Low-Rank Matrix Approximation Using Point-Wise Operators

Arash Amini, Amin Karbasi, *Student Member, IEEE*, and Farokh Marvasti, *Senior Member, IEEE*

***Abstract*—The problem of extracting low-dimensional structure from high-dimensional data arises in many applications such as machine learning, statistical pattern recognition, wireless sensor networks, and data compression. If the data is restricted to a lower dimensional subspace, then simple algorithms using linear projections can find the subspace and consequently estimate its dimensionality. However, if the data lies on a low-dimensional but nonlinear space (e.g., manifolds), then its structure may be highly nonlinear and, hence, linear methods are doomed to fail. In this paper, we introduce a new technique for dimensionality reduction based on point-wise operators. More precisely, let $\mathbf{A}_{n\times n}$ be a matrix of rank $k \ll n$ and assume that the matrix $\mathbf{B}_{n\times n}$ is generated by taking the elements of $\mathbf{A}$ to some real power p. In this paper, we show that based on the values of the data matrix $\mathbf{B}$, one can estimate the value p and, therefore, the underlying low-rank matrix $\mathbf{A}$; i.e., we are reducing the dimensionality of $\mathbf{B}$ by using point-wise operators. Moreover, the estimation algorithm does not need to know the rank of $\mathbf{A}$. We also provide bounds on the quality of the approximation and validate the stability of the proposed algorithm with simulations in noisy environments.**

***Index Terms*—Dimensionality reduction, low-rank matrix, point-wise operator.**

I. Introduction

A FUNDAMENTAL question regarding the real-world data is the degrees of freedom involved in their generation. The term *intrinsic dimension*, which is the main focus of dimensionality reduction methods, refers to the number of independent variables for describing a model or a class of signals. Investigation of the intrinsic dimension arises in a variety of settings such as machine learning [1], computer vision [2], sensor networks [3], bandwidth compression [4], and compressed learning [5]. The purpose of reducing the dimension is to obtain a more compact representation of the data with no or little loss of information. Simplification of the data structure enables us to use the traditional machinery provided for low-dimensional data, whereas oversimplifying, results in the loss of crucial information. Hence, there has been significant interest over the last few decades in finding dimensionality reductions that preserve as much information as possible.

Manuscript received November 27, 2010; revised July 13, 2011; accepted July 26, 2011. Date of publication September 12, 2011; date of current version January 06, 2012. A. Amini was supported in part by the Iranian Telecommunication Research Center (ITRC).

A. Amini and F. Marvasti are with the Advanced Communication Research Institute (ACRI), EE department, Sharif University of Technology, Tehran, Iran (e-mail: arashsil@ee.sharif.edu; marvasti@sharif.edu).

A. Karbasi is with the School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland (e-mail: amin.karbasi@epfl.ch).

Communicated by D. Palomar, Associate Editor for Detection and Estimation.

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.2011.2167714

One way to formalize this problem is by representing the data as a matrix $\mathbf{B}$ and then find a low-rank matrix $\mathbf{A}$ that well approximates $\mathbf{B}$. The main intuition behind this method is that the rank of a matrix roughly represents its degrees of freedom.

It is known that the best low-rank approximation (in terms of the Frobenius norm) of $\mathbf{B}$ is given by the truncated singular value decomposition (SVD). More precisely, let $\mathbf{B}_k$ denote the matrix obtained by keeping the k largest singular values of $\mathbf{B}$ and setting the rest to zero. Then one can show that for all $\mathbf{M}$ for which $\mathrm{rank}(\mathbf{M}) = k$

$$\|\mathbf{B} - \mathbf{B}_k\|_F \leq \|\mathbf{M} - \mathbf{B}\|_F$$

where $\|\cdot\|_F$ denotes the Frobenius norm. This approach is shown to be successful in a variety of areas such as information retrieval [6], face recognition [2] and matrix completion [7], [8] where the matrix $\mathbf{B}$ is approximately low-rank.

In this paper, we look at dimensionality reduction from the signal processing point of view. More specifically, in applications such as wireless sensor networks and ultrasound tomography, where the system contains a number of separated nodes and the communication channel distorts the transmitted signals among the nodes. In many cases, if the matrix $\mathbf{A}$ models the transmitted (ideal) signals from each node to the rest of the network, $\mathbf{A}$ is low-rank due to the structure of the network and the type of the transmitted signals. Now, the corresponding matrix of the received signals, $\mathbf{B}$, is formed from $\mathbf{A}$ through point-wise but nonlinear operations. Hence, the linear approximation of $\mathbf{B}$ will no longer lead to a faithful representation of the data, while finding the inverse operator will do the job. Thus, the fundamental question is how to find this inverse operator which leads to the approximation of the desired low-rank matrix.

To illustrate this point, below we will describe two applications as typical examples.

A. Calibration in Ultrasound Tomography: Known Operator

In ultrasound tomography with circular apertures, there are transmitters and receivers installed on the interior edge of a circular ring surrounding an unknown object. At each time instance, a transmitter is fired and all the other sensors on the ring record the scattered signals and the process is repeated for all the transmitters. Each sensor on the ring is capable of transmitting and receiving ultrasound signals. The aim of tomography, in general, is to reconstruct the characteristics of the 2-D (or 3-D) enclosed object based on the recorded data (e.g., sound speed, sound attenuation, etc.). The general setup for such a tomography device is depicted in Fig. 1.

Before the start of the recording process, the apparatus should be calibrated. To this end, the relative time of flights between each transmitter-receiver pair are measured. More formally, let t_{ij} denote the flight time between transmitter i and receiver

0018-9448/$26.00 © 2011 IEEE

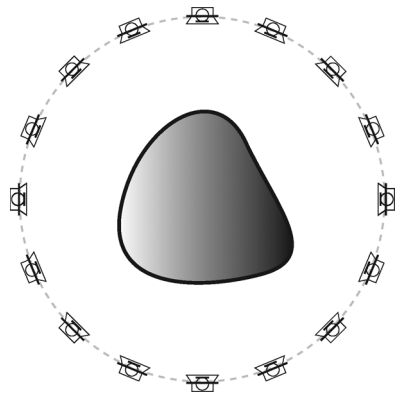

Fig. 1. Circular setup for ultrasound tomography. Sensors are fired each in turn and the remaining sensors record the arriving ultrasound signals.

j and let $\mathbf{T}$ denote the matrix of flight times given by $\mathbf{T} = [t_{ij}]_{n\times n}$. Because of the circular placement of the sensors, it is easy to check that only a small fraction of the total $\binom{n}{2}$ flight times suffices to estimate the rest; i.e., the intrinsic dimension is much less than the nominal dimension. In general, $\mathbf{T}$ is a full-rank matrix; however, the matrix $\tilde{\mathbf{T}} = [t_{i,j}^2]_{n\times n}$ is shown to have rank 3 [9]. Therefore, the linear projection of $\tilde{\mathbf{T}}$ onto the rank 3 matrices will provide us with a much better calibration result than the received matrix $\mathbf{T}$.

This is a situation where both the intrinsic dimension (rank=3) and the point-wise operator (x^2) are known a priori. The same phenomenon also happens in the Time of Arrival estimation (TOA) of a sound source when a series of microphones is deployed [10].

B. Sensor Localization

In radio communications, the received signal power decreases as the distance between the transmitter and the receiver increases; this phenomenon is called path-loss. In general, the received signal power from a transmitter at distance r is proportional to

$$\frac{1}{r^\alpha}$$

where the exponent α varies between 2 in free space to 6 in heavily built urban areas [11].

In a sensor localization problem, n sensor nodes are located in a d-dimensional space where each sensor measures its distance to other (probably neighboring) sensors. In practice, exact distance measurements are not directly available and must be estimated using the Received Signal Strength (RSS). More formally, let d_{ij} denote the Euclidean distance between the nodes i and j, and let $\mathbf{D}$ denote the distance matrix given by $\mathbf{D} = [d_{i,j}]_{n\times n}$. Furthermore, let p_{ij} denote the received signal power by sensor i from sensor j and let $\mathbf{P} = [p_{i,j}]_{n\times n}$. Given the pairwise distances, the positions of the sensors can be found using MultiDimensional Scaling (MDS) [12]. Unfortunately, in practice the matrix $\mathbf{D}$ is not available and should be estimated through $\mathbf{P}$. Although the matrix $\mathbf{P}$ is full-rank in general, it is not difficult to see that its point-wise transformed matrix $\tilde{\mathbf{P}} = [p_{i,j}^{-2\alpha}]_{n\times n}$ has a rank not exceeding $d+2$. Therefore, estimating the matrix $\tilde{\mathbf{P}}$ will provide us with a low-rank matrix that can later be used for many kinds of sensor localization algorithms [9], [13].

Since α is a property of the environment, it is usually unknown. Hence, we are in a situation where the intrinsic dimension is known $(\mathrm{d}+2)$ but the point-wise operator linking the high rank matrix to the low-rank one is unknown ($x^{-2\alpha}$).

In this work, we consider this problem in its general form, namely, two matrices, one low-rank and the other full-rank (or with a higher rank than the first one) that are linked through a polynomial point-wise operator. Given the full-rank matrix, we would like to obtain the low-rank one without a priori knowledge of the rank or the point-wise operator.

C. Related Work

Truncated singular value decomposition became popular by the pioneer work of Papadimitriou *et al.* [14] who proved that latent semantic analysis works under the context of a simplified model. This method generates faithful low-dimensional representations when the high-dimensional input patterns are mainly confined to a low-dimensional subspace.

The nuclear norm of a Matrix is defined as the sum of the singular values. For a partially known matrix (e.g., we know some of the entries or their linear combinations) it is shown in [15] that by constraint nuclear norm minimization we can achieve the matrix with the minimum rank satisfying the constraints. This can be interpreted as the matrix form of the compressed sensing where low rank matrices are considered as the 2-D generalization of sparse vectors and ℓ_1-norm minimization is replaced with nuclear norm minimization[16]. Beside the nuclear norm, there are other minimization problems such as log-det penalty function which heuristically lead to the matrix with the minimum rank [17], [18].

Graph-based methods have recently received some attention as a powerful tool for analyzing high-dimensional data which is sampled from a low-dimensional sub-manifold. These methods begin by constructing a sparse graph in which nodes represent input patterns and edges represent neighborhood relations. The resulting graph can be viewed as a discretized approximation of the sub-manifold sampled by the input patterns. From these graphs, one can then construct matrices whose spectral decompositions reveal the low-dimensional structure of the sub-manifold (and sometimes even the dimensionality itself). A detailed survey of many of these algorithms is given in [19]. These algorithms find the low-dimensional embedding using the properties of the manifold. Isomap [20] is based on computing the low-dimensional representation of a high-dimensional data set that most faithfully preserves the pairwise distances between input patterns as measured along the sub-manifold from which they were sampled. Maximum variance unfolding [21] tries to maintain the distances and angles between nearby input patterns. The main goal in locally linear embedding is to keep the local linear structure of nearby input patterns [22]. Finally, Laplacian eigenmaps map nearby input patterns to the nearby outputs by preserving proximity relations [23].

In the context of bandwidth reduction, the point-wise operators of the form x^p for $p \in \mathbb{N}$ are previously studied in [4]. In fact, the point-wise operator x^p extends the bandwidth of a lowpass signal; based on the root-multiplicities of the derivative

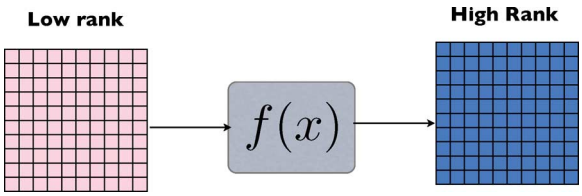

Fig. 2. Distortion function linking matrix $\mathbf{A}$ to $\mathbf{B}$.

of a bandlimited signal, it is shown in [4] that one can estimate integer ps and, therefore, reduce the effective bandwidth.

Our approach differs significantly from the previous works. Our work is closer to the recent work [24] on compressed sensing where the authors consider two correlated signals linked by a sparse filter.

The rest of this paper is organized as follows. In Section II, we introduce formally the problem and state the main results in Section III. Section IV provides simulation and numerical results and Section V concludes the paper.

II. Notations and Statement of the Problem

We begin by definitions and notations which are used throughout the paper.

A. Notations

These notations are consistently used throughout the paper: we use $\odot$ to indicate an element-wise operation on a matrix or a vector; for instance, if $\mathbf{A} = [a_{i,j}]_{m\times n}$ then $\mathbf{A}^{\odot 2} = \left[a_{i,j}^2\right]_{m\times n}$ and $\odot f(\mathbf{A}) = \left[f(a_{i,j})\right]_{m\times n}$. The rank-deficient matrix and its rank are represented by $\mathbf{A}$ and k, respectively, and if $\mathbf{B} = \odot f(\mathbf{A})$, $f(.)$ is called the *distorting operator* relating $\mathbf{B}$ to $\mathbf{A}$. Furthermore, the inverse of $f(.)$ (if exists) is called the rank minimizing operator. We also represent the noisy version of $\mathbf{B}$ by $\tilde{\mathbf{B}}$.

Since we frequently use the determinant of some specific matrices, we define the following notations:

$$T_{\mathbf{B}}(x) = \det\left(\mathbf{B}^{\odot x}\right) \tag{1}$$

$$\bar{T}_{\mathbf{B}}(q_1,\ldots,q_n) = \det\left(\left[(\ln b_{i,j})^{q_i}\right]\right). \tag{2}$$

In this paper, $\mathbb{Z}$ and $\mathbb{Z}^+$ denote the set of integers and non-negative integers (zero included), respectively.

Problem Statement: Let $f(x) = x^p$ be the polynomial distortion function with unknown but fixed $p > 0$. Moreover, let $f(x)$ link the rank deficient square matrix $\mathbf{A}$ to the square matrix $\mathbf{B}$ as follows:

$$\mathbf{B} = \odot f(\mathbf{A}) = \left[f(a_{i,j})\right].$$

The goal is to estimate the point-wise operator $f(x)$ and consequently the rank deficient matrix $\mathbf{A}$ given the matrix $\mathbf{B}$ (see Fig. 2).

III. Main Results

In this section, we explain our results regarding the mentioned problem in form of lemmas and theorems.

A. Monomials With Integer-Valued Power

The following theorem reveals the effect of polynomial-type distorting functions on a rank-deficient matrix:

Theorem 1: Let $\mathbf{A}_{n\times n}$ be a matrix of rank k and p be an arbitrary positive integer, we have

$$\operatorname{rank}\left(\mathbf{A}^{\odot p}\right) \le \min\left\{n, \binom{k+p-1}{p}\right\}. \tag{3}$$

Proof: Since $rank(\mathbf{A}) = k$, we can select k linearly independent row vectors $\{\mathbf{v}_i\}_{i=1}^k$ among the rows of $\mathbf{A}$. This means that the rows of $\mathbf{A}$ can be written as the linear combination of these vectors

$$\begin{aligned}\mathbf{A}_{n\times n} &= \underbrace{\begin{bmatrix} c_{1,1} & \cdots & c_{1,k} \\ \vdots & \ddots & \vdots \\ c_{n,1} & \cdots & c_{n,k}\end{bmatrix}}_{\mathbf{C}_{n\times k}} \underbrace{\begin{bmatrix}\mathbf{v}_1 \\ \vdots \\ \mathbf{v}_k\end{bmatrix}}_{\mathbf{V}_{k\times n}} \\ &= \left[\sum_{l=1}^k c_{i,l} v_{l,j}\right].\end{aligned} \tag{4}$$

Therefore, we have

$$\begin{aligned}\mathbf{A}^{\odot p} &= \left[\Big(\sum_{l=1}^k c_{i,l} v_{l,j}\Big)^p\right] \\ &= \left[\sum_{\substack{p_1+\ldots+p_k=p \\ p_i\in\mathbb{Z}^+}} \binom{p}{p_1,\ldots,p_k} \prod_{l=1}^k (c_{i,l}v_{l,j})^{p_l}\right] \\ &= \sum_{\substack{p_1+\ldots+p_k=p \\ p_i\in\mathbb{Z}^+}} \binom{p}{p_1,\ldots,p_k} \left[\prod_{l=1}^k (c_{i,l}v_{l,j})^{p_l}\right]\end{aligned} \tag{5}$$

where $\left[\prod_{l=1}^k (c_{i,l}v_{l,j})^{p_l}\right]$ stands for the $n\times n$ matrix for which $\prod_{l=1}^k (c_{i,l}v_{l,j})^{p_l}$ is the i,j element and $\mathbb{Z}^+$ represents the set of non-negative integers. Thus

$$\operatorname{rank}(\mathbf{A}^{\odot p}) \le \sum_{\substack{p_1+\ldots+p_k=p \\ p_i\in\mathbb{Z}^+}} \operatorname{rank}\left(\left[\prod_{l=1}^k (c_{i,l}v_{l,j})^{p_l}\right]\right). \tag{6}$$

Note that

$$\begin{aligned}&\Big[\prod_{l=1}^k (c_{i,l}v_{l,j})^{p_l}\Big] = \\ &\begin{bmatrix}\prod_{l=1}^k c_{1,l}^{p_l} \\ \vdots \\ \prod_{l=1}^k c_{n,l}^{p_l}\end{bmatrix} \begin{bmatrix}\prod_{l=1}^k v_{l,1}^{p_l} & \cdots & \prod_{l=1}^k v_{l,n}^{p_l}\end{bmatrix},\end{aligned} \tag{7}$$

which suggests

$$\operatorname{rank}\left(\left[\prod_{l=1}^k (c_{i,l}v_{l,j})^{p_l}\right]\right) = 1. \tag{8}$$

Combining this result with (6), we get

$$\operatorname{rank}(\mathbf{A}^{\odot p}) \leq \sum_{\substack{p_1+\ldots+p_k=p \\ p_i\in\mathbb{Z}^+}} 1 = \binom{k+p-1}{p} \tag{9}$$

and the proof is complete. ■

Remark 1: If $\mathbf{A}$ is a circulant matrix such that the first row has only k nonzero and consecutive DFT coefficients

$$\operatorname{rank}(\mathbf{A}^{\odot p}) \leq p(k-1)+1. \tag{10}$$

According to the properties of the Fourier series, if $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are vectors of the same size such that $\mathbf{a} = \mathbf{b} \odot \mathbf{c}$, the DFT coefficients of $\mathbf{a}$ are obtained by circularly convolving the DFT coefficients of $\mathbf{b}, \mathbf{c}$. In addition, we know that circulant matrices can be diagonally decomposed using DFT and IDFT unitary matrices where the diagonal matrix contains the DFT coefficients of the first row on its main diagonal (eigen-values). This suggests that the eigen-values of $\mathbf{A}^{\odot p}$ are found by p-fold circular convolution of the DFT coefficients of the first row of $\mathbf{A}$ which results in (10).

Remark 2: The distorting operator in Theorem 1 can be considered as a special case of the polynomial operator with $\mathbf{B} = \odot f(\mathbf{A})$ where $f(x) = \sum_{i=0}^{p} f_i x^i$; in fact, $f(x)$ is a monic monomial in Theorem 1. For the general polynomial operator, we have

$$\odot f(\mathbf{A}) = \sum_{i=0}^{p} f_i \mathbf{A}^{\odot i} \tag{11}$$

thus

$$\begin{aligned}\operatorname{rank}\big(\odot f(\mathbf{A})\big) &\leq \sum_{i=0}^{p} \operatorname{rank}\big(\mathbf{A}^{\odot i}\big) \\ &\leq \sum_{i=0}^{p} \binom{k+i-1}{i} = \binom{k+p}{p}.\end{aligned} \tag{12}$$

Remark 3: The bound in Theorem 1 is often achieved. This is helpful for detecting polynomially distorted low-rank matrices: assume $f(x)$ is a polynomial of degree p and $\mathbf{A}_{n\times n}$ is a matrix of rank $k_{\mathbf{A}}$ where $k_{\mathbf{A}}$ is small compared to n. Moreover, let $\mathbf{B} = \odot f(\mathbf{A})$ be the distorted version with rank $k_{\mathbf{B}} \leq \binom{k_{\mathbf{A}}+p}{p}$. If $\mathbf{B}$ is a general rank-deficient matrix of rank $k_{\mathbf{B}}$, the matrix $\mathbf{B}^{\odot i}$ ($i \in \mathbb{N}$) will most likely have the rank $\binom{k_{\mathbf{B}}+i-1}{i}$. However, $\mathbf{B}^{\odot i}$ can be related to $\mathbf{A}$ using a polynomial of degree $p+i$ which implies that the rank of this matrix is upper-bounded by $\binom{k_{\mathbf{A}}+p+i}{p+i}$. It is not hard to check that the latter upper-bound is less than the general upper-bound $\binom{k_{\mathbf{B}}+i-1}{i}$; this fact, simply distinguishes $\mathbf{B}$ from a general rank-deficient matrix subject to the condition that $k_{\mathbf{A}}$ is small enough compared to n, otherwise, $\mathbf{B}^{\odot i}$ or even $\mathbf{B}$ are probably full-rank matrices. Furthermore, the trend of $rank\big(\mathbf{B}^{\odot i}\big)$ with respect to i can reveal the degree of the distorting polynomial (f): if $rank\big(\mathbf{A}^{\odot i}\big) \approx \binom{k_{\mathbf{A}}+i}{i}$, then $\frac{rank\big(\mathbf{B}^{\odot i+1}\big)}{rank\big(\mathbf{B}^{\odot i}\big)} \approx 1 + \frac{k_{\mathbf{A}}}{p+i+1}$ for a range of i values. Now it is easy to estimate p and $k_{\mathbf{A}}$ by having $rank\big(\mathbf{B}^{\odot i}\big)$ for a number of consecutive values of i.

B. Monomial Operators With Real-Valued Power

In the rest of the paper, we focus on the special case of $\mathbf{B} = \mathbf{A}^{\odot p}$ where we assume that $.^p$ is an invertible function (that we can recover the original matrix $\mathbf{A}$). For example, consider the case of $\mathbf{A}^{\odot \frac{3}{5}}$; if the elements of $\mathbf{A}$ are real, both $.^{\frac{3}{5}}$ and $.^{\frac{5}{3}}$ are well defined. For a rank-deficient matrix $\mathbf{A}$, it is very likely that the matrix $\mathbf{B} = \mathbf{A}^{\odot \frac{3}{5}}$ is full-rank. Here, by observing $\mathbf{B}_{n\times n}$, we aim to decide whether this matrix is originated from a rank-deficient matrix using an operator of the form $.^p$ and if yes we would like to estimate p and the rank-deficient matrix. Note that $\mathbf{B}^{\odot \frac{1}{p}}$ is the original rank-deficient matrix; however, if x is a good approximation of $\frac{1}{p}$ (but $x \neq \frac{1}{p}$), $\mathbf{B}^{\odot x}$ is still full-rank; i.e., even good estimates of $\frac{1}{p}$ do not decrease the rank. This difficulty is mainly due to the discrete nature of the rank value; therefore, we should introduce continuous measures to evaluate the rank deficiency of the matrices. For this purpose, we employ the function $T_{\mathbf{B}}(x)$ defined in (1).

It is clear that $\mathbf{B}^{\odot 0} = \mathbf{1}_{n\times n}$ (if there are no zeros in $\mathbf{B}$); thus, if $n > 1$, we have $T_{\mathbf{B}}(0) = 0$. Moreover, if $\mathbf{B} = \mathbf{A}^{\odot p}$ where $\mathbf{A}_{n\times n}$ is a rank-deficient matrix, $T_{\mathbf{B}}(\frac{1}{p}) = T_{\mathbf{A}}(1) = 0$. Note that $T_{\mathbf{B}}(x)$ is a continuous function of x which implies that if x is close to $\frac{1}{p}$, $T_{\mathbf{B}}(x)$ is also close to zero. This means that the roots of $T_{\mathbf{B}}(x)$ (except the trivial case of $x = 0$) play an important role in detecting the rank-deficient structure behind $\mathbf{B}$; nonetheless, finding the roots of $T_{\mathbf{B}}(x)$ is not an easy task. For this purpose we try to approximate the function with its truncated Taylor series.

Lemma 1: The function $T_{\mathbf{B}}(x)$ has convergent Taylor series at each point and the qth Taylor coefficient in series expansion around $x = 0$ ($T_{\mathbf{B}}(x) = \sum_{q=0}^{\infty} t_q x^q$) is given by

$$t_q = \frac{\sum_{\pi\in S_n} \operatorname{sgn}(\pi)\big(\sum_{i=1}^{n} \ln b_{i,\pi(i)}\big)^q}{q!} \tag{13}$$

where S_n denotes the set of all permutations of $\{1,\ldots,n\}$ ($|S_n| = n!$) and for each element $\pi \in S_n$, the sign of π (denoted by $\operatorname{sgn}(\pi)$) is defined as $(-1)^{N(\pi)}$ where $N(\pi)$ represents the number of inversions in the permutation.

Proof: We consider the expanded version of the determinant function to find the Taylor coefficients

$$\begin{aligned}T_{\mathbf{B}}(x) = \det\left[b_{i,j}^x\right] &= \sum_{\pi\in S_n} \operatorname{sgn}(\pi) \prod_{i=1}^{n} b_{i,\pi(i)}^x \\ &= \sum_{\pi\in S_n} \operatorname{sgn}(\pi) e^{x\sum_{i=1}^{n} \ln b_{i,\pi(i)}} \\ &= \sum_{\pi\in S_n} \operatorname{sgn}(\pi) \sum_{q=0}^{\infty} \frac{\big(\sum_{i=1}^{n} \ln b_{i,\pi(i)}\big)^q}{q!} x^q \\ &= \sum_{q=0}^{\infty} \frac{x^q}{q!} \sum_{\pi\in S_n} \operatorname{sgn}(\pi)\big(\sum_{i=1}^{n} \ln b_{i,\pi(i)}\big)^q.\end{aligned} \tag{14}$$

The last equation shows the Taylor coefficients. Note that since $\det(\mathbf{M}_1 + \mathbf{M}_2)$ is not necessarily equal to $\det(\mathbf{M}_1) +$

$\det(\mathbf{M}_2)$, the term $\frac{\partial^q}{\partial x^q} T_{\mathbf{B}}(x)$ is not necessarily $\det\left(\frac{\partial^q}{\partial x^q}\mathbf{B}^{\odot x}\right)$ (which shows the importance of (14) and its derivation). Also note that, due to the representation of $T_{\mathbf{B}}(x)$ as the sum of finite number of exponentials, the Taylor series is convergent for every x. ■

Although Lemma 1 describes the Taylor coefficients, there is a more useful representation of the terms which we later exploit to demonstrate bounds on the truncation error.

Theorem 2: The Taylor series of $T_{\mathbf{B}}(x)$ has $n-1$ vanishing terms and the expansion can be reformulated as

$$T_{\mathbf{B}}(x) = \sum_{q=n-1}^{\infty} \frac{x^q}{q!} \sum_{\substack{q_1,\ldots,q_n=q \\ q_i\in\mathbb{Z}^+}} \binom{q}{q_1,\ldots,q_n} \times \bar{T}_{\mathbf{B}}(q_1,\ldots,q_n). \tag{15}$$

where the function $\bar{T}_{\mathbf{B}}$ is previously defined in (2).

Proof: For a given permutation $\pi \in S_n$ we know

$$\Big(\sum_{i=1}^{n} \ln b_{i,\pi(i)}\Big)^q = \sum_{\substack{q_1,\ldots,q_n=q \\ q_i\in\mathbb{Z}^+}} \binom{q}{q_1,\ldots,q_n} \prod_{i=1}^{n} \big(\ln b_{i,\pi(i)}\big)^{q_i}. \tag{16}$$

Now, by using Lemma 1 and the definition (2) in (16), we have

$$T_{\mathbf{B}}(x) = \sum_{q=0}^{\infty} \frac{x^q}{q!} \sum_{\substack{q_1+\ldots+q_n=q \\ q_i\in\mathbb{Z}^+}} \binom{q}{q_1,\ldots,q_n} \times \bar{T}_{\mathbf{B}}(q_1,\ldots,q_n). \tag{17}$$

If there are 2 or more zeros in an n-tuple $(q_1,\ldots,q_n)$ (non-negative integers q_i where $\sum_{i=1}^{n} q_i = q$), then $\left[\left(\ln b_{i,j}\right)^{q_i}\right]$ includes two or more rows completely filled with ones and, thus, $\bar{T}_{\mathbf{B}}(q_1,\ldots,q_n) = 0$. Hence, only n-tuples appear in the coefficients that contain at most one zero. Consequently, the coefficients of x^q for $q < n-1$ vanish and the Taylor series start with x^{n-1}; i.e., $x = 0$ is a multiple root of $T_{\mathbf{B}}$ with multiplicity at least $n-1$. ■

Remark 4: The summations in (17) for finding the Taylor coefficients involve an increasing number of summands which becomes computationally impractical for large n. In these cases, it might be possible to numerically approximate the Taylor coefficients; for instance, one might look for the Fourier series of the function $T_{\mathbf{B}}\left(e^{j\omega}\right)$ (it is easy to show that $T_{\mathbf{B}}(.)$ is an analytic function and it is well-defined over the complex plane) with respect to ω which yields the Taylor coefficients.

Theorem 2 and Lemma 1 suggest that in order to have a good approximation of $T_{\mathbf{B}}$ at a given x, it suffices to include only a finite number of the terms in the Taylor series; i.e., for a limited range of x, $T_{\mathbf{B}}$ can be properly approximated by a polynomial of finite degree. This is our main key to evaluate the roots of $T_{\mathbf{B}}(x)$; we truncate the Taylor series at the Nth term and find the roots of the resultant polynomial. We then calculate the range of x for which the N-term approximation of the Taylor series yields acceptable (prespecified upper-bound for error) results. Finally, we discard those roots which do not belong to this range.

In the following theorem, we demonstrate an upper-bound on the truncation error of the Taylor series.

Theorem 3: Let $E_N(x) = \sum_{q=N+1}^{\infty} t_q x^q$ denote the truncation error of the Taylor series approximated by the first $N+1$ terms and let $M_{\mathbf{B}}$ be the maximum modulus of the elements of $\odot \ln \mathbf{B}$. For an arbitrary value x and $N \geq \lceil eM_{\mathbf{B}}x \rceil$, we have

$$|E_N(x)| \leq \frac{n^{\frac{n}{2}}\left(eM_{\mathbf{B}}x\right)^{N+1}}{\sqrt{2\pi}(N+1)^{N+1.5-n}\left(1-\frac{eM_{\mathbf{B}}x}{N+1}\right)}. \tag{18}$$

Proof: Using Hadamard's inequality, we have

$$\begin{aligned} \left|\bar{T}_{\mathbf{B}}(q_1,\ldots,q_n)\right| &= \left|\det\left(\left[(\ln b_{i,j})^{q_i}\right]\right)\right| \\ &\leq \prod_{i=1}^{n}\left(\sum_{j=1}^{n} |\ln b_{i,j}|^{2q_i}\right)^{0.5} \\ &\leq \prod_{i=1}^{n} \sqrt{n} M_{\mathbf{B}}^{q_i} = n^{\frac{n}{2}} M_{\mathbf{B}}^{\sum_{i=1}^{n} q_i}. \end{aligned} \tag{19}$$

Note that

$$\begin{aligned} |E_N(x)| &= \Big|\sum_{q=N+1}^{\infty} \frac{x^q}{q!} \sum_{\substack{q_1,\ldots,q_n=q \\ q_i\in\mathbb{Z}^+}} \binom{q}{q_1,\ldots,q_n} \times \bar{T}_{\mathbf{B}}(q_1,\ldots,q_n)\Big| \\ &\leq \sum_{q=N+1}^{\infty} \frac{x^q}{q!} \sum_{\substack{q_1,\ldots,q_n=q \\ q_i\in\mathbb{Z}^+}} \binom{q}{q_1,\ldots,q_n} \times \left|\bar{T}_{\mathbf{B}}(q_1,\ldots,q_n)\right|. \end{aligned} \tag{20}$$

Therefore, from (19), we get

$$\begin{aligned} |E_N(x)| &\leq \sum_{q=N+1}^{\infty} \frac{x^q}{q!} \sum_{\substack{q_1,\ldots,q_n=q \\ q_i\in\mathbb{Z}^+}} \binom{q}{q_1,\ldots,q_n} n^{\frac{n}{2}} M_{\mathbf{B}}^{q} \\ &= n^{\frac{n}{2}} \sum_{q=N+1}^{\infty} \frac{\left(M_{\mathbf{B}}x\right)^q q^n}{q!}. \end{aligned} \tag{21}$$

Employing $n! > \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$, we obtain

$$\begin{aligned} |E_N(x)| &\leq \frac{n^{\frac{n}{2}}(eM_{\mathbf{B}}x)^{n-0.5}}{\sqrt{2\pi}} \sum_{q=N+1}^{\infty} \left(\frac{eM_{\mathbf{B}}x}{q}\right)^{q+0.5-n} \\ &\leq \frac{n^{\frac{n}{2}}(eM_{\mathbf{B}}x)^{n-0.5}}{\sqrt{2\pi}} \sum_{q=N+1}^{\infty} \left(\frac{eM_{\mathbf{B}}x}{N+1}\right)^{q+0.5-n}. \end{aligned} \tag{22}$$

Due to our assumption that $N \geq \lceil eM_{\mathbf{B}}x \rceil$, we have $\frac{eM_{\mathbf{B}}x}{N+1} < 1$, and therefore

$$\sum_{q=N+1}^{\infty} \left(\frac{eM_{\mathbf{B}}x}{N+1}\right)^{q+0.5-n} = \left(\frac{eM_{\mathbf{B}}x}{N+1}\right)^{N+1.5-n} \frac{1}{1-\frac{eM_{\mathbf{B}}x}{N+1}} \tag{23}$$

which completes the proof. ■

It should be mentioned that multiplying or dividing the elements of $\mathbf{B}$ by a scalar does not change the roots of $T_{\mathbf{B}}(x)$; however, it does affect the value $M_{\mathbf{B}}$ and, consequently, the upper-bound for the truncation error. Thus, in order to improve the numerical results, it is desirable to find the optimum scalar

TABLE I
ESTIMATING THE RANK MINIMIZING OPERATOR

Input:

- $\mathbf{B}_{n\times n}$.

Outputs:

- p, the power of the rank minimizing monomial.
- k_{max}, maximum possible rank of the original matrix.
- $\hat{\mathbf{A}}_{n\times n}$, estimated rank-deficient matrix.

Steps:

1) Initialize N, the number of Taylor coefficients to keep, and x_{max}, the maximum expected root of $T_{\mathbf{B}}(.)$.
2) [optional] Instead of working with $\mathbf{B}$ it is better to divide all the elements by $M_{\mathbf{B}}$, the geometric mean between the minimum and the maximum entries of $\mathbf{B}$.
3) Compute the first N Taylor coefficients (coefficients of $x^{n-1}, \ldots, x^{n+N-2}$) of $T_{\mathbf{B}}(.)$, either directly by using (17) or indirectly by an approximation method (for large n).
4) Find the roots of the truncated Taylor series by one of the root-finding methods such as *Splitting Circle Method* [25].
5) Check if the roots belong to the confidence interval of the N-term Taylor approximation. The safest way is to use the upper-bound on the error but more practical is to check if it is also the root of the truncated Taylor series with $N-1$ and $N-2$ terms.
6) If there are $l > 1$ acceptable roots (including $x = 0$), set $x_{max} \leftarrow \frac{l+1}{l-1} r_{max}$ where r_{max} is the maximum acceptable root; otherwise, do not change x_{max}.
7) If N is large enough to yield good enough approximations for the whole range of $[0, x_{max}]$, terminate the loop; otherwise increase N and return to step 2.
8) Is there any arithmetic progression of length at least 3 among the acceptable roots?
   - No (or if there are no non-trivial roots): The matrix $\mathbf{B}$ is unlikely to be produced by a low rank matrix.
   - Yes: set $\frac{1}{p}$ as the smallest positive element in the arithmetic progression, set k_{max} as the maximum value k such that $\binom{l+k-2}{l-1} < n$ where l is the length of the arithmetic progression, and define $\hat{\mathbf{A}}$ by setting the $n - k_{max}$ smallest singular values of $\mathbf{B}^{\odot\frac{1}{p}}$ to zero.

value (it is easy to check that this value is the inverse of the geometric mean between the minimum and maximum values in the main matrix).

The last thing to discuss is the rank (k) of the original matrix $\mathbf{A}_{n\times n}$ from which the matrix $\mathbf{B}_{n\times n}$ is generated. As Theorem 1 indicates, for all values of m that $\binom{m+k-1}{m} < n$ (let m_{max} denote the maximum of these m's), the matrix $\mathbf{A}^{\odot m}$ is rank-deficient. Thus, if $\mathbf{B} = \mathbf{A}^{\odot p}$ ($p \in \mathbb{R}$), all the elements of the set $\{\frac{i}{p}\}_{i=0}^{m_{max}}$ are the roots of $T_{\mathbf{B}}(x)$; i.e., the set of the roots contains an arithmetic progression of length $m_{max}+1$ starting from zero and with step size $\frac{1}{p}$. In fact this is a helpful tool for both detecting the rank-deficient structure behind $\mathbf{B}$ and denoising the estimates of p or its inverse $\frac{1}{p}$; i.e., since we are using the truncated version of the Taylor series (probably using even noisy matrix elements), the roots are not exact but a pattern similar to an arithmetic progression helps denoising the step size and consequently recovering the original matrix. Furthermore, the length of the detected arithmetic progression can be used to estimate the original rank value (k). If, the length is l we have

$$\binom{l+k-2}{l-1} < n \leq \binom{l+k-1}{l}. \tag{24}$$

It is again useful to denoise the estimate of the matrix $\mathbf{A}$ from $\mathbf{B}$. In fact, we can map the estimated matrix $\hat{\mathbf{A}}$ to the closest rank-deficient matrix using the upper-bound for k by means of setting some of the smallest singular values of $\hat{\mathbf{A}}$ to zero. Table I shows the step-wise procedure of the proposed method.

1) A Special Case: We showed in Remark 1 that circulant matrices are special cases for the monomial distorting operators with integer power. Here we also develop special methods for finding the real-valued power of a distorting operator acting on a circulant matrix. Let us denote the first row of the distorted circulant matrix $\mathbf{B}$ (which is also circulant) as $[b_1, \ldots, b_n]$. We know

$$\mathbf{B}^{\odot x} = \begin{bmatrix} b_1^x & b_2^x & \ldots & b_n^x \\ b_n^x & b_1^x & \ldots & b_{n-1}^x \\ \vdots & \vdots & \ddots & \vdots \\ b_2^x & b_3^x & \ldots & b_1^x \end{bmatrix} = \mathbf{D} \begin{bmatrix} \hat{b}_1(x) & 0 & \ldots & 0 \\ 0 & \hat{b}_2(x) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \hat{b}_n(x) \end{bmatrix} \mathbf{D}^H \tag{25}$$

where $\mathbf{D}$ represents the unitary DFT matrix and $\hat{b}_i(x)$'s are the DFT coefficients of the vector $[b_1^x, \ldots, b_n^x]$

$$\begin{bmatrix} \hat{b}_1(x) \\ \vdots \\ \hat{b}_n(x) \end{bmatrix} = \mathbf{D} \begin{bmatrix} b_1^x \\ \vdots \\ b_n^x \end{bmatrix} = \sum_{k=0}^{\infty} \mathbf{D} \begin{bmatrix} (\ln b_1)^k \\ \vdots \\ (\ln b_n)^k \end{bmatrix} \frac{x^k}{k!}. \tag{26}$$

From (25) it follows that

$$T_{\mathbf{B}}(x) = \det\left(\mathbf{B}^{\odot x}\right) = \prod_{i=1}^{n} \hat{b}_i(x). \tag{27}$$

This shows that we have a simple factorization of the function $T_{\mathbf{B}}(x)$ in which we can simply calculate the Taylor coefficients of each term; according to (26), the kth Taylor coefficient of $\hat{b}_i(x)$ is equal to the ith DFT coefficient of $\frac{1}{k!}[(\ln b_1)^k, \ldots, (\ln b_n)^k]$. In simple words, instead of approximating $T_{\mathbf{B}}(x)$ with its truncated Taylor series, we can approximate it by truncating the Taylor series of its components. The advantage is that unlike (17) where the summation is over a large and increasing number of summands, the Taylor coefficients of the components can be obtained using the Fast Fourier Transform (FFT).

IV. NUMERICAL RESULTS

For the purpose of simulation results, we have implemented the algorithms for both the integer and the real-valued powers in MATLAB. In the first scenario, we have generated a low rank matrix by multiplying two 1000×5 and 5×1000 random matrices with i.i.d. elements uniformly distributed in $[0, 1]$. The resultant matrix which is of rank 5 is used as the original low rank matrix while its distorted version is constructed as $\mathbf{B} = \mathbf{A}^{\odot 2}$. According to Theorem 1, we should have $rank\left(\mathbf{A}^{\odot p}\right) \leq \binom{4+p}{p}$ for a positive integer p; in fact, Fig. 3 confirms that the equality happens for this matrix.[1] On the other hand, for the matrix $\mathbf{B}$ where we have $rank(\mathbf{B}) = 15$, there is a gap between the upper bound $\binom{14+p}{p}$ and the actual rank. It is easy to check that the rank curve of $\mathbf{B}^{\odot p}$ coincides with $\binom{4+2p}{2p}$; this is in fact the key to find the rank of $\mathbf{A}$ by having $\mathbf{B}$ (as explained in Remark 3).

[1]Due to the numerical errors, MATLAB's rank function is inaccurate for large rank-deficient matrices; for determining the rank, we have used the gap between the singular values to distinguish between the zero and nonzero values.

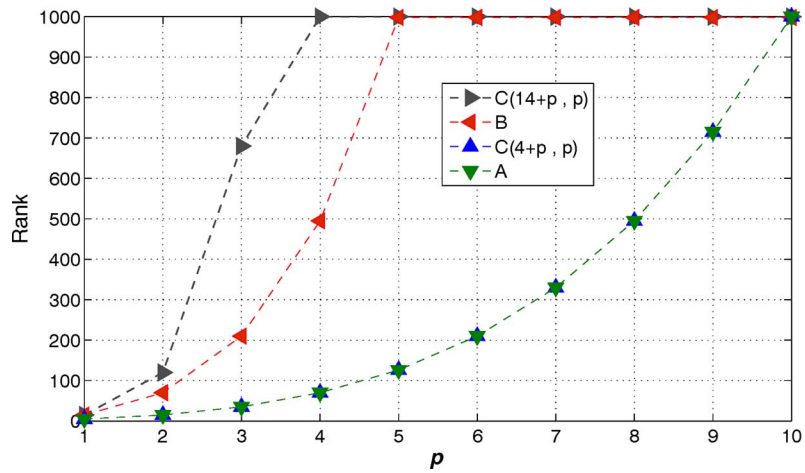

Fig. 3. Rank of the 1000×1000 matrices $\mathbf{A}^{\odot p}, \mathbf{B}^{\odot p}$ versus integer p where $\mathbf{B} = \mathbf{A}^{\odot 2}$. The rank of the matrices $\mathbf{A}$ and $\mathbf{B}$ are 5 and 15, respectively, which imply upper bounds of the form $\binom{4+p}{p}$ and $\binom{14+p}{p}$ on the ranks of $\mathbf{A}^{\odot p}, \mathbf{B}^{\odot p}$.

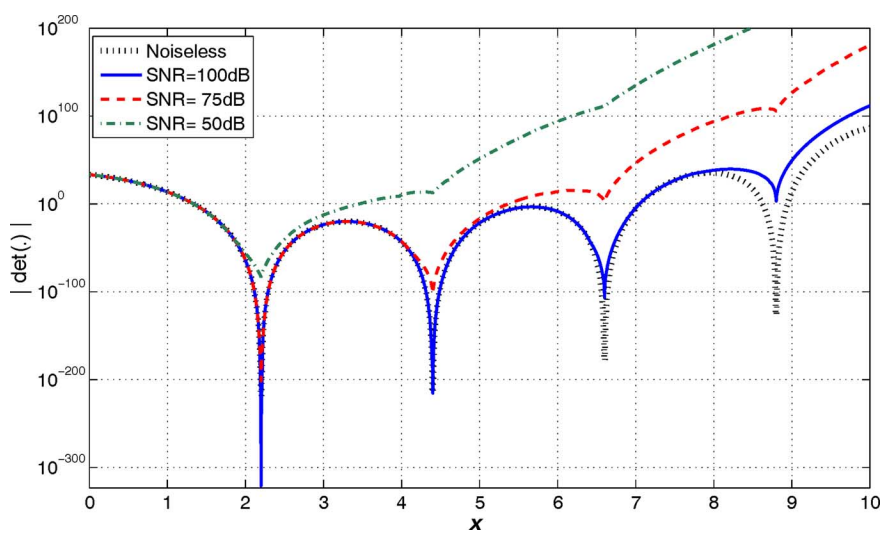

Fig. 4. Determinant of $B^{\odot x}$ and its noisy versions versus x; the large negative peaks in the log scale indicate the existence of roots (rank minimizing power).

In the second scenario, we have considered a circulant rank-deficient matrix. The first row of this 100×100 matrix is generated in such a way that it has only 5 nonzero and consecutive DFT coefficients; in fact the nonzero coefficients are generated by realizations of i.i.d. zero-mean normal random variables with variance 10. The applied distorting operator here is $.^{\odot \frac{5}{11}}$; i.e., $p = \frac{5}{11}$ and the original rank-deficient matrix $\mathbf{A}_{100\times 100}$ is recovered by taking the elements of the distorted matrix $\mathbf{B}_{100\times 100}$ to the power $\frac{1}{p} = 2.2$. In order to include the noise effect, in addition to implementing the techniques on the noiseless matrix, the noisy versions of $\mathbf{B}$ are also studied: the elements of $\mathbf{B}$ are subject to Additive White Gaussian Noise (AWGN) with difference noise variances resulting in SNR values 100 *dB*, 75 *dB* and 50 *dB*. Before we conduct the experiments to find the rank minimizing operator, it is interesting to examine the suitability of the determinant function for this purpose. Fig. 4 depicts the curve of the function $T_{\mathbf{B}}(x)$ and the corresponding functions for the noisy versions of $\mathbf{B}$. As discussed before, we expect the functions to have roots at $x = 2.2, 4.4, 6.6, \ldots$; nevertheless, due to the additive noise, the roots are perturbed and they may no longer lay on the real axis. As a result, instead of roots, the functions have local minima at the desired values which can still be used for approximating the original roots. Now it is interesting to replace the determinant measure by the condition number of the matrix; the roots of the determinant are mapped to the sharp peaks in the curve of the condition number. Although this is the case for the noiseless matrix, Fig. 5 reveals that there are many fake peaks in the curves of the noisy matrices. In other words, it is more difficult to locate the original roots by focusing on the condition number when the observations are noisy.

Figs. 6–9 show the approximations of the determinant functions obtained by truncating the Taylor series as explained in Section III-B.I. The curves for the truncated Taylor approximations with degrees higher than 40 almost coincide and, therefore, are not reported in these figures. This fact shows that for the mentioned matrix, only 40 Taylor coefficients are sufficient for an accurate estimation of the determinant for $x \in [0, 10]$ and consequently, the roots in this interval.

To evaluate the computational complexity of the propose method, note that as explained in Section III-B.I, for each

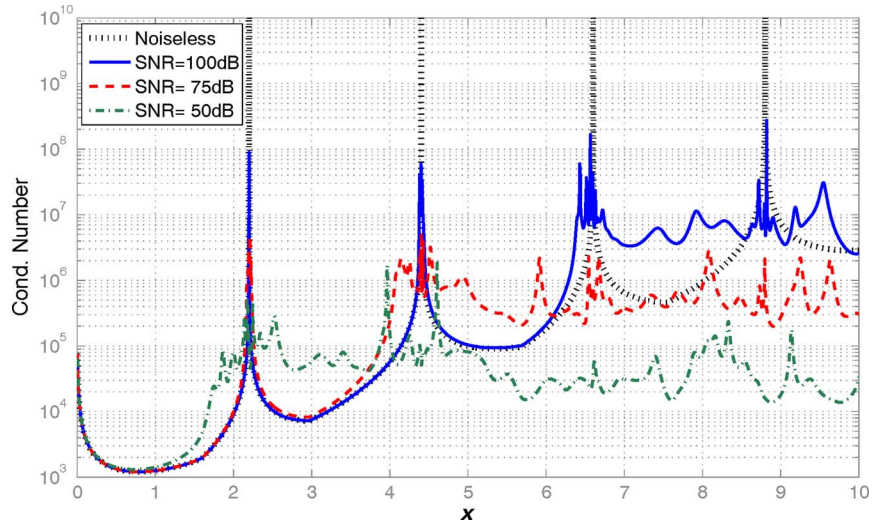

Fig. 5. Condition number of $B^{\odot x}$ and its noisy versions versus x; the large positive peaks might indicate the existence of roots (rank minimizing power).

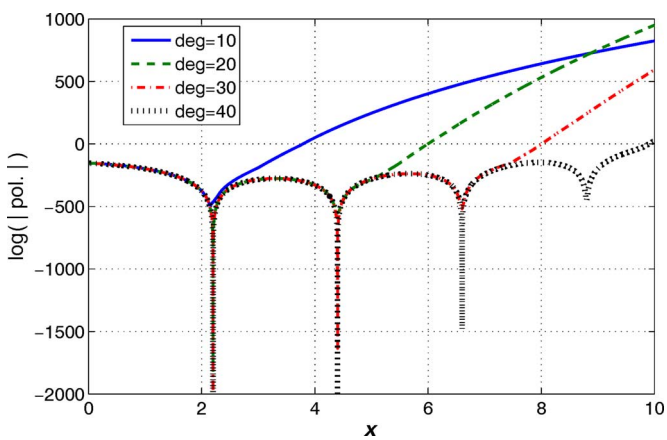

Fig. 6. Taylor approximations of the determinant for the noiseless 100×100 circulant matrix with various number of terms.

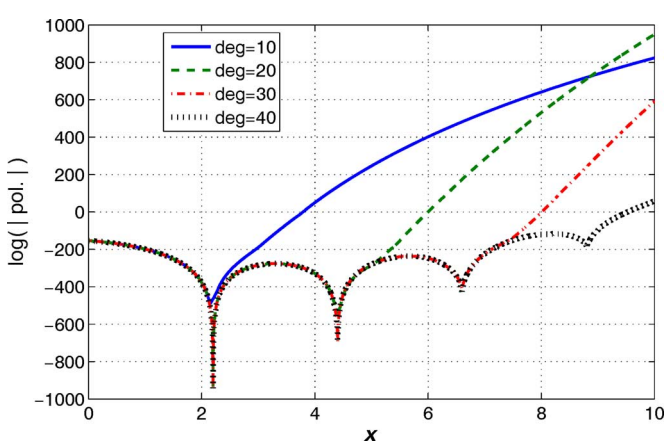

Fig. 7. Taylor approximations of the determinant for the noisy 100×100 circulant matrix with $SNR = 100\ dB$ and various number of terms.

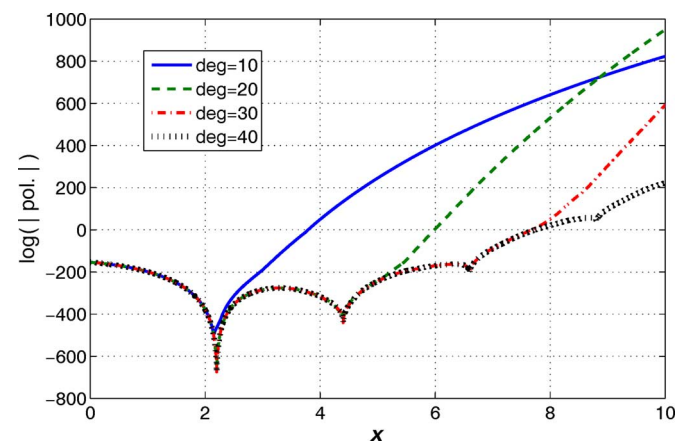

Fig. 8. Taylor approximations of the determinant for the noisy 100×100 circulant matrix with $SNR = 70\ dB$ and various number of terms.

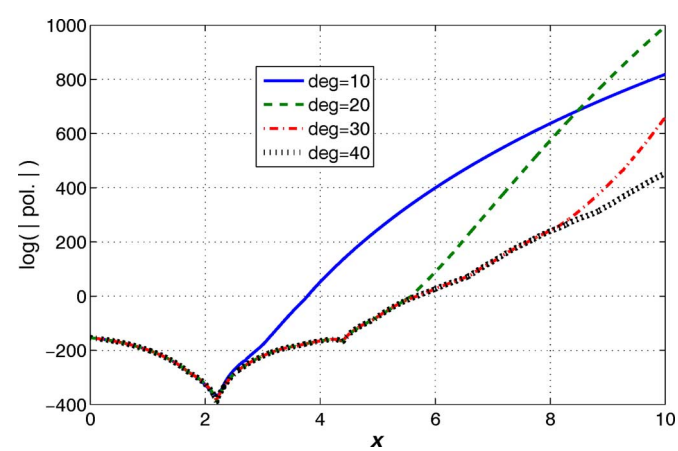

Fig. 9. Taylor approximations of the determinant for the noisy 100×100 circulant matrix with $SNR = 50\ dB$ and various number of terms.

Taylor coefficient of an $n \times n$ circulant matrix we require an FFT operation which requires $\mathcal{O}(n \log n)$ multiplications and suggests the complexity of $\mathcal{O}(dn \log n)$ for a degree d approximation of the Taylor series. Generally, finding the roots of a polynomial of degree d with ϵ accuracy requires $\mathcal{O}\big(d^2 \log d\ |\log \epsilon|\big)$ operations [26] which shows that the total cost of the proposed method is $\mathcal{O}\big(d^2 \log d\ |\log \epsilon| + dn \log n\big)$. Now to compare this result with the brute-force search, it should be mentioned that each determinant evaluation for a

circulant matrix costs $\mathcal{O}(n^2 \log n)$ multiplications (it is more than $\mathcal{O}(n^3)$ for general matrices) and to have ϵ accuracy in the roots, $\mathcal{O}(\frac{1}{\epsilon})$ determinant evaluations are needed; i.e., the complexity of $\mathcal{O}(\frac{n^2 \log n}{\epsilon})$ for the whole procedure. It is evident that for the cases in which a good Taylor approximation is obtained with $d < n$ (such as the mentioned example where $n = 100$ and $d = 40$), the proposed method is much faster than the simple brute-force search. It is also important to point out that the desired accuracy in the roots (ϵ) influences the complexities with different order of magnitudes; the complexity of the proposed method scales with $|\log \epsilon|$ (which is inherent in all common root-finding methods) while the scaling factor is $\frac{1}{\epsilon}$ in the brute-force search method.

V. Conclusion

We have considered the problem of detecting the existence of a function of the form $f(x) = x^p$ ($p \in \mathbf{R}$) that converts a given matrix $\mathbf{B}_{n\times n}$ into a rank-deficient matrix $\mathbf{A}$. Furthermore, we have introduced means to estimate p by considering a determinant-type function of $\mathbf{B}$; this function has p as its root and the problem of estimating p reduces to the problem of finding the roots of this generally nonpolynomial function. We have shown that when $rank(\mathbf{A}) \ll n$, there exists an arithmetic progression among the roots starting from zero with step size p. Since finding the roots of this function is computationally hard, we used its truncated Taylor series which are shown to be good approximations. Simulation results show that these polynomials yield acceptable results for finding the roots when the noise level on the elements of $\mathbf{B}$ is small.

References

[1] I. Jolliffe, *Principal Component Analysis*. New York: Springer, 1986.

[2] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.

[3] S. Oh, A. Karbasi, and A. Montanari, "Sensor network localization from local connectivity : Performance analysis for the mds-map algorithm," presented at the IEEE Information Theory Workshop (ITW2010), 2010.

[4] F. Marvasti and A. Jain, "Zero-crossings, bandwidth compression and restoration of bandlimited signals distorted by nonlinear systems," *J. Opt. Soc. Amer.*, vol. 3, no. 5, pp. 651–654, 1986.

[5] R. Calderbank, S. Jafarpour, and R. Schapire, "Compressed Learning: Universal Sparse Dimensionality Reduction and Learning in the Measurement Domain," Tech. Rep. 2009.

[6] M. W. Berry, Z. Drmac, and E. R. Jessup, "Matrices, vector spaces, and information retrieval," *SIAM Rev.*, vol. 41, no. 2, pp. 335–362, 1999.

[7] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.

[8] R. H. Keshavan, A. Montanari, and S. Oh, Matrix Completion from Noisy Entries, 2009 [Online]. Available: http://arxiv.org/PScache/arxiv/pdf/0906/0906.2027v1.pdf

[9] A. Karbasi, S. Oh, R. Parhizkar, and M. Vetterli, "Ultrasound tomography calibration using structured matrix completion," presented at the Int. Congr. Acoustics (ICA2010) 2010.

[10] J. Chen, K. Yao, and R. E. Hudson, "Source localization and beamforming," *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 30–39, May 2002.

[11] T. Rappaport, *Wireless Communications: Principles and Practice*. Upper Saddle River, NJ: Prentice-Hall, 2001.

[12] T. Cox and M. A. A. Cox, *Multidimensional Scaling*. New York: Chapman and Hall, 1994.

[13] P. Drineas, A. Javed, M. Magdon-Ismail, G. Pandurangant, R. Virrankoski, and A. Savvides, "Distance matrix reconstruction from incomplete distance information for sensor network localization," in *Proc. Sensor and Ad-Hoc Communications and Networks Conf. (SECON)*, Sep. 2006, vol. 2, pp. 536–544.

[14] C. H. Papadimitriou, P. Raghavan, H. Tamaki, and S. Vempala, "Latent semantic indexing: A probabilistic analysis," *J. Comput. Syst. Sci.*, vol. 61, no. 2, pp. 217–235, 2000.

[15] M. Fazel, "Matrix Rank Minimization with Applications," Ph.D. dissertation, Stanford Univ., , 2002.

[16] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, 2009.

[17] M. Fazel, H. Hindi, and S. Boyd, "Log-det heuristic for matrix rank minimization with applications to hankel and euclidean distance matrices," in *Proc. American Control Conf.*, 2003, vol. 3, pp. 2156–2162.

[18] M. Fazel, H. Hindi, and S. Boyd, "Rank minimization and applications in system theory," in *Proc. American Control Conf.*, 2004, vol. 4, pp. 3273–3278.

[19] J. C. Burges, *Geometric Methods for Feature Extraction and Dimensional Reduction in Data Mining and Knowledge Discovery Handbook*. Boston, MA: Kluwer, 2004.

[20] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.

[21] Q. Weinberger, B. D. Packer, and L. K. Saul, "Nonlinear dimensionality reduction by semidefinite programming and kernel matrix factorization," presented at the Workshop on AI and Statistics (AISTATS2005), Jan. 2005.

[22] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[23] M. Belkin, I. Matveeva, and P. Niyogi, "Regularization and semisupervised learning on large graphs," in *Proc. Annu. Conf. Computational Learning Theory (COLT2004)*, 2004, vol. 3120, pp. 624–638.

[24] A. Hormati, O. Roy, Y. M. Lu, and M. Vetterli, "Distributed sampling of correlated signals linked by sparse filtering: Theory and applications," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1095–1109, Mar. 2010.

[25] V. Y. Pan, "Solving a polynomial equation: Some history and recent progresses," *SIAM Rev.*, vol. 39, no. 2, pp. 187–220, Jun. 1997.

[26] M. H. Kim and S. Sutherland, "Polynomial root-finding algorithms and branched covers," *SIAM J. Comput.*, vol. 23, pp. 415–436, 1994.

Arash Amini received B.Sc., M.Sc., and Ph.D. degrees in electrical engineering (Communications and Signal Processing) in 2005, 2007, and 2011, respectively, and the B.Sc. degree in petroleum engineering (Reservoir) in 2005, all from Sharif University of Technology (SUT), Tehran, Iran. Since April 2011, he has been a postdoc researcher in Biomedical Imaging Group (BIG),École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. His research interests include different aspects of sampling, especially compressed sensing.

Amin Karbasi (S'10) received the B.Sc. degree in electrical engineering in 2004 and M.Sc. degree in communication systems in 2007 from EPFL , Lausanne, Switzerland. Since March 2008, he has been a Ph.D. student at EPFL. He was the recipient of the ICASSP 2011 best student paper award and ACM/Sigmetrics 2010 best student paper award. His research interests include graphical models, large scale networks, compressed sensing and information theory.

Farokh Marvasti (SM'70) received his B.S., M.S., and Ph.D. degrees all from Rensselaer Polytechnic Institute in 1970, 1971, and 1973, respectively. He has worked, consulted and taught in various industries and academic institutions since 1972, among which include Bell Labs, University of California Davis, Illinois Institute of Technology, University of London, and King's College. He was an editor and associate editor for the IEEE Transactions on Communications and the IEEE Transactions on Signal Processing from 1990–1997. He has about 100 Journal publications and has written several reference books; he also holds several international patents. His authored the book *Nonuniform Sampling: Theory and Practice* (Kluwer, 2001). He was also a guest editor for the Special Issue on Nonuniform Sampling of the *Sampling Theory & Signal and Image Processing* journal, May 2008. Besides being the co-founders of two international conferences (ICT and SampTA), he has been an organizer and special session chair of many IEEEE conferences, including ICASSP conferences. Dr. Marvasti is currently a Professor at the Sharif University of Technology and the Director of the Advanced Communications Research Institute (ACRI) and Head of the Center for Multi-Access Communications Systems.