A STUDY OF THE SINGULAR SPECTRUM ANALYSIS WITH THE GLOBAL OPTIMIZATION TECHNIQUES

DRAFT AS OF November 10, 2012

MOODY T. CHU*, MATTHEW M. LIN†, AND LIQI WANG‡

Abstract. Singular spectral analysis has recently become an attractive tool in a broad range of applications. Its main mechanism of alternating between rank reduction and Hankel projection, however, deserves further mathematical justification. Exploiting a natural parametrization of a general Hankel matrix via its Vandermonde factorization, this paper recasts the numerical scheme for the singular spectral analysis as a semi-linear least squares problem over a compact feasible set. Such a formulation makes it possible to employ global optimization techniques to find the absolute best approximation which then is used to assess the quality of the approximation produced from the singular spectral analysis. **What is the main conclusion at the end of this investigation?**

Key words. singular spectral analysis, time series, Hankel operator, semi-infinite matrix, low rank approximation, Vandermonde factorization, global optimization

AMS subject classifications. 47B35, , 65F25, 65F30, 15A23, 15B05

1. Introduction. Singular spectrum analysis (SSA) is a novel non-parametric technique based on principles of multivariate statistics. Its utilization is more as an exploratory model-building tool than an affirmative procedure. The development of the SSA is usually attributed to Broomhead and King [7] whose original intention was to identify qualitative information in the analysis of dynamical systems. Actually, similar ideas also came to surface independently at about the same time in other disciplines, such as that in [11] for the dimensional analysis of weather and climate attractors. Over years, the methodologies and applications of the SSA have been refined and broadened significantly in the literature. A quick Google search for the term "singular spectral analysis" shows up more than 1.5 million results as of the date this note is written. Far from being complete, we mention only a few references [2, 4, 9, 14, 25, 36, 39] representing some interesting applications. We also suggest two monographs [10, 16] for an elementary introduction to this subject.

Given a finite time series $\mathbf{z} = (z_0, \ldots, z_n)$ of length $n+1$, the practicalities of the SSA usually consist of four basic parts [16]:

Embedding. Choose an integer $1 \le p < n$, referred to as the *embedding dimension*, and let $q = n - p + 2$. Define *p-lagged vectors* $\mathbf{z}_0, \ldots, \mathbf{z}_{q-1}$ by $\mathbf{z}_i = [z_i, \ldots, z_{i+p-1}]^\top \in \mathbb{C}^p$ and the associated *trajectory matrix* Z by

$$Z := [\mathbf{z}_0, \ldots, \mathbf{z}_{q-1}] = \begin{bmatrix} z_0 & z_1 & \cdots & z_{q-1} \\ z_1 & z_2 & \cdots & z_q \\ \vdots & \vdots & \ddots & \vdots \\ z_{p-1} & z_p & \cdots & z_n \end{bmatrix}. \tag{1.1}$$

Note that Z is an $p \times q$ Hankel matrix. This process of embedding $\mathbf{z}$ into Z, fundamental in time series analysis, creates a handle for manipulating rank reduction.

The choice of p has been more or less on a heuristic basis [10, 16, 17]. The recommended threshold is somewhere between $n/4$ and $n/2$.

*Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205. (chu@math.ncsu.edu) This research was supported in part by the National Science Foundation under grant DMS-1014666.

†Department of Mathematics, National Chung Cheng University, Chiayi 62012 Taiwan. (mlin@math.ccu.edu.tw). This research was supported in part by the National Science Council of Taiwan under grant 99-2115-M-194-010-MY2.

‡Department of Mathematics, Dalian University of Technology, (lizzy@mail.dlut.edu.cn)

Rank reduction. Suppose that

$$Z = U\Sigma V^* = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^* \tag{1.2}$$

is the singular value decomposition (SVD) of the trajectory matrix Z, where nonzero singular values are arranged in descending order $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r$ and $r = \text{rank}(Z)$. A fundamental mechanism used for rank reduction is the truncated singular value decomposition. A truncated SVD of Z is a partial sum of the form

$$\widetilde{Z} := \sum_{i=1}^{d} \sigma_i \mathbf{u}_i \mathbf{v}_i^*,$$

where $d \leq r$ is a preselected positive integer serving as the target rank for the analysis.

There is a statistical meaning of the truncated singular value decomposition. With regard to the *lag-covariance* matrix $Z^\top Z$ when columns of Z have been centered, it can be argued that $\widetilde{Z}$ represents the best, unbiased, linear, minimum-variance estimate of Z among all possible d-dimensional sample subspaces in $\mathbb{C}^p$. Because of this, if d is properly chosen, then those components of Z corresponding to "small" singular values are considered as less significant and often are disregarded.

Group selection. In order to identify possibly disparate physical behaviors of the time series, the above process can further be refined. Without being specific, we simply mention that different rank-one matrices from the SVD of Z can be chosen and collected together, and be processed separately, with the hope that properly partitioned groups would reflect different dynamical components of the original time series. See [16, 26] for some practical concerns and applications.

Signal recovery: It is almost always true that the truncated SVD of a structured matrix cannot preserve the original structure [8, 28]. As such, $\widetilde{Z}$ generally is not a Hankel matrix. The nearest Hankel matrix to $\widetilde{Z}$ can be obtained by *diagonally averaging* entries of $\widetilde{Z}$ along the anti-diagonals. In doing so, however, the rank of the resulting Hankel matrix approximation is no longer d. So the above four steps have to be repeated with the hope that when a limit point which is a Hankel matrix of rank d is finally determined.

In short, the basic idea of the SSA is to decompose the original time series into the sum of a number of interpretable components, each of which should be easily identified as part of a certain modulated signals or, in reciprocity, the random noises. The workings of the SSA are by repeatedly sifting the time series through the truncated SVD of the embedded trajectory matrix with the hope that a cleaner, more representative time series showing critical traits, such as low-frequency trends, high-frequency variability, cyclic components, or other structures of the original series, can be extracted.

The notion of the SSA is attractive because of its simplicity in implementation and its broadness in applications. The main mechanism whereby the scheme alternates between the rank reduction and the Hankel projection, however, deserves further understanding. We can think of at least three reasons. Firstly, since the method is within the general frame of lift-and-project, at best the linear rate of convergence is expected [8]. How would the strength of noise or the inherent intricacy of the original signal affects the convergence? Secondly, the the iteration may not achieve a Hankel approximation of rank d if the alternating projection gets stagnant at a local minimizer of the distance between the manifold of matrices of rank d and the subspace of Hankel matrices. What does the limit point of the SSA really represent? Thirdly, there might be multiple local minimizers. How does this local solution affect the interpretation of physical properties being investigated? The purpose of this paper is to study, by means of global optimization techniques on parameters of the Vandermonde factorization, what the absolutely best and low rank Hankel approximation can be achieved for a general time series. In return, we are interested in using this information as a reference point for assessing the quality of the SSA solutions.

2. Sinusoidal signal. To motivate the connection of a time series to its trajectory matrix, we first consider a noise-free time-domain signal comprising d components of exponentially decaying sinusoids

$$\mathfrak{s}(t) = \sum_{\ell=1}^{d} a_\ell e^{-\alpha_\ell t} e^{\imath(2\pi\nu_\ell t + \phi_\ell)}, \tag{2.1}$$

where a_ℓ, α_ℓ, ν_ℓ, and ϕ_ℓ are real numbers denoting the magnitude, the decay rate, the frequency, and the phase angle, respectively. Starting with $t_0 = 0$ and sampling this signal at uniformly spaced nodes $t_0, t_1, \ldots$ with fixed interval length Δt (so $1/\Delta t$ is the so called sampling rate), we obtain an infinite sequence

$$\mathfrak{s}(t_k) = \sum_{\ell=1}^{d} a_\ell e^{-\alpha_\ell k\Delta t} e^{\imath(2\pi\nu_\ell k\Delta t + \phi_\ell)} = \sum_{\ell=1}^{d} a_\ell e^{\imath\phi_\ell} \left(e^{(-\alpha_\ell + \imath 2\pi\nu_\ell)\Delta t}\right)^k.$$

For simplicity, denote $s_k := \mathfrak{s}(t_k)$ and define

$$\beta_\ell := a_\ell e^{\imath\phi_\ell}, \tag{2.2}$$

$$\lambda_\ell := e^{(-\alpha_\ell + \imath 2\pi\nu_\ell)\Delta t}. \tag{2.3}$$

Then the time series $\{s_0, s_1, \ldots\}$ enjoys the relationship

$$s_k = \sum_{\ell=1}^{d} \beta_\ell \lambda_\ell^k. \tag{2.4}$$

We can rewrite the corresponding trajectory matrix S as

$$\begin{aligned} S &= \sum_{\ell=1}^{d} \beta_\ell \begin{bmatrix} \lambda_\ell^0 & \lambda_\ell^1 & \lambda_\ell^2 & \cdots \\ \lambda_\ell^1 & \lambda_\ell^2 & & \\ \lambda_\ell^2 & & & \\ \vdots & & & \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1^0 & \lambda_2^0 & \cdots & \lambda_d^0 \\ \lambda_1^1 & \lambda_2^1 & & \lambda_d^1 \\ \lambda_1^2 & & & \\ \vdots & & & \vdots \end{bmatrix} \begin{bmatrix} \beta_1 & 0 & \cdots & 0 \\ 0 & \beta_2 & & \\ \vdots & & \ddots & \\ 0 & & & \beta_d \end{bmatrix} \begin{bmatrix} \lambda_1^0 & \lambda_1^1 & \lambda_1^2 & \cdots \\ \lambda_2^0 & \lambda_2^1 & & \\ \vdots & & & \\ \lambda_d^0 & \lambda_d^1 & & \cdots \end{bmatrix}, \end{aligned} \tag{2.5}$$

which easily shows that the matrix S is of rank d. For strictly decaying signals, we should expect that $\alpha_\ell > 0$ and, hence, $|\lambda_\ell| < 1$, implying that S is a bounded operator. The decomposition in the last equation of (2.5) is known as the Vandermonde factorization of S. Note that the trajectory matrix S is naturally a semi-infinite matrix. If we take only finitely many samples, then it means only a leading principal submatrix of size $p \times q$ extracted from S. A classical question immediately arises is how information of the original signal can be retrieved from a finite number of samples. Key to our global optimization approach, as will be explained below, are the variables β_ℓ, λ_ℓ, $\ell = 1, \ldots, d$, which serve as a parameterization of any Hankel matrices of rank d.

The converse of the above characterization is worth noting. Given a Vandermonde factorization in the form of (2.5), then by (2.2) the polar form of the complex number β_ℓ determines the magnitude a_ℓ and phase angle ϕ_ℓ. In the meantime, by (2.3) the polar form the quantity λ_ℓ determines the decay rate α_ℓ and the frequency ν_ℓ up to a scaling by Δt. The dependence on the scaling Δt is due to the fact that whenever a measurement of decay rate and frequency is referred to, the meaning of a unit

time ought to be defined first. Once Δt is specified, then a composite signal is completely determined from a given Vandermonde decomposition. In other words, we can interpret interchangeably between a bounded low rank Hankel matrix and a sinusoidal signal through the relationship (2.4).

When sampling an unknown signal $\mathfrak{s}(t)$ in practice, often the observed data $\{z_k\}$ are contaminated. Since the samples are taken independently, we may assume the popular AWGN channel model,

$$z_k := s_k + \epsilon_k, \tag{2.6}$$

i.e., white Gaussian noise ϵ_k with specified signal-to-noise ratio (SNR) is added linearly per sample. Through embedding, the corresponding trajectory matrix Z can be forced to maintain the Hankel structure. But the property of rank d is generally lost. Indeed, Z could easily be of "full" rank. In the SSA, the idea is to approximate Z by a Hankel approximation of rank d, whence the signal information of $\mathfrak{s}(t)$ is partially retrieved. As we have pointed out earlier, the SSA does not even make use of the semi-infinite matrix Z in its entirety. Only a segment of finite length n is fed into the procedure. To what extent this process of alternating projection can recover the essential information is the question we want to investigate in this study.

3. Vandermonde parametrization. The Vandermonde factorization of a trajectory matrix is not idiosyncratic for sinusoidal signals only. In this section, we briefly review the fact that any Hankel operator, of finite dimension or not, always enjoys such a factorization so long as it is of finite rank. The factorization offers an effective characterization in $2d$ parameters, if the operator is of rank d.

A comprehensive account of Hankel operators, their applications, and other related development can be found in the monograph [30]. Indeed, the importance of Hankel operators is evinced in their frequent appearance in many seemingly disparate areas of classical mathematics [1, 20, 29], as well as their critical applications in wide range of disciplines outside mathematics [12, 15]. So. other than the general description of Hankel structure that $H = [h_{ij}]$ where $h_{ij} = h_m$ for all indices $i + j = m$, it is essential to characterize a Hankel operator in the least possible parameters.

For finite dimensional problems, the Hankel structure allows effective implementations of various algorithmic schemes. For semi-infinite Hankel matrices, numerical computation remains challenging. For instance, though spectral properties of Hankel matrices are generally well understood [38], to retrieve the SVD information for semi-infinite dimensional matrices requires considerable efforts [23, 40]. In the event that an semi-infinite H is known a priori to have finite rank, such as that derived from a sinusoidal signal (2.5), endeavors have been made to approximate H by a finite-dimensional matrix. Careful readers then must question whether such a finite-dimensional matrix is able to preserve other innate properties of H.

One straightforward example, referred to as the Caratheodory-Fejer method [19], is to approximate the singular values of H by those of the principal submatrices of H. The idea is that the Courant-Fisher min-max principle enforces an interlacing property among the singular values of the gradually expanding matrices, which seems to serve as reasonable bounds for the singular values of H. That this notion of approximation by truncation has the danger of misconception can be seen from the example

$$H(a) = \begin{bmatrix} 1 & a & a^2 & \dots \\ a & a^2 & a^3 & \\ a^2 & a^3 & & \\ \vdots & & & \end{bmatrix}$$

given in [40]. It is obvious that the matrix $H(a)$ is of rank one. Let $H_k(a)$ denote the $k \times k$ leading principal submatrix of $H(a)$. Then $H_k(a)$ is of rank-one and its sole nonzero singular value is trivially given by $\frac{1-a^{2k}}{1-a^2}$, if $a \neq 1$; or k, if $a = 1$. As we gradually increase k to infinity, $H_k(a)$ should become a better approximation to $H(a)$, but its sole nonzero singular value also grows to infinity, if $a \geq 1$. In this way, even though $H(a)$ is of rank one, we will be forced to say that $H(a)$ has a singular value at infinity.

Indeed, a concern already arises at a more fundamental level when dealing with matrix-to-vector multiplication involving the semi-infinite matrix H. It inevitably involves infinite sums. For the infinite series to converge, we must impose the condition that H is a bounded linear operator. In most literature, H is considered as an operator over the space ℓ^2 of square summable (semi-infinite) sequences [32]. It turns that there is an interesting connection between bounded Hankel operators over ℓ^2 and bounded functions over the unit circle [27].

THEOREM 3.1. *The Hankel matrix H represents a bounded operator over ℓ^2 if and only if there exists a function $\psi \in L^\infty$ on the unit circle such that its*

$$h_m = \frac{1}{2\pi} \int_0^{2\pi} \psi(\theta) e^{-im\theta} d\theta, \quad m = 0, 1, \ldots. \tag{3.1}$$

In this case, the operator norm of H is bounded above by $\|\psi\|_\infty$.

In other words, the problem of whether the relevant sequence $\{h_0, h_1, \ldots\}$ determines a bounded operator H on ℓ^2 is equivalent to whether the sequence itself represents the Fourier coefficients of an essentially bounded function over the unit disk. See also [29, Theorem 2.1] for an elegant proof invoking the Hahn-Banach theorem. Since L^∞ is dense in L^2 over the compact set $[0, 2\pi]$, we may assume without loss of generality that all rows and columns of H are sequences in ℓ^2 themselves [6].

We are most interested in Hankel operators with finite rank. Toward this end, a classical result by Kronecker for checking the finite rank of a Hankel matrix is to examine the convergence of the associated generating function defined by the formal power series [29, Theorem 4.1]

$$G(z; \{h_n\}) := \sum_{n=0}^{\infty} h_n z^n. \tag{3.2}$$

THEOREM 3.2. *The Hankel matrix H has finite rank if and only the power series (3.2) determines a rational function. In this case,*

$$\operatorname{rank}(H) = \deg(zG(z; \{h_n\})), \tag{3.3}$$

where the degree of a rational function is the maximum of the degrees of its minimal constituent polynomials.

The above theorem nicely tells the precise rank, but finding the limit of the generating function $G(z; \{h_n\})$ is not always an easy task. An equivalent but more direct observation of finite rank is through a recursive relationship among elements of H [13, Chapter XV, Theorem 7].

THEOREM 3.3. *The Hankel matrix H is of finite rank d if and only if there exist constants $\gamma_0, \ldots, \gamma_{d-1}$ such that*

$$h_i = \gamma_{d-1} h_{i-1} + \gamma_{d-2} h_{i-2} + \ldots \gamma_o h_{i-d}, \quad i = d, d+1, \ldots \tag{3.4}$$

and d is the least integer having this property.

Assume the scenario that the semi-infinite Hankel matrix H in Theorem 3.3 is already known to be of rank d. Then it can be argued that the $d \times d$ leading principal submatrix $\widehat{H}$, i.e., the trajectory matrix of $h_0, \ldots, h_{2d-2}$, is necessarily nonsingular. Rewrite the finite difference equation (3.4) in the form of a linear system

$$\begin{bmatrix} h_0 & h_1 & h_2 & \ldots & h_{d-1} \\ h_1 & h_2 & h_3 & & h_d \\ h_2 & h_3 & & & h_{d+1} \\ \vdots & & & & \\ h_{d-1} & h_d & & & h_{2d-2} \end{bmatrix} \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{d-1} \end{bmatrix} = \begin{bmatrix} h_d \\ h_{d+1} \\ \vdots \\ h_{2d-1} \end{bmatrix} \tag{3.5}$$

for the coefficients $\gamma_0, \gamma_1, \ldots, \gamma_{d-1}$. We thus see that the parameters $\gamma_0, \ldots, \gamma_{d-1}$ and consequently the entire semi-infinite time series are uniquely determined by $\widehat{H}$ and one extra element h_{2d-1}. In other words, to uniquely determine a Hankel matrix of rank d, we need to know exactly $2d$ many elements $\{h_0, h_1, \ldots, h_{2d-1}\}$. Obviously, the linear system (3.5) can be generalized to reflect any segment of $2d$ successive elements, so long as the corresponding trajectory matrix is nonsingular.

Once the the parameters $\gamma_0, \ldots, \gamma_{d-1}$ are in hand, then by exploiting the roots of the characteristic polynomial

$$p(\lambda) := \lambda^d - \gamma_{d-1}\lambda^{d-1} - \ldots - \gamma_1\lambda - \gamma_0, \tag{3.6}$$

we should be able to directly generate the entry of H at any location without recursive reference to other entries. This formulation leads to a natural parameterization of H which will be used to reassess the rationales behind the SSA.

To fix the idea, let λ_ℓ, $\ell = 1, \ldots, r$, denote the distinct roots of $p(\lambda)$ in (3.6), each of which has multiplicity ρ_ℓ. So $\sum_{\ell=1}^{r} \rho_\ell = d$. We formulate a general solution to the difference equation (3.4) as follows. Let

$$\mathscr{V}_{p(\lambda)} = [\mathscr{V}^{(1)}, \ldots, \mathscr{V}^{(r)}] \in \mathbb{C}^{d\times d},$$

be a block matrix with the block $\mathscr{V}^{(\ell)} = [v_{ij}^{(\ell)}]$ of size $d \times \rho_\ell$, $\ell = 1, \ldots, r$, where

$$v_{ij}^{(\ell)} = c_{ij}\lambda_\ell^{i-j}, \quad i = 0, 1, \ldots, d-1, \quad j = 0, 1, \ldots, \rho_\ell - 1, \tag{3.7}$$

and $c_{ij} = 0$ if $i < j$; and $c_{ij} = \frac{i!}{(i-j)!j!}$ otherwise. A typical block looks like

$$\mathscr{V}^{(\ell)} := \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ \lambda_\ell & 1 & 0 & \ldots & \\ \lambda_\ell^2 & 2\lambda_\ell & 1 & \ldots & \\ \lambda_\ell^3 & 3\lambda_\ell^2 & 3\lambda_\ell & & \\ \vdots & & & \ddots & \\ \lambda_\ell^{\rho_\ell-1} & & & & 1 \\ \lambda_\ell^{\rho_\ell} & & & & \rho_\ell\lambda_\ell \\ \vdots & \vdots & & & \vdots \\ \lambda_\ell^{d-1} & (d-1)\lambda_\ell^{d-2} & \frac{(d-1)(d-2)}{2}\lambda_\ell^{d-3} & \ldots & \begin{pmatrix} d-1 \\ \rho_\ell - 1 \end{pmatrix}\lambda_\ell^{d-\rho_\ell} \end{bmatrix} \in \mathbb{C}^{d\times\rho_\ell}.$$

Note that for convenience the indices of the entries are counted starting from $(0, 0)$. Introducing the confluent Vandermonde system [21]

$$\mathscr{V}_{p(\lambda)} \begin{bmatrix} \beta_1^{(0)} \\ \vdots \\ \beta_1^{(\rho_1-1)} \\ \vdots \\ \beta_r^{(0)} \\ \vdots \\ \beta_r^{(\rho_r-1)} \end{bmatrix} = \begin{bmatrix} h_0 \\ h_1 \\ h_2 \\ \vdots \\ h_{d-1} \end{bmatrix} \tag{3.8}$$

so as to enforce the initial values $h_0, \ldots, h_{d-1}$ on the general solution of (3.4). Trivially, the equations in (3.8) can be expressed as

$$h_i = \sum_{\ell=1}^{r} \sum_{j=0}^{\rho_\ell - 1} \beta_\ell^{(j)} c_{ij} \lambda_\ell^{i-j}, \quad i = 0, 1, \ldots, d-1.$$

It can further be verified recursively that

$$\begin{aligned} h_i &= \sum_{s=0}^{d-1} \gamma_s h_{i-d+s} = \sum_{s=0}^{d-1} \gamma_s \left(\sum_{\ell=1}^{r} \sum_{j=0}^{\rho_\ell - 1} \beta_\ell^{(j)} c_{i-d+s,j} \lambda_\ell^{i-d+s-j} \right) \\ &= \sum_{\ell=1}^{r} \sum_{j=0}^{\rho_\ell - 1} \beta_\ell^{(j)} \left(\sum_{s=0}^{d-1} \left(c_{i-d+s,j} \gamma_s \lambda_\ell^{i-d+s-j} \right) \right) \\ &= \sum_{\ell=1}^{r} \sum_{j=0}^{\rho_\ell - 1} \beta_\ell^{(j)} c_{ij} \lambda_\ell^{i-j}, \quad i = d, d+1, \ldots . \end{aligned} \tag{3.9}$$

Let V_∞ denote the confluent Vandermonde matrix of size $\infty \times d$ obtained by extending the matrix $\mathscr{V}_{p(\lambda)}$ defined in (3.7) downward to infinite length. Upon carefully examining the general solution (3.9), we can rearrange the expression to establish the Vandermonde factorization for a semi-infinite Hankel matrix H as follows [5].

THEOREM 3.4. *Suppose H is a semi-infinite Hankel matrix of rank d. Then there exist a $d \times d$ block diagonal matrix D_∞ whose ℓ-th block is of size $\rho_\ell \times \rho_\ell$ and is Hankel and upper anti-triangular such that*

$$H = V_\infty D_\infty V_\infty^\top. \tag{3.10}$$

It is important to point out that the Vandermonde factorization described above is valid for any semi-infinite Hankel matrix of rank d. For a given H, the factorization furnishes a set of parameters $\beta_\ell^{(j)}$ and λ_ℓ, referred to henceforth as the Vandermonde parameters. If H as an operator is further required to be bounded, then the root conditions must be satisfied, i.e., all roots λ_ℓ's of the polynomial (3.6) must have modulus less than or equal to one and those of modulus one must be simple.

For our application, we shall assume henceforth the generic case that all roots of $p(\lambda)$ are distinct. The system (3.8) is reduced to $h_i = \sum_{\ell=1}^{d} \beta_\ell^{(0)} \lambda_\ell^i$ for $i = 0, 1, \ldots, d-1$. It follows that (3.4) can be written recursively as

$$\begin{aligned} h_i &= \sum_{s=0}^{d-1} \gamma_s h_{i-d+s} = \sum_{s=1}^{d-1} \gamma_s \left(\sum_{\ell=1}^{d} \beta_\ell^{(0)} \lambda_\ell^{i-d+s} \right) = \sum_{\ell=1}^{d} \beta_\ell^{(0)} \left(\sum_{s=0}^{d-1} \gamma_s \lambda_\ell^{i-d+s} \right) \\ &= \sum_{\ell=1}^{d} \beta_\ell^{(0)} \lambda_\ell^{i-d} \left(\sum_{s=0}^{d-1} \gamma_s \lambda_\ell^{s} \right) = \sum_{\ell=1}^{d} \beta_\ell^{(0)} \lambda_\ell^{i}, \quad i = d, d+1, \ldots, \end{aligned} \tag{3.11}$$

which, in turn, establishes the Vandermonde factorization (3.10) with the specific structure

$$D_\infty := \operatorname{diag}\{\beta_1^{(0)}, \ldots, \beta_d^{(0)}\}, \tag{3.12}$$

$$V_\infty^\top := \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \ldots \\ 1 & \lambda_2 & \lambda_2^2 & \ldots \\ \vdots & & & \\ 1 & \lambda_d & \lambda_d^2 & \ldots \end{bmatrix}. \tag{3.13}$$

4. Least squares approximation. In this section, we set up the least squares approximation by using the Vandermonde parameters. We have already argued that corresponding to a noise-free sinusoidal signal with d components (2.1), the time series $\{s_0, s_1, \ldots\}$ of samples at a fixed time interval Δt enjoys a Vandermonde factorization (2.5). When the noise is present, we have in hand only a contaminated sequence $\{z_0, z_1, \ldots\}$. Seeking out the noise-free signal (2.1) means finding the suitable parameters $\{\beta_1, \ldots, \beta_d\}$ and $\{\lambda_1, \ldots, \lambda_d\}$ for its Vandermonde factorization. The question is what counts as the best approximation.

Consider the scenario that the SSA is applied to the finite time series $\{z_0, z_1, \ldots, z_{2n-2}\}$ with a square Hankel embedding, i.e., suppose that the corresponding trajectory matrix Z is of size precisely $n \times n$. The iteration behind the SSA alternates between the best rank d approximation and the best Hankel approximation with respect to the Frobenius matrix norm. Since we already know how a Hankel matrix of rank d should look like, the SSA application amounts to solving the minimization problem

$$\min_{\substack{\beta_\ell, \lambda_\ell \in \mathbb{C} \\ |\lambda_\ell| \le 1,\ \ell = 1, \ldots, d}} \quad \sum_{i=0}^{n-1} (i+1) \left| z_i - \sum_{\ell=1}^{d} \beta_\ell \lambda_\ell^i \right|^2 + \sum_{i=n}^{2n-2} (2n-1-i) \left| z_i - \sum_{\ell=1}^{d} \beta_\ell \lambda_\ell^i \right|^2 . \tag{4.1}$$

As $n \to \infty$, we eventually are dealing with the problem

$$\min_{\substack{\beta_\ell, \lambda_\ell \in \mathbb{C} \\ |\lambda_\ell| \le 1,\ \ell = 1, \ldots, d}} \quad \sum_{i=0}^{\infty} (i+1) \left| z_i - \sum_{\ell=1}^{d} \beta_\ell \lambda_\ell^i \right|^2 . \tag{4.2}$$

Noticeable in both formulations is that the term-wise errors are not equally weighted. Does this make any sense? Take the extreme case (4.2) as a point of argument. We see that the weight is increased at a constant rate even when the signal gradually decays to zero. In a sense we are imposing more penalty on the noise for weaker signals. There are two concerns, however. One is that the weights grow unboundedly and there is no satisfactory theory to guarantee the convergence of such a weighted infinite series in general. The other is that the weight scheme might have biased the outcome of the optimizers $(\beta_\ell, \lambda_\ell)$ for overly correcting the higher order terms which are composed mostly of noises. On the other hand, we think that the bell-shape weight distribution $1, 2, \ldots, n, n-1, \ldots 1$ in the finite case (4.1) is more an artifact due to our obstinacy in maintaining the Hankel structure than any other mathematical reason [8, 28].

In what follows, we consider to fit the data by means of solving

$$\min_{\substack{\beta_\ell, \lambda_\ell \in \mathbb{C} \\ |\lambda_\ell| \le 1,\ \ell = 1, \ldots, d}} \quad \sum_{i=0}^{2n-2} \left| z_i - \sum_{\ell=1}^{d} \beta_\ell \lambda_\ell^i \right|^2 \tag{4.3}$$

from the first $2n-1$ samples, i.e., we regard all noises in (2.6) as being uncorrelated, independent, and having equal variance. For the SSA computation, it is preferred to keep n as small as is necessary because the size of the associated trajectory matrix does affect the overhead in the sequence of matrix decomposition. But in our formulation, the size n matters only to the summation. Note that there are only $2d$ parameters to be estimated and we allow λ_ℓ to range over a closed unit disk. The following arguments assure that the minimum in (4.3) does exist.

Define

$$\Lambda := \Lambda(\lambda_1, \ldots, \lambda_d) := \begin{bmatrix} I & I & \ldots & I \\ \lambda_1 & \lambda_2 & \ldots & \lambda_d \\ \lambda_1^2 & \lambda_2^2 & & \lambda_d^2 \\ \vdots & & & \\ \lambda_1^{2n-2} & \lambda_2^{2n-2} & & \lambda_d^{2n-2} \end{bmatrix}, \quad \boldsymbol{\beta} := \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{bmatrix}, \quad \mathbf{z} := \begin{bmatrix} z_0 \\ z_1 \\ \vdots \\ z_{2n-2} \end{bmatrix}.$$

Trivially, we can rewrite (4.3) in vector form

$$\min_{\substack{\beta_\ell,\lambda_\ell\in\mathbb{C}\\ |\lambda_\ell|\le 1,\ \ell=1,\dots,d}} \quad \|\Lambda\boldsymbol{\beta}-\mathbf{z}\|^2. \tag{4.4}$$

This is a nonlinear least squares problem over the complex field. For each given Λ of full column rank, the unique optimal solution for $\boldsymbol{\beta}$ is given by [24]

$$\boldsymbol{\beta} := \mathcal{L}(\Lambda;\mathbf{z}) := (\Lambda^*\Lambda)^{-1}\Lambda^*\mathbf{z} \tag{4.5}$$

where * stands for the conjugate transpose. The problem is thus reduced to finding λ for the minimization problem

$$\min_{|\lambda_1|,\dots,|\lambda_d|\le 1} \quad \|\Lambda(\Lambda^*\Lambda)^{-1}\Lambda^*\mathbf{z}-\mathbf{z}\|^2. \tag{4.6}$$

An example of such an objective function for the case $d=1$ with randomly generated noise is illustrated in Figure 4.1, showing the presence of multiple minimizers. Clearly, in all cases a minimizer over the compact d-dimensional torus must exist. Since we are dealing with a compact set, it is possible to exploit global optimization techniques to search for the absolute minimizer, which will be delineated later. We are interested in using the global solution to compare the result constructed from the SSA algorithm.

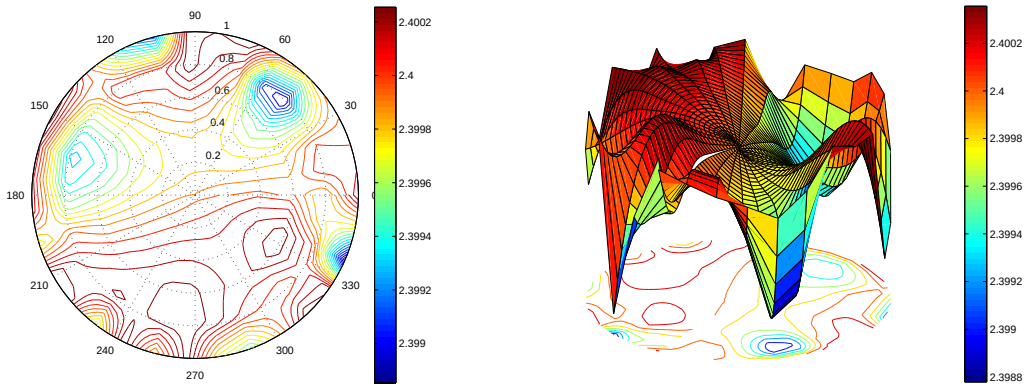


FIG. 4.1. *An example of objective function for (4.6). (The surface has been rotated with AZ=124 and EL=38.)*

It might be useful to include the gradient information of the objective function in (4.6) for the global optimization method. Let the objective function be redefined as

$$f(\lambda_1,\dots,\lambda_d) := \frac{1}{2}\|\Lambda\mathcal{L}(\Lambda;\mathbf{z})-\mathbf{z}\|^2 \tag{4.7}$$

which we rewrite in the form

$$f(\lambda_1,\dots,\lambda_d) = \frac{1}{2}\langle\Lambda\mathcal{L}(\Lambda;\mathbf{z})-\mathbf{z},\Lambda\mathcal{L}(\Lambda;\mathbf{z})-\mathbf{z}\rangle,$$

with inner product for complex vectors defined by $\langle\mathbf{p},\mathbf{q}\rangle := \sum_i p_i\overline{q}_i$. So that only box constraints are involved, we employ polar coordinates (ρ_i,θ_i) when $\lambda_i = \rho_i e^{i\theta_i}$ as variables in our global method. That is, we identify $\Lambda = \Lambda(\lambda_1,\dots,\lambda_d) = \Lambda(\rho_1,\dots,\rho_d,\theta_1,\dots,\theta_d)$. We now calculate the gradient of f

with respect to the real variables ρ_i and θ_i. This is not complex differentiation, but rather the notion of Frèchet derivative prevails.

Using the fact that $\Lambda^* \left(\Lambda\mathcal{L}(\Lambda;\mathbf{z}) - \mathbf{z}\right) = 0$, the action of the Fréchet derivative of f with respect to Λ at a complex matrix H is given by

$$\frac{\partial f}{\partial \Lambda}.H = \Re\left(\langle H, \left(\Lambda\mathcal{L}(\Lambda;\mathbf{z}) - \mathbf{z}\right)\mathcal{L}(\Lambda;\mathbf{z})^*\rangle\right),$$

where $\Re$ stands for the real part of a complex-valued quantity and the same notation $\langle\cdot,\cdot\rangle$ denotes the generalization to the Frobenius inner product of complex matrices. On the other hand, the action of the Fréchet derivative of Λ with respect to variables $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_d]$ and $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_d]$ at vectors $\mathbf{h}, \mathbf{k} \in \mathbb{R}^d$ can be expressed as

$$\begin{aligned}
\frac{\partial \Lambda}{\partial \boldsymbol{\rho}}.\mathbf{h} &= \underbrace{\mathrm{diag}\{0, 1, 2, \ldots, 2n-2\}}_{\Xi} \Lambda \mathrm{diag}\left\{\boldsymbol{\rho}\right\}^{-1} \mathrm{diag}\{\mathbf{h}\}, \\
\frac{\partial \Lambda}{\partial \boldsymbol{\theta}}.\mathbf{k} &= \imath \Xi \Lambda \mathrm{diag}\{\mathbf{k}\},
\end{aligned}$$

respectively. By the chain rule, we obtain the actions of the gradient of f as follows:

$$\begin{aligned}
\frac{\partial f}{\partial \boldsymbol{\rho}}.\mathbf{h} = \frac{\partial f}{\partial \Lambda}.\left(\frac{\partial \Lambda}{\partial \boldsymbol{\rho}}.\mathbf{h}\right) &= \Re\left(\langle \Xi\Lambda \mathrm{diag}\left\{\boldsymbol{\rho}\right\}^{-1} \mathrm{diag}\{\mathbf{h}\}, \left(\Lambda\mathcal{L}(\Lambda;\mathbf{z}) - \mathbf{z}\right)\mathcal{L}(\Lambda;\mathbf{z})^*\rangle\right), \\
&= \Re\left(\langle \mathrm{diag}\{\mathbf{h}\}, \mathrm{diag}\left\{\boldsymbol{\rho}\right\}^{-1} \Lambda^* \Xi \left(\Lambda\mathcal{L}(\Lambda;\mathbf{z}) - \mathbf{z}\right)\mathcal{L}(\Lambda;\mathbf{z})^*\rangle\right), \\
\frac{\partial f}{\partial \boldsymbol{\theta}}.\mathbf{k} = \frac{\partial f}{\partial \Lambda}.\left(\frac{\partial \Lambda}{\partial \boldsymbol{\theta}}.\mathbf{k}\right) &= \Re\left(\langle \imath\Xi\Lambda \mathrm{diag}\{\mathbf{k}\}, \left(\Lambda\mathcal{L}(\Lambda;\mathbf{z}) - \mathbf{z}\right)\mathcal{L}(\Lambda;\mathbf{z})^*\rangle\right) \\
&= \Re\left(\langle \mathrm{diag}\{\mathbf{k}\}, -\imath\Lambda^*\Xi \left(\Lambda\mathcal{L}(\Lambda;\mathbf{z}) - \mathbf{z}\right)\mathcal{L}(\Lambda;\mathbf{z})^*\rangle\right)
\end{aligned}$$

By the Riesz representation theorem, we see that the gradient of f can be expressed as

$$\begin{aligned}
\frac{\partial f}{\partial \boldsymbol{\rho}} &= \Re\left(\mathrm{diag}\left\{\mathrm{diag}\left\{\boldsymbol{\rho}\right\}^{-1} \Lambda^* \Xi \left(\Lambda\mathcal{L}(\Lambda;\mathbf{z}) - \mathbf{z}\right)\mathcal{L}(\Lambda;\mathbf{z})^*\right\}\right), \\
\frac{\partial f}{\partial \boldsymbol{\theta}} &= \Re\left(\mathrm{diag}\left\{-\imath\Lambda^* \Xi \left(\Lambda\mathcal{L}(\Lambda;\mathbf{z}) - \mathbf{z}\right)\mathcal{L}(\Lambda;\mathbf{z})^*\right\}\right).
\end{aligned}$$

5. Numerical experiment. The feasible set in our least squares formulation for the SSA is compact, so the absolute minimizer always exists. In this section, we report some numerical experiments on employing the global optimization techniques to find the absolute minimizer and make comparison with results from the SSA algorithm.

The objective of global optimization is to find the globally best solution in the presence of multiple local optima. Over the years, many strategies for global optimization have been proposed. A comprehensive survey on continuous global optimization can be found in [31]. One prevailing scheme that allows general structural constraints and makes possible straightforward generalizations and extensions of existent local algorithms is the notion of adaptive partition.

Some practical methods with a view towards efficient computer-based implementations include multistart framework equipped with sophisticated scatter search algorithm [34]; generic algorithms motivated by the process of natural selection and the survival of the fittest principle [18, 37]; simulated annealing based upon the physical analogy of cooling crystal structures that spontaneously attempt to arrive at some stable and, hopefully, global equilibrium [33, 35]; and pattern search methods which compare instantiations and explore new regions with the goal of conducting a global search [3, 22]. Taking advantage of the global optimization toolbox available in MATLAB, we choose to use the MULTISTART method with solver FMINCON to carry out our experiments.

We carry out our experiment by setting up test data in the following way. We assume that the expected rank d is known a priori, which itself is a difficult problem in practice. We generate a sinusoidal signal $\mathfrak{s}(t)$ in the form of (2.1) with d components as the basis for comparison. This exact signal is random in the sense that its parameters $a_\ell \in [0, 6]$, $\alpha_\ell \in [0, 5]$, and $\phi_\ell \in [0, 1]$ are uniformly distributed pseudorandom numbers, while the frequencies ν_ℓ are pseudorandom integers from a uniform discrete distribution over $[0, 400]$. Sampling rate is set at 2^9 and we take $2^{10} - 1$ samples $s_k := \mathfrak{s}(t_k)$ over the time interval $[0, 2)$. Additive white Gaussian noises based on the AWGN channel model with fixed SNR are then added to the signal in the way of (2.6) as the observed signal. We apply both the SSA algorithm and the MultiStart method to the observed signals. The goal is to examine the closeness of the reconstructed signals by the algorithms to the original signal.

We first experiment with the effect of error strength on the performance of the SSA method. For simplicity of demonstration, we choose not to perform the group selection and consider only the case with embedding dimension 512, i.e., the trajectory matrix Z is always of size 512×512. With each fixed $d = 1, \ldots, 5$, we randomly generate 100 signals. The alternating projection scheme is applied to each randomly perturbed time series until two consecutive Hankel projections differ by less than 10^{-12} when we regard that convergence has been achieved. Each iteration involves the SVD of a 512×512 Hankel matrix, which is not cheap. Depicted Figure 5.1 are the boxplots for numbers of iterations needed by the SSA to attain a rank d Hankel approximation of the original Z. It seems to suggest that the strength of the noise (in dB) is not critical for convergence. However, from Figure 5.2, it is quite obvious that stronger noise (negative SNR) does affect the quality of approximation.

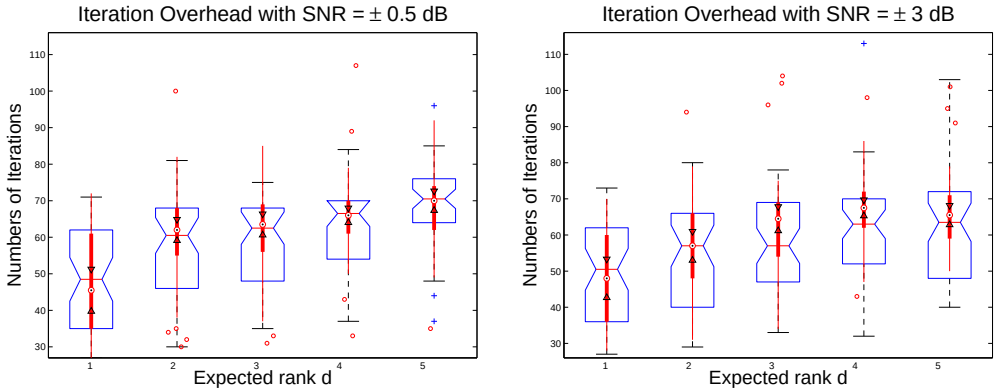


FIG. 5.1. *Statistics of SSA on overhead in terms of iterations needed for convergence.* SNR > 0 *(red compact box style);* SNR < 0 *(blue empty box style).*

We caution that, because the manifold of rank d matrices is not convex, the alternating projection scheme, central to the SSA algorithm, may not return at its convergence a rank d Hankel approximation to the given trajectory matrix Z. Even if it does, the limit point may not be the nearest rank d Hankel approximation to Z. See a counterexample in [8]. In contrast, our formulation directly searches for the best fitting signal in the sense of (4.4).

In yet another experiment, we randomly generate 50 signals for each $d = 1, \ldots, 5$ with AWGN noises at SNR $= \pm 3$ dB and compare the errors of the signals reconstructed by the SSA algorithm and our global optimization approach. From Figure 5.3, we see that in general the global optimization approach gives rise to better approximation to the original time series $\{s_k\}$ than the SSA algorithm. The advantage of the global optimization approach to the SSA algorithm becomes more obvious when the noise is relatively stronger (e.g., dB = -3). For weakly perturbed signals, the SSA performs almost equally with the global optimization approach.

The boxplots in Figure 5.3 indicates to us the overall trend. To compare the errors produced by

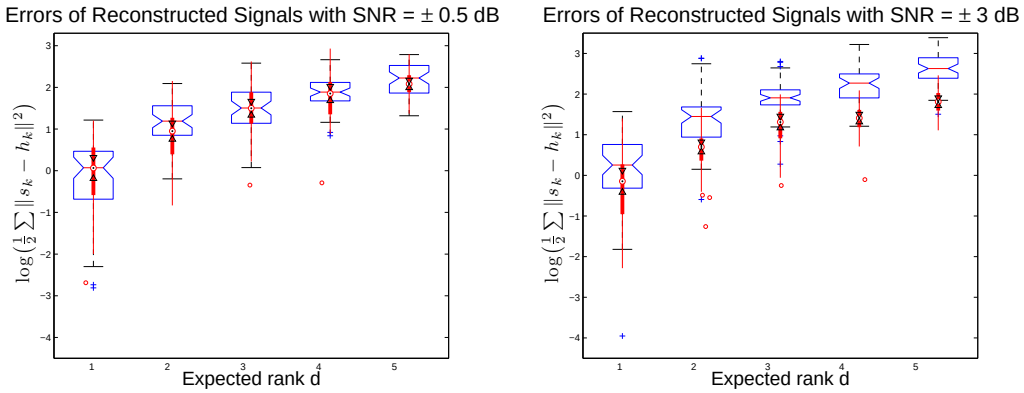


FIG. 5.2. *Statistics of SSA on errors between the true time series* $\{s_k\}$ *and the reconstructed time series* $\{h_k\}$. SNR > 0 *(red compact box style);* SNR < 0 *(blue empty box style).*

each independent random test, we take the ratio of the SSA error to the global optimization error sample by sample and plot the histogram in Figure 5.4. A negative logarithm of the ratio indicates that the SSA reconstruction is superior to the global optimization reconstruction. We observe that, except for the case $d = 1$, the global optimization usually has a higher percentage of producing a better reconstruction when the noise is strong. In contrast, when the noise is relatively weak (e.g., dB $= 3$), the simpler SSA performs amazingly well, comparing to the more complicated global optimization.

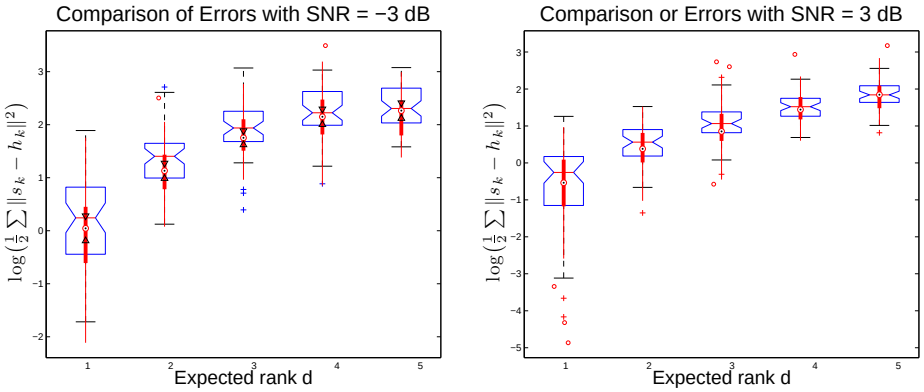


FIG. 5.3. *Statistics of errors between the true time series* $\{s_k\}$ *and the reconstructed time series* $\{h_k\}$ *by the global optimization method (red compact box style) and by the SSA algorithm (blue empty box style).*

One might wonder why sometimes the SSA algorithm gives better approximation to $\{s_k\}$. Should not the global optimization does the best? There are multiple factors that cause this to happen. Firstly, the quantity being minimized in the global optimization is the sum $\sum_{i=0}^{2n-2} \left| z_i - \sum_{\ell=1}^{d} \beta_\ell \lambda_\ell^i \right|^2$ (see (4.3)), not the absolute error $\sum_{i=0}^{2n-2} \left| s_i - \sum_{\ell=1}^{d} \beta_\ell \lambda_\ell^i \right|^2$. Because the true time series $\{s_k\}$ is practically not available in reality, the global optimization offers at its best an optimal approximation to the series $\{z_k\}$, not $\{s_k\}$. Secondly, the machinery we employed in the global optimization is the scheme MULTISTART which uses uniformly distributed start points within predefined bounds. In our experiment we allow merely 40 start points. It is possible to improve the objective value, or to confirm

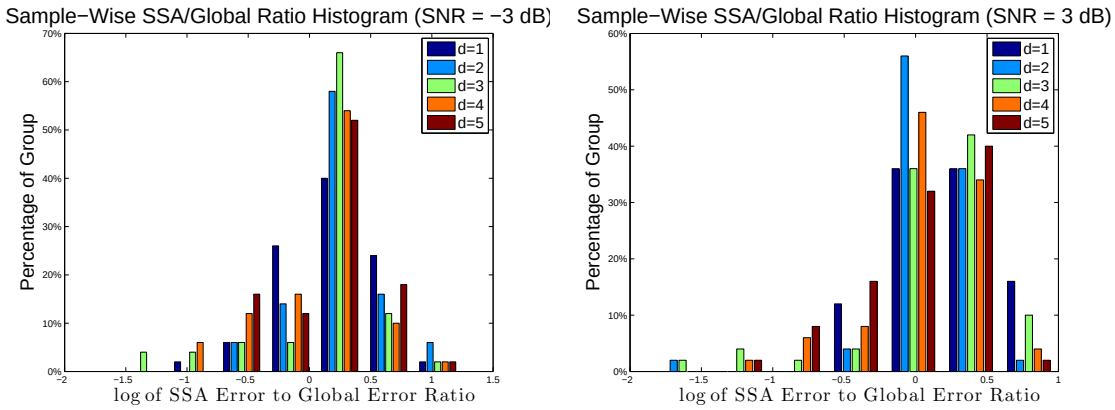


FIG. 5.4. *Histogram of sample-by-sample SSA error to global optimization error ratios.*

the global solution, by more start points. Thirdly, we have mentioned that not necessarily the SSA algorithm will produce the best rank d Hankel approximation at its convergence. Even if it does, the SSA is a specially weighted least squares problem (see (4.1)), whereas our formulation (4.3) is not. So we are comparing nearness to a "blackbox" $\{s_k\}$ from two different objective functions.

If we really want to consider

$$\min_{\substack{\beta_\ell,\lambda_\ell\in\mathbb{C}\\ |\lambda_\ell|\le 1,\ \ell=1,\dots,d}} \quad \sum_{i=0}^{2n-2} \omega_i \left| z_i - \sum_{\ell=1}^{d} \beta_\ell \lambda_\ell^i \right|^2 \tag{5.1}$$

with weights $\omega_i > 0$, the the normal equation (4.5) per fixed Λ becomes

$$\Lambda^* W \Lambda \boldsymbol{\beta} = \Lambda^* W \mathbf{z}, \tag{5.2}$$

with $W := \mathrm{diag}\{\omega_0, \omega_0, \dots, \omega_{2n-2}, \omega_{2n-2}\}$. Equivalently, with the change of variables by diagonal scaling

$$\Upsilon := W^{\frac{1}{2}}\Lambda, \quad \mathbf{c} := W^{\frac{1}{2}}\mathbf{z}, \tag{5.3}$$

the weighted least squares problem (5.1) can be reformulated as

$$\min_{|\lambda_1|,\dots,|\lambda_d|\le 1} \quad \|\Upsilon(\Upsilon^*\Upsilon)^{-1}\Upsilon^*\mathbf{c} - \mathbf{c}\|^2, \tag{5.4}$$

which is analogous to (4.4) and our code can easily be modified accordingly. If in particular we use the special SSA weight scheme, we find surprisingly a reversal situation, as seen in Figure 5.5, that the SSA generally performs better than the global optimization at the presence of stronger noise..

6. Conclusion. Alternating projection between two manifolds with the hope of seeking out a point of intersection, or a pair of points yielding minimum distance, has long been employed across the fields in a variety of applications. Starting with a Hankel matrix embedded with a contaminated time series, the SSA exploits this notion to find, in particular, a low rank Hankel approximation. The idea is simple enough for quick implementation and has gained popularity among practitioners.

This paper investigates the effectuality of the SSA algorithm by comparing its reconstructed time series with the absolute best approximation obtained from global optimization techniques. The framework is built upon Vandermonde factorization of Hankel operators, which provides natural parameters over compact feasible set and, hence, guarantees the existence of a global solution.

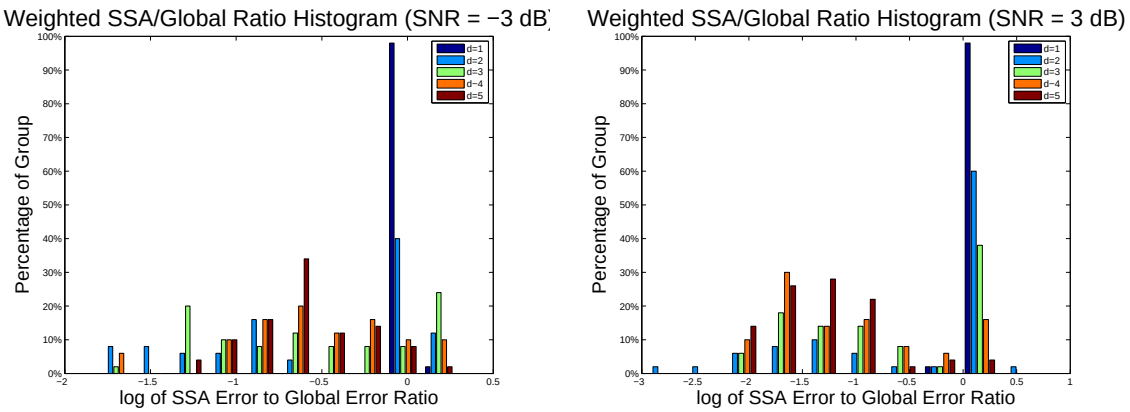


FIG. 5.5. *Histogram of weighted sample-by-sample SSA error to global optimization error ratios.*

Setting the stopping criterion at the nearness of two successive iterates within a uniformly specified bound, we find that the convergence behavior of the SSA algorithm is generally insensitive to the strength of the added noise. That is, so long as the size of the embedding is fixed, it usually takes about the same order of iterations for the SSA to converge, regardless of the SNR. On the other hand, as is expected, stronger noise does degrade the quality of the final approximation to the original, uncontaminated signal.

What is most interesting is the empirical evidence that, despite of its simplicity, the SSA does indeed perform remarkably well when comparing to the results from the more complicated global methods. This is especially so when the SNR is relatively high. Since the global methods, if successful, produce the best possible approximation to the Vandermonde parameters from the contaminated data, such a comparison justifies therefore that, for exploratory model-building purpose, the simpler SSA algorithm might be sufficient as a handy tool. For more rigorous affirmative procedure, however, we must stress that the SSA does not always give rise to an optimal solution.

REFERENCES

[1] V. M. ADAMJAN, D. Z. AROV, AND M. G. KREĬN, *Infinite Hankel matrices and generalized Carathéodory-Fejér and I. Schur problems*, Funkcional. Anal. i Priložen., 2 (1968), pp. 1–17.

[2] M. R. ALLEN AND L. A. SMITH, *Monte carlo ssa: detecting irregular oscillations in the presence of coloured noise*, Journal of Climate, 9 (1996), pp. 3372–3404.

[3] C. AUDET, J. E. DENNIS, JR., AND S. LE DIGABEL, *Globalization strategies for mesh adaptive direct search*, Comput. Optim. Appl., 46 (2010), pp. 193–215.

[4] M. AUVERGNE, *Singular value analysis applied to phase space reconstruction of pulsating stars*, Astronomy and Astrophysics, 204 (1988), pp. 341–348.

[5] D. L. BOLEY, F. T. LUK, AND D. VANDEVOORDE, *Vandermonde factorization of a Hankel matrix*, in Scientific computing (Hong Kong, 1997), Springer, Singapore, 1997, pp. 27–39.

[6] C. M. BOWDEN, *Boundedness of linear operators in the space ℓ^2*, Int. J. Quantum Chem., 2 (1968), pp. 363–371.

[7] D. S. BROOMHEAD AND G. P. KING, *Extracting qualitative dynamics from experimental data*, Phys. D, 20 (1986), pp. 217–236.

[8] M. T. CHU, R. E. FUNDERLIC, AND R. J. PLEMMONS, *Structured low rank approximation*, Linear Algebra Appl., 366 (2003), pp. 157–172. Special issue on structured matrices: analysis, algorithms and applications (Cortona, 2000).

[9] D. DANILOV AND A. E. ZHIGLJAVSKY, *Principal components of time series: the 'caterpillar' method*, University of St. Petersburg Press. (In Russian), (1997).

[10] J. ELSNER AND A. TSONIS, *Singular Spectrum Analysis: A New Tool in Time Series Analysis*, Language of Science, Plenum Press, 1996.

[11] K. FRAEDRICH, *Estimating the dimensions of weather and climate attractors*, J. Atmospheric Sci., 43 (1986),

pp. 419–432.
[12] B. A. Francis, *A course in H_∞ control theory*, vol. 88 of Lecture Notes in Control and Information Sciences, Springer-Verlag, Berlin, 1987.
[13] F. R. Gantmacher, *The theory of matrices. Vols. 1, 2*, Translated by K. A. Hirsch, Chelsea Publishing Co., New York, 1959.
[14] M. Ghil and C. Taricco, *Advanced spectral analysis methods*, In Past and Present Variability of the Solar-Terrestrial System: Measurement, Data Analysis and Theoretical Models, (1997), pp. 137–159.
[15] K. Glover, *All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds*, Internat. J. Control, 39 (1984), pp. 1115–1193.
[16] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *Analysis of time series structure*, vol. 90 of Monographs on Statistics and Applied Probability, Chapman & Hall/CRC, Boca Raton, FL, 2001. SSA and related techniques.
[17] H. Hassani, R. Mahmoudvand, and M. Zokaei, *Separability and window length in singular spectrum analysis*, C. R. Math. Acad. Sci. Paris, 349 (2011), pp. 987–990.
[18] R. L. Haupt and S. E. Haupt, *Practical genetic algorithms*, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, second ed., 2004. With 1 CD-ROM (Windows).
[19] J. W. Helton and N. J. Young, *Approximation of Hankel operators: truncation error in an H^∞ design method*, in Signal Processing, Part II, vol. 23 of IMA Vol. Math. Appl., Springer, New York, 1990, pp. 115–137.
[20] I. S. Iohvidov, *Hankel and Toeplitz matrices and forms*, Birkhäuser Boston, Mass., 1982. Algebraic theory, Translated from the Russian by G. Philip A. Thijsse, With an introduction by I. Gohberg.
[21] D. Kalman, *The generalized Vandermonde matrix*, Math. Mag., 57 (1984), pp. 15–21.
[22] T. G. Kolda, R. M. Lewis, and V. Torczon, *Optimization by direct search: new perspectives on some classical and modern methods*, SIAM Rev., 45 (2003), pp. 385–482.
[23] S. Y. Kung, *A new identification and model reduction algorithm via singular value decomposition*, in Proceedings of the 12th Asilomar Conference on Circuits, Systems and Computers, Pacific Grove, CA, November 1978, pp. 705–714.
[24] K. S. Miller, *Complex linear least squares*, SIAM Rev., 15 (1973), pp. 706–726.
[25] A. Mineva and D. Popivanov, *Method for single-trial readiness potential identification, based on singular spectrum analysis*, Journal of Neuroscience Methods, 68 (1996), pp. 91 – 99.
[26] V. Moskvina and K. M. Schmidt, *Approximate projectors in singular spectrum analysis*, SIAM J. Matrix Anal. Appl., 24 (2003), pp. 932–942 (electronic).
[27] Z. Nehari, *On bounded bilinear forms*, Ann. of Math. (2), 65 (1957), pp. 153–162.
[28] H. Park, L. Zhang, and J. B. Rosen, *Low rank approximation of a Hankel matrix by structured total least norm*, BIT, 39 (1999), pp. 757–779.
[29] V. V. Peller, *An excursion into the theory of Hankel operators*, in Holomorphic spaces (Berkeley, CA, 1995), vol. 33 of Math. Sci. Res. Inst. Publ., Cambridge Univ. Press, Cambridge, 1998, pp. 65–120.
[30] ———, *Hankel operators and their applications*, Springer Monographs in Mathematics, Springer-Verlag, New York, 2003.
[31] J. Pintér, *Global optimization in action*, vol. 6 of Nonconvex Optimization and its Applications, Kluwer Academic Publishers, Dordrecht, 1996. Continuous and Lipschitz optimization: algorithms, implementations and applications.
[32] S. C. Power, *Hankel operators on Hilbert space*, vol. 64 of Research Notes in Mathematics, Pitman (Advanced Publishing Program), Boston, Mass., 1982.
[33] P. Salamon, P. Sibani, and R. Frost, *Facts, conjectures, and improvements for simulated annealing*, SIAM Monographs on Mathematical Modeling and Computation, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002.
[34] Z. Ugray, L. Lasdon, J. Plummer, F. Glover, J. Kelly, and R. Martí, *Scatter search and local NLP solvers: a multistart framework for global optimization*, INFORMS J. Comput., 19 (2007), pp. 328–340.
[35] P. J. M. van Laarhoven and E. H. L. Aarts, *Simulated annealing: theory and applications*, vol. 37 of Mathematics and its Applications, D. Reidel Publishing Co., Dordrecht, 1987.
[36] R. Vautard, P. Yiou, and M. Ghil, *Singular-spectrum analysis: A toolkit for short, noisy chaotic signals*, Physica D: Nonlinear Phenomena, 58 (1992), pp. 95 – 126.
[37] M. D. Vose, *The simple genetic algorithm*, Complex Adaptive Systems, MIT Press, Cambridge, MA, 1999. Foundations and theory, A Bradford Book.
[38] H. Widom, *Hankel matrices*, Trans. Amer. Math. Soc., 121 (1966), pp. 1–35.
[39] P. Yiou, D. Sornette, and M. Ghil, *Data-adaptive wavelets and multi-scale singular-spectrum analysis*, Phys. D, 142 (2000), pp. 254–290.
[40] N. J. Young, *The singular-value decomposition of an infinite Hankel matrix*, Linear Algebra Appl., 50 (1983), pp. 639–656.