (1/k, \ldots, 1/k, 0, \ldots, 0)$, i.e. $\sigma = (\|\sigma\|_1/k, \ldots, \|\sigma\|_1/k, 0, \ldots, 0)^T$. ■

The above property shows that $\mathrm{erank}(A)$ is upper bounded by $\mathrm{rank}(A)$ and that equality holds when the singular value distribution is uniform over its support. An important observation is that the effective rank can take all possible values in the interval $[1, Q]$ as opposed to the integer value of the rank in the set $\{1,2,\ldots,Q\}$. This makes the use of numerical optimization methods on the effective rank feasible. Let us now denote by A^* and $\bar{A}$ the Hermitian transpose and the complex conjugate of the matrix A, respectively. We have the following result.

Property 2 *It holds that*

$$\mathrm{erank}(A) = \mathrm{erank}(A^*) = \mathrm{erank}(A^T) = \mathrm{erank}(\bar{A}) = \mathrm{erank}(cA)$$

for all $c \neq 0$.

Proof: The property simply follows from the fact that the p_k's defined by equation (1) are the same for the matrices A, A^*, A^T, $\bar{A}$ and cA for all $c \neq 0$. ■

The following property also holds.

Property 3 *A unitary transformation on A does not change its effective rank.*

Proof: Let us assume without lost of generality that $M \leq N$. The singular values of A are the (principal) square roots of the eigenvalues of the matrix AA^*. Let U denote an $M \times M$ unitary transform matrix. We have from the determinant formula $\det(AB+I) = \det(BA+I)$ that

$$\det\left((UA)(UA)^* - \lambda I_M\right) = \det\left(AA^* - \lambda I_M\right)\,,$$

i.e. the eigenvalues of $(UA)(UA)^*$ and AA^* are the same. The effective rank thus remains unchanged. ■

As a special case of the above property, the only elementary operation that preserves the effective rank corresponds to the interchange of two rows or two columns of A. Finally, similarly to the rank inequality $\mathrm{rank}(A+B) \leq \mathrm{rank}(A) + \mathrm{rank}(B)$ [8, Section 0.4.5], we can state the following property.

Property 4 *Let A and B be two positive semidefinite Hermitian matrices of size* $N \times N$. *It holds that*

$$\mathrm{erank}(A+B) \leq \mathrm{erank}(A) + \mathrm{erank}(B)\,.$$

Proof: Let us denote the singular values arranged in decreasing order of A, B and $A+B$ by $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_N)^T$, $\mu = (\mu_1, \mu_2, \ldots, \mu_N)^T$ and $\nu = (\nu_1, \nu_2, \ldots, \nu_N)^T$, respectively. Since A and B are positive semidefinite Hermitian matrices, their singular values correspond to their eigenvalues and

$$\|\sigma\|_1 + \|\mu\|_1 = \mathrm{tr}(A) + \mathrm{tr}(B) = \mathrm{tr}(A+B) = \|\nu\|_1 .$$

Let $p_k = \sigma_k/\|\sigma\|_1$, $q_k = \mu_k/\|\mu\|_1$ and $r_k = \nu_k/\|\nu\|_1$ for $k = 1,2,\ldots,N$. Since $\exp(x)$ is a convex function, for all $x_1, x_2 \in \mathbb{R}$ and $\lambda \in [0,1]$, we have that

$$\exp(\lambda x_1 + (1-\lambda)x_2) \le \lambda \exp(x_1) + (1-\lambda)\exp(x_2) . \quad (2)$$

In particular, if we set

$$\begin{aligned} x_1 &= H(p_1,p_2,\ldots,p_N) - \log\lambda , \\ x_2 &= H(q_1,q_2,\ldots,q_N) - \log(1-\lambda) \quad \text{and} \\ \lambda &= \|\sigma\|_1/(\|\sigma\|_1 + \|\mu\|_1) , \end{aligned}$$

we obtain that

$$\begin{aligned} &\lambda \exp(x_1) + (1-\lambda)\exp(x_2) \\ &= \exp\{H(p_1,p_2,\ldots,p_N)\} + \exp\{H(q_1,q_2,\ldots,q_N)\} \\ &= \mathrm{erank}(A) + \mathrm{erank}(B) . \quad (3) \end{aligned}$$

We can also write

$$\begin{aligned} &\lambda x_1 + (1-\lambda)x_2 \quad (4) \\ &= \lambda\,(H(p_1,p_2,\ldots,p_N) - \log\lambda) \\ &\quad +(1-\lambda)\,(H(q_1,q_2,\ldots,q_N) - \log(1-\lambda)) \\ &= -\sum_{k=1}^{N} \frac{\sigma_k}{\|\sigma\|_1 + \|\mu\|_1} \log \frac{\sigma_k}{\|\sigma\|_1 + \|\mu\|_1} \\ &\quad -\sum_{k=1}^{N} \frac{\mu_k}{\|\sigma\|_1 + \|\mu\|_1} \log \frac{\mu_k}{\|\sigma\|_1 + \|\mu\|_1} . \end{aligned}$$

Furthermore, it follows from [9, Theorem G.1.b] that

$$\frac{(\sigma,\mu)}{\|\sigma\|_1 + \|\mu\|_1} \prec \frac{(\nu,0)}{\|\sigma\|_1 + \|\mu\|_1} = \frac{(\nu,0)}{\|\nu\|_1} , \quad (5)$$

where $\prec$ denotes majorization. Since the function $f(x) = -x\log x$ is concave on $(0,1]$, we can use [9, Proposition B.1] to lower bound the left-hand side of (2) as

$$\begin{aligned} \exp(\lambda x_1 + (1-\lambda)x_2) &\ge \exp\left(-\sum_{k=1}^{N} \frac{\nu_k}{\|\nu\|_1} \log \frac{\nu_k}{\|\nu\|_1}\right) \\ &= \exp\{H(r_1,r_2,\ldots,r_N)\} \\ &= \mathrm{erank}(A+B) . \quad (6) \end{aligned}$$

Combining equations (2), (3) and (6) yields the desired result. ■

It is not clear whether Property 4 still holds for arbitrary $M \times N$ matrices A and B. In general, the vector $\sigma + \mu$ only weakly majorizes ν and the last step of the proof cannot be applied. Furthermore, one would need to find positive semidefinite Hermitian matrices with prescribed eigenvalues (see e.g. [10, Theorem 1]) such as to satisfy equation (5).

We also remark that, with minor modifications, Properties 1 to 3 still hold if the ℓ_1-norm in Definition 1 is replaced by the ℓ_p-norm ($p \ge 1$)

$$\|\sigma\|_p = \left(\Sigma_{k=1}^{Q} |\sigma_k|^p\right)^{\frac{1}{p}} .$$

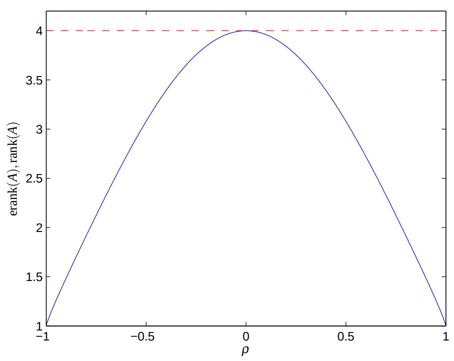


Figure 1: The effective rank (plain) and the rank (dashed) of the matrix A of Section 2.3 as a function of the correlation parameter ρ. As the correlation increases, erank(A) decreases whereas rank(A) remains the same.

Interestingly, if one uses the ℓ_0-norm (which simply counts the number of non-zero singular values) the effective rank becomes equivalent to the rank. In other words, the rank can be seen as an effective rank with a particular vector norm.

2.3 Example

We now compute the effective rank of a simple matrix to illustrate the theory developed previously. Let us consider the 4×4 positive semidefinite Hermitian circulant matrix A defined as

$$A = \begin{pmatrix} 1 & \rho & \rho^2 & \rho \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho & \rho^2 & \rho & 1 \end{pmatrix} ,$$

where $\rho \in [-1,1]$ is a correlation parameter. Its singular values (eigenvalues) are easily computed as $(1+|\rho|)^2, 1-|\rho|^2, 1-|\rho|^2$ and $(1-|\rho|)^2$. Using Definition 1, a straightforward derivation reveals that

$$\begin{aligned} &\mathrm{erank}(A) \\ &= \exp\left\{H\left(\frac{(1+|\rho|)^2}{4}, \frac{1-|\rho|^2}{4}, \frac{1-|\rho|^2}{4}, \frac{(1-|\rho|)^2}{4}\right)\right\} \\ &= \exp\left\{-(1+|\rho|)\log\frac{1+|\rho|}{2} - (1-|\rho|)\log\frac{1-|\rho|}{2}\right\} \\ &= \exp\left\{2H\left(\frac{1+|\rho|}{2}, \frac{1-|\rho|}{2}\right)\right\} . \end{aligned}$$

As illustrated in Figure 1, the effective rank is maximized when $\rho = 0$ and corresponds to the rank of the matrix A. However, as $|\rho|$ increases, the rank remains the same whereas the effective rank decreases. It hence provides the range of A with an "effective dimension".

3. OPERATIONAL MEANING

As pointed out previously, the effective rank is closely related to the concept of coefficient rate introduced by Campbell in [4]. In order to provide the effective rank with an operational meaning, we present in the sequel a similar derivation to that in [4] (see also [7]) for the case of random vectors. To this end, we first note that to every $M \times N$ matrix A, we can associate the $M \times M$ positive semidefinite Hermitian matrix $\sqrt{AA^*}$ which has the same singular values, possibly with additional zeros. It thus follows that

$$\mathrm{erank}(A) = \mathrm{erank}(\sqrt{AA^*})$$

and the operational meaning can be equivalently given in terms of the matrix $\sqrt{AA^*}$. Let us assume without lost of generality that $M \leq N$ and denote by C the Karhunen-Loève transform (KLT) of the matrix $\sqrt{AA^*}$, i.e. the unitary matrix satisfying

$$C^* \sqrt{AA^*} C = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_m)\,.$$

We now consider R i.i.d. random vectors $X_1, X_2, \ldots, X_R$ of size M with mean zero and covariance matrix $\sqrt{AA^*}$. Their Karhunen-Loève expansion is given by

$$X_r = \sum_{k=1}^{M} Y_{r,k} c_k\,, \quad \text{for } r = 1, 2, \ldots, R, \tag{7}$$

where c_k denotes the k-th column of the matrix C and where the $Y_{r,k}$'s are uncorrelated random variables with mean zero and variance $\mathrm{E}[|Y_{r,k}|^2] = \sigma_k$. We then define the product of R components of these random vectors as

$$Z(l_1, l_2, \ldots, l_R) = \prod_{r=1}^{R} X_r(l_r)\,, \tag{8}$$

where $X_r(l_r)$ denotes the l_r-th component of the random vector X_r with $l_r \in \{1, 2, \ldots, M\}$. In an analogous manner to [7, Section II], the R-dimensional random process defined by equation (8) can be expanded using (7) as

$$\begin{aligned} Z(l_1, l_2, \ldots, l_R) &= \prod_{r=1}^{R} \sum_{k=1}^{M} Y_{r,k} c_k(l_r) \\ &= \sum_{k=1}^{M^R} Y^{(k)} c^{(k)}\,, \end{aligned}$$

where we define $c^{(k)} = c_{k_1}(l_1) c_{k_2}(l_2) \cdots c_{k_R}(l_R)$, with k indexing all possible R-tuples $(k_1, k_2, \ldots, k_R) \in \{1, 2, \ldots, M\}^R$ such that the coefficients $Y^{(k)} = Y_{1,k_1} Y_{2,k_2} \cdots Y_{R,k_R}$ are arranged in decreasing order of their variance. Note that the dependance of $c^{(k)}$ on $l_1, l_2, \ldots, l_R$ is implicit. The goal then is to approximate $Z(l_1, l_2, \ldots, l_R)$ using only the first K coefficients, that is

$$\hat{Z}_K(l_1, l_2, \ldots, l_R) = \sum_{k=1}^{K} Y^{(k)} c^{(k)}\,. \tag{9}$$

The resulting mean-squared error can be expressed using (8) and (9) as

$$\begin{aligned} &\frac{1}{\|\sigma\|_1^R} \sum_{l_1, l_2, \ldots, l_R = 1}^{M} \mathrm{E}\left[|Z(l_1, l_2, \ldots, l_R) - \hat{Z}_K(l_1, l_2, \ldots, l_R)|^2 \right] \\ &\quad = \frac{1}{\|\sigma\|_1^R} \sum_{l_1, l_2, \ldots, l_R = 1}^{M} \sum_{k=K+1}^{M^R} \mathrm{E}\left[|Y^{(k)} c^{(k)}|^2 \right] \\ &\quad = \frac{1}{\|\sigma\|_1^R} \sum_{k=K+1}^{M^R} \mathrm{E}\left[|Y^{(k)}|^2 \right] \sum_{l_1, l_2, \ldots, l_R = 1}^{M} |c^{(k)}|^2 \\ &\quad = \frac{1}{\|\sigma\|_1^R} \sum_{k=K+1}^{M^R} \mathrm{E}\left[|Y^{(k)}|^2 \right]\,, \end{aligned}$$

where the first equality follows from the fact that the $Y^{(k)}$'s are uncorrelated and the third one from the fact that $\|c_k\|_2^2 = 1$ for $k = 1, 2, \ldots, M$ since the matrix C is unitary. In [4], Campbell shows that it is possible to find a value K (that depends on R) such that, in the limit when R goes to infinity, the above approximation error vanishes. Furthermore, this K satisfies the asymptotic relation

$$K^{\frac{1}{R}} \xrightarrow{R \to \infty} \exp\{H(p_1, p_2, \ldots, p_M)\}\,,$$

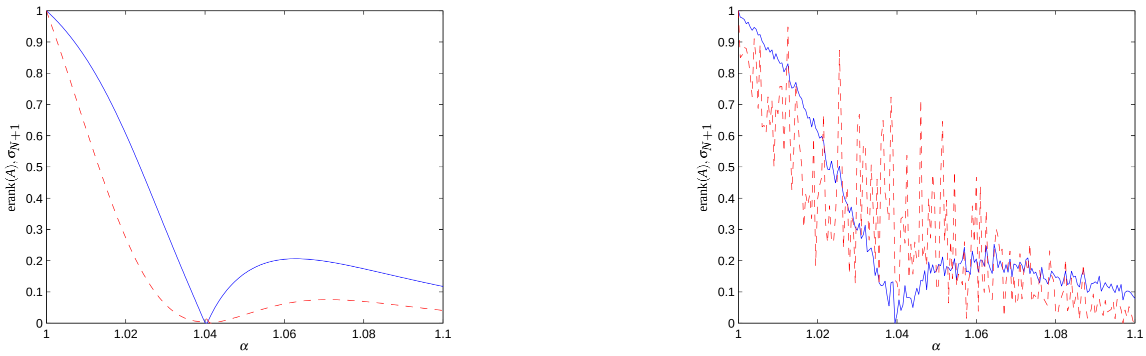
where the term on the right-hand side is recognized as the effective rank of the matrix A.

Campbell's result can be interpreted as follows. Each vector X_r can be represented by an M-dimensional random vector. The product Z defined by equation (8) thus admits a representation in a space with M^R dimensions, out of which only K are significant (in the sense that they contribute to the above approximation error in the limit of large R). Hence, on average, only $K^{1/R}$ coefficients out of M are significant in the expansion of X_r. In light of the above interpretation, the effective rank of a matrix A thus represents the average number of significant dimensions in the range of A, hence the terminology of "effective dimension".

Finally, the connection between effective rank and the coefficient rate of a stationary random process is established as follows. Assume that the matrix $\sqrt{AA^*}$ is of Toeplitz form with an absolutely summable generating sequence $\{a_k\}_{k \in \mathbb{Z}}$ such that $\|\sigma\|_1 = M$ (i.e. with appropriate normalization). Associate to it the power spectral density (PSD) $\Phi_A(\omega) = \sum_{k \in \mathbb{Z}} a_k e^{-j\omega k}$. The normalized version of the effective rank then satisfies, in the limit of large matrix size M,

$$\begin{aligned} &\lim_{M \to \infty} \frac{1}{M} \mathrm{erank}(A) \\ &\quad = \lim_{M \to \infty} \frac{1}{M} \exp\left(-\sum_{k=1}^{M} \frac{\sigma_k}{\|\sigma\|_1} \log \frac{\sigma_k}{\|\sigma\|_1} \right) \\ &\quad = \lim_{M \to \infty} \exp\left(-\frac{1}{M} \sum_{k=1}^{M} \sigma_k \log \sigma_k \right) \\ &\quad = \exp\left(-\int_{\omega \in [0, 2\pi]} \Phi_A(\omega) \log \Phi_A(\omega)\, d\omega \right)\,, \end{aligned} \tag{10}$$

where the last equality follows from the Toeplitz distribution theorem [11, Theorem 4.2]. The term in (10) corresponds to the coefficient rate of the discrete-time stationary random process with PSD $\Phi_A(\omega)$ defined in [4].



(a) (b)

Figure 2: Effective rank (plain) and smallest singular value (dashed) of the matrix A as a function of the parameters α. (a) Noiseless case. (b) Noisy case (SNR=50 [dB]). We observe that in both scenarios the effective rank provides the best results. Note that the matrix A is normalized and the results are scaled to the interval $[0,1]$ for comparison purposes.

4. APPLICATIONS

The effective rank proves useful in applications where multiple signals are coherently related through a finite number of unknown parameters (see e.g. [2, 12]). It may also be used to assess the loss incurred by dimensionality reduction methods, such as principal component analysis (PCA).

As a means to illustrate the potential of the effective rank in a practical scenario, we consider here the specific problem addressed in [3]. The goal is to estimate the parameters of local diffusive sources using a finite number of tomographic measurements. If N sources are present, it is shown in [3] that this task can be accomplished by finding the parameter $\alpha \geq 1$ such that the $(N+1)\times(N+1)$ matrix

$$A = \begin{pmatrix} r_N \alpha^{N^2} & r_{N-1}\alpha^{(N-1)^2} & \cdots & r_0 \alpha^0 \\ r_{N+1}\alpha^{(N+1)^2} & r_N \alpha^{N^2} & \cdots & r_1 \alpha^1 \\ \vdots & \vdots & \ddots & \vdots \\ r_{2N}\alpha^{(2N)^2} & r_{2N-1}\alpha^{(2N-1)^2} & \cdots & r_N \alpha^{N^2} \end{pmatrix}$$

is of rank N. Here r_n is a fixed scalar value ($n = 0,1,\ldots,2N$). This can be achieved either by minimizing the smallest singular value σ_{N+1} of A or by minimizing its effective rank. We plot in Figure 2 the two quantities for $N = 2$ as a function of the parameter α, in both a noiseless and a noisy case. For comparison purposes, the matrix A is normalized and the results are scaled to the interval $[0,1]$. In the noiseless case [Figure 2 (a)], the two methods provide the correct answer $\alpha_{opt} \simeq 1.04$. The minima obtained by the effective rank is however more precise. In the noisy scenario [Figure 2 (b)], the effective rank method clearly outperforms the singular value approach which basically provides no insight about the optimal solution.

REFERENCES

[1] A. van der Veen, E. Deprettere, and A. Swindlehurst, "Subspace-based signal analysis using singular value decomposition," *Proceedings of the IEEE*, vol. 81, pp. 1277–1308, September 1993.

[2] P. Vandewalle, L. Sbaiz, J. Vandewalle, and M. Vetterli, "Super-resolution from unregistered and totally aliased signals using subspace methods," *to appear in IEEE Transactions on Signal Processing*, 2007.

[3] I. Jovanović, L. Sbaiz, and M. Vetterli, "Tomographic approach for parametric estimation of local diffusive sources and applications to heat diffusion," *to appear in IEEE International Conference on Image Processing*, 2007.

[4] L. L. Campbell, "Minimum coefficient rate for stationary random processes," *Information and Control*, vol. 3, pp. 360–371, December 1960.

[5] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Transactions on Information Theory*, vol. 38, pp. 713–718, March 1992.

[6] J. D. Gibson, S. P. Stanners, and S. A. McClellan, "Spectral entropy and coefficient rate for speech coding," *Twenty-Seventh Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 925–929, November 1993.

[7] W. Yang, J. D. Gibson, and T. He, "Coefficient rate and lossy source coding," *IEEE Transactions on Information Theory*, vol. 51, pp. 381–386, January 2005.

[8] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, Cambridge, 1985.

[9] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and its Applications*, vol. 143 of *Mathematics in Science and Engineering*. Academic Press, New-York, 5th ed., 1979.

[10] W. Fulton, "Eigenvalues, invariant factors, highest weights, and Schubert calculus," *Bulletin of the American Mathematical Society*, vol. 36, no. 3, pp. 209–249, 2000.

[11] R. M. Gray, "Toeplitz and circulant matrices: a review," *Foundations and Trends in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.

[12] J. Chen, J. Benesty, and Y. Huang, "Robust time delay estimation exploiting redundancy among multiple microphones," *IEEE Transactions on Speech and Audio Processing*, vol. 11, pp. 549–557, November 2003.