Robust Initialization Strategies for Hankel Structured Low-Rank Approximation via Variable Projection

Natsuki Yoshino* and Akira Tanaka†
* Graduate School of Information Science and Technology, Hokkaido University, Japan
E-mail: n_yoshino@ist.hokudai.ac.jp
† Faculty of Information Science and Technology, Hokkaido University, Japan
E-mail: takira@ist.hokudai.ac.jp

***Abstract*—We address the problem of Hankel Structured Low-Rank Approximation (HSLRA) using the Variable Projection (VP) method, which has been explored in signal processing tasks such as denoising and spectral estimation. A notable issue in VP-based HSLRA is its sensitivity to initialization, particularly under noisy conditions. We identify a phenomenon termed *singular value switching*, where the singular vector associated with the smallest singular value may not best reflect the true signal structure due to noise-induced perturbations. To mitigate this, we propose two initialization strategies: (M1) selecting among multiple singular vectors based on VP objective evaluation, and (M2) incorporating random linear combinations of singular vectors to expand the candidate set. Numerical experiments on exponentially decaying sinusoidal signals demonstrate that the proposed methods enhance reconstruction accuracy and robustness compared to conventional VP initialization.**

I. Introduction

In this study, we address an optimization problem known as the structured low-rank approximation (SLRA) of a Hankel matrix, referred to as HSLRA [1]. For the given signal $\boldsymbol{p} = [p_1 \cdots p_N]^\top \in \mathbb{R}^N$ and the positive definite weighting matrix $\mathbf{W} \in \mathbb{R}^{N\times N}$, the problem is formulated as

$$\begin{aligned} &\min_{\hat{\boldsymbol{p}}} \|\boldsymbol{p}-\hat{\boldsymbol{p}}\|_{\mathbf{W}}^2 \\ &\text{subject to } \operatorname{rank} \boldsymbol{\mathcal{H}}_L(\hat{\boldsymbol{p}}) \le r, \end{aligned} \tag{1}$$

where $\hat{\boldsymbol{p}} \in \mathbb{R}^N$ is the approximation to $\boldsymbol{p}$ and $\boldsymbol{\mathcal{H}}_L(\boldsymbol{p}) \in \mathbb{R}^{(N-L+1)\times L}$ denotes the Hankel matrix of window length L, defined as

$$\boldsymbol{\mathcal{H}}_L(\boldsymbol{p}) := \begin{bmatrix} p_1 & p_2 & \cdots & p_L \\ p_2 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ p_{N-L+1} & \cdots & \cdots & p_N \end{bmatrix}. \tag{2}$$

A signal $\hat{\boldsymbol{p}}$ that parameterizes a low-rank Hankel matrix $\boldsymbol{\mathcal{H}}_L(\hat{\boldsymbol{p}})$ with $\operatorname{rank} \boldsymbol{\mathcal{H}}_L(\hat{\boldsymbol{p}}) = r$ is closely related to a linear recurrence relation (LRR) [2]. It is well known that any real-valued signal satisfying such a relation can be represented as a sum of products of polynomials and exponentially decaying sinusoids:

$$\hat{p}_t = \sum_{j=1}^{R} g_j(t)\exp(d_j t)\sin(2\pi\omega_j t + \phi_j),\ t = 1,\dots,N, \tag{3}$$

where $g_j(t)$ is a real-valued polynomial of polynomial degree m_k and $r = 2R$ [3]. Therefore, the optimization problem (1) for $L > 2R$ seeks an approximattion to a given signal $\boldsymbol{p}$ by a model described in (3).

Two distinct signal behaviors—exponentially decaying sinusoids and purely polynomial components—are captured by this formulation, which serves as a general modeling structure in signal processing through a unified framework. This modeling capability has led to a wide range of applications of HSLRA in signal processing, including spectral estimation, where HSLRA helps extract dominant frequency components from noisy data [4], [5]; sparse signal recovery, where HSLRA aids in denoising and interpolation [6], [7]; system identification, where HSLRA corresponds to fitting a linear time-invariant system with bounded model complexity; and approximate Prony methods, which estimate signal parameters such as frequencies and damping factors, as described in section II [8], [9]. Beyond these classical applications, the inherent property of low-rank Hankel matrices has also been utilized in advanced domains such as magnetic resonance imaging (MRI) [10], [11] and seismic data analysis [12], [13]. These scenarios typically involve multidimensional signals and therefore employ block Hankel matrices to capture more intricate dependencies.

A. Related Work

Although the HSLRA problem (1) is inherently non-convex with many local minima [14], local optimization methods have been widely adopted due to their computational efficiency. Among them, Cadzow's iterative algorithm [15] is one of the earliest methods, an alternating projection technique. Other approaches include structured total least-norm [16], Newton-like iterations [17], symbolic methods [14], proximal approaches [7], and ODE-based method [18]. Among them, the method by Usevich [19], based on the variable projection principle [20] and Gauss-Newton optimization, has gained attention. This method is referred to as VP in the remainder of this paper. It supports a wide range of structured matrices and achieves linear iteration complexity for certain weight matrices. Its MATLAB implementation [21] is also available. In this study, we adopt the VP method as the baseline approach for solving the HSLRA problem (1). A variant of VP specifically designed for HSLRA is also proposed in [22].

979-8-3315-7206-8/25/$31.00

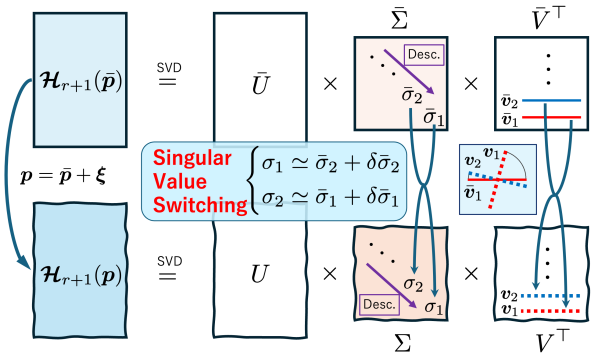

Fig. 1: Singular value switching example: a perturbation swaps the smallest and second smallest singular values, shifting alignment from $\boldsymbol{v}_1$ to $\boldsymbol{v}_2$.

B. Proposed Approach

Existing initialization strategies of VP often rely on the right singular vector corresponding to the smallest singular value of the Hankel matrix. However, this approach can be unreliable under parameter perturbations, as the singular vector most aligned with the true structure may no longer correspond to the smallest singular value—a phenomenon we refer to as singular value switching (see Fig. 1).

This misalignment can lead to degraded optimization performance. To address this issue, we propose two initialization strategies that enhance the robustness of the VP method. Both strategies are inspired by the multi-start approach commonly used in local optimization, where multiple initial points are evaluated to avoid convergence to suboptimal local minima. In our context, we generate multiple candidate initial vectors based on the right singular vectors of the Hankel matrix and select the one that minimizes the VP objective function. The first method (M1) directly evaluates a fixed number of singular vectors, while the second method (M2) extends this idea by incorporating random linear combinations of these vectors to diversify the candidate set further.

These strategies build upon the concepts introduced in our previous work [23], where similar initialization techniques were proposed in the context of Sylvester matrices. The present study extends these ideas to Hankel matrices, demonstrating their effectiveness in a broader class of structured low-rank approximation problems. While this paper focuses on the theoretical development of robust initialization strategies for HSLRA, the proposed methods are designed to improve the reliability and convergence of the solver used in practical signal processing tasks introduced earlier.

II. Low-rank Hankel matrix and sum of damuped sinusoids

For a fixed signal length N and varying window length L, the following lemma is known:

Lemma 1 ([2, Corollary 5.1])**.** For any sequence $\boldsymbol{p} \in \mathbb{C}^N$, let $L' = \lfloor (N-1)/2 \rfloor$ and $d = \operatorname{rank} \boldsymbol{\mathcal{H}}_{L'}(\boldsymbol{p})$. Then, we obtain

$$\operatorname{rank} \boldsymbol{\mathcal{H}}_L(\boldsymbol{p}) = \min(L, N-L+1, d).$$

This lemma guarantees the existence of signal $\boldsymbol{p}$ whose associated Hankel matrices $\boldsymbol{\mathcal{H}}_L(\boldsymbol{p})$ attain a constant rank over a broad range of window lengths L, excluding cases where the matrix becomes extremely tall or wide—that is, when L is very small or very close to N. This result indicates that such sequences possess an intrinsic structural property that manifests as rank invariance of the Hankel matrix across a balanced range of window lengths. Here, the rank of a signal $\boldsymbol{p}$ is defined as the integer r such that $\operatorname{rank} \boldsymbol{\mathcal{H}}_{r+1}(\boldsymbol{p}) = r$, provided $r < N/2$. To avoid ambiguity in the aspect ratio of the Hankel matrix, we restrict our attention to the case $L \leq N - L + 1$, that is, $L \leq \frac{N+1}{2}$. This assumption simplifies the formulation without loss of generality.

A topic closely related to low-rank Hankel matrices is the linear recurrence relation (LRR). If $\boldsymbol{\mathcal{H}}_{r+1}(\boldsymbol{p})$ is not of full column rank, then there exists a nonzero vector $\boldsymbol{\theta} = [\theta_1 \cdots \theta_{r+1}]^\top$ that satisfies the homogeneous linear system $\boldsymbol{\mathcal{H}}_{r+1}(\boldsymbol{p})\boldsymbol{\theta} = \boldsymbol{0}$. By expanding this equation element-wise, we obtain

$$p_n\theta_1 + p_{n+1}\theta_2 + \cdots + p_{n+r}\theta_{r+1} = 0, \tag{4}$$

which holds for all n in the domain where the indices are defined. In general, if there exists a nonzero vector $\boldsymbol{\theta}$ such that (4) is satisfied, the sequence $\boldsymbol{p}$ is said to satisfy an LRR of order r, and $\boldsymbol{\theta}$ is referred to as the coefficient vector of the LRR. Clearly, any signal of rank r satisfies an LRR of order r. Among these, a natural choice is the one with the smallest order, referred to as the minimal LRR. This minimal LRR plays an important role, as it provides a canonical characterization of the signal.

Lemma 2 ([3, Lemma 3.2])**.** For a signal of finite rank r, the minimal order LRR is of order r. The coefficient vector of the LRR is unique up to multiplication by a constant factor, and can be found from the nullspace of the associated Hankel matrix, as $\boldsymbol{\mathcal{H}}_{r+1}(\boldsymbol{p})\boldsymbol{\theta} = \boldsymbol{0}$.

The following lemma shows that the rank of any signal in the form (3) is determined by the paremeters ω_k and the polynomial degree of $g_k(t)$ provided that the signal length N is sufficiently large.

Lemma 3. A real-valued sequence $\boldsymbol{p}$ satisfies an LRR (4) with real coefficients $\boldsymbol{\theta} \in \mathbb{R}^{r+1}$ such that $\theta_1, \theta_{r+1} \neq 0$ if and only if it can be represented in the form (3). Moreover, suppose that the parameter pairs (d_k, ω_k) are all distinct, and define $r_k = 2$ if $0 < \omega_k < 0.5$ and $r_k = 1$ otherwise. Then the rank r of the sequence $\boldsymbol{p}$ is given by $r = \sum_{k=1}^{R}(m_k + 1)r_k$.

The first part of this lemma follows from [3, Corollary 3.2], and the second part is derived in [22, Appendix].

Moreover, it is known that the coefficient vector $\boldsymbol{\theta}$ of the minimal LRR of order r not only encodes a recurrence satisfied by the signal but also enables recovery of the essential parameters d_k and ω_k appearing in (3), provided that $\boldsymbol{p}$ is a

sum of exponentially decaying sinusoids. Assuming the signal $\boldsymbol{p}$ follows the form (3) with $g_j(t) = 1$, and there exists a nonzero vector $\boldsymbol{\theta}$ such that $\mathcal{H}_{r+1}(\boldsymbol{p})\boldsymbol{\theta} = \mathbf{0}$. The minimal LRR with coefficient vector $\boldsymbol{\theta}$ leads to the characteristic polynomial:

$$\psi(z) = \theta_1 + \theta_2 z + \cdots + \theta_{r+1} z^r.$$

The roots of $\psi(z)$, denoted by z_k, correspond to the complex exponential components z_k^n present in the signal. Specifically, each root can be expressed as $z_k = \exp(d_k + \mathrm{i}\omega_k)$, from which the decay rate d_k and frequency ω_k can be obtained as the real and imaginary parts of $\log z_k$, respectively. This approach is well known as Prony's method [8].

In this study, we consider the setting in which a noisy observation $\boldsymbol{p} = \bar{\boldsymbol{p}} + \boldsymbol{\xi}$ is obtained from a true signal $\bar{\boldsymbol{p}}$ of rank r that follows the model form given in (3). Our approach is based on the assumption that a reasonable estimate of $\bar{\boldsymbol{p}}$ can be obtained by selecting, among all signals $\hat{\boldsymbol{p}}$ of the same length and rank r, the one that deviates the least from the observed signal $\boldsymbol{p}$. The optimization problem introduced in (1) arises naturally from this estimation strategy.

III. Variable Projection method

This section presents a brief overview of the VP method and discusses its conventional initialization strategy.

A. Formulation of VP

To make the rank constraint in (1) more explicit, we reformulate it as the existence of a nonzero vector $\boldsymbol{x} \in \mathbb{R}^{r+1}$ satisfying

$$\mathcal{H}_{r+1}(\hat{\boldsymbol{p}})\boldsymbol{x} = \mathbf{0}. \tag{5}$$

This is justified by the fact that any matrix satisfying the rank constraint must have a nontrivial nullspace. Let

$$\Delta\boldsymbol{p} := \mathbf{C}_{\mathbf{W}}(\boldsymbol{p} - \hat{\boldsymbol{p}}) \in \mathbb{R}^N \tag{6}$$

be a new optimization variable, where $\mathbf{C}_{\mathbf{W}} \in \mathbb{R}^{N\times N}$ is the Choresky factor of $\mathbf{W}$ (i.e., $\mathbf{W} = \mathbf{C}_{\mathbf{W}}^\top \mathbf{C}_{\mathbf{W}}$). This variable substitution simplifies the weighted norm in the objective function of (1) as follows:

$$\|\hat{\boldsymbol{p}} - \boldsymbol{p}\|_{\mathbf{W}}^2 = \|\Delta\boldsymbol{p}\|_2^2. \tag{7}$$

Next, we rewrite the constraint condition of (1) using the new variable $\Delta\boldsymbol{p}$. Since $\hat{\boldsymbol{p}} = \boldsymbol{p} - \mathbf{C}_{\mathbf{W}}^{-1}\Delta\boldsymbol{p}$, the constraint $\mathcal{H}_{r+1}(\hat{\boldsymbol{p}})\boldsymbol{x} = \mathbf{0}$ becomes

$$\begin{aligned}\mathcal{H}_{r+1}(\hat{\boldsymbol{p}})\boldsymbol{x} &= \left(\mathcal{H}_{r+1}(\boldsymbol{p}) - \mathcal{H}_{r+1}(\mathbf{C}_{\mathbf{W}}^{-1}\Delta\boldsymbol{p})\right)\boldsymbol{x} = \mathbf{0} \\ &\iff \mathcal{H}_{r+1}(\mathbf{C}_{\mathbf{W}}^{-1}\Delta\boldsymbol{p})\boldsymbol{x} = \mathcal{H}_{r+1}(\boldsymbol{p})\boldsymbol{x}. \end{aligned} \tag{8}$$

To express the constraint in terms of the vector $\boldsymbol{p}$, we rewrite $\mathcal{H}_{r+1}(\boldsymbol{p})\boldsymbol{x}$ using the vectorization operator (vec) and the Kronecker product ($\otimes$) as follows:

$$\begin{aligned}\mathcal{H}_{r+1}(\boldsymbol{p})\boldsymbol{x} &= \operatorname{vec}\left((\mathcal{H}_{r+1}(\boldsymbol{p})\boldsymbol{x})^\top\right) \\ &= (\mathbf{I}_K \otimes \boldsymbol{x}^\top)\operatorname{vec}(\mathcal{H}_{r+1}^\top(\boldsymbol{p})) \\ &= (\mathbf{I}_K \otimes \boldsymbol{x}^\top)\operatorname{vec}\sum_{i=1}^{N}(\mathcal{H}_{r+1}^\top(\boldsymbol{e}_i)p_i) \\ &= (\mathbf{I}_K \otimes \boldsymbol{x}^\top)\boldsymbol{\mathcal{S}}\boldsymbol{p},\end{aligned}$$

where $\boldsymbol{e}_i$ is the unit vector whose i-th element equals to 1, $K := N - r$ denotes the column dimension of $\mathcal{H}_{r+1}(\boldsymbol{p})$, and $\boldsymbol{\mathcal{S}} := \left[\operatorname{vec}\mathcal{H}_{r+1}^\top(\boldsymbol{e}_1) \cdots \operatorname{vec}\mathcal{H}_{r+1}^\top(\boldsymbol{e}_N)\right] \in \mathbb{R}^{K(r+1)\times N}$.

Based on the variable projection principle [20], the original problem (1) can now be reformulated as a two-level minimization problem:

$$\underset{\boldsymbol{x}}{\text{minimize}}\, \mathcal{L}(\boldsymbol{x}) \tag{9}$$

$$\mathcal{L}(\boldsymbol{x}) := \min_{\Delta\boldsymbol{p}} \|\Delta\boldsymbol{p}\|_2^2 \;\text{ subject to }\; \mathbf{G}(\boldsymbol{x})\Delta\boldsymbol{p} = \boldsymbol{h}(\boldsymbol{x}), \tag{10}$$

where $\mathbf{G}(\boldsymbol{x}) := (\mathbf{I}_K \otimes \boldsymbol{x}^\top)\boldsymbol{\mathcal{S}}\mathbf{C}_{\mathbf{W}}^{-1} \in \mathbb{R}^{K\times N}$ and $\boldsymbol{h}(\boldsymbol{x}) = \mathcal{H}_{r+1}(\boldsymbol{p})\boldsymbol{x} \in \mathbb{R}^K$. If $\mathbf{G}(\boldsymbol{x})$ is of full row rank, the solution of constrained least-squares problem (10) and the corresponding objective value are given by

$$\Delta\boldsymbol{p}_*(\boldsymbol{x}) := \mathbf{G}^\top(\boldsymbol{x})\boldsymbol{\Gamma}^{-1}\boldsymbol{h}(\boldsymbol{x}), \tag{11}$$

$$\mathcal{L}(\boldsymbol{x}) = \|\Delta\boldsymbol{p}_*(\boldsymbol{x})\|_2^2 \tag{12}$$

where $\boldsymbol{\Gamma} = \boldsymbol{\Gamma}(\boldsymbol{x}) := \mathbf{G}(\boldsymbol{x})\mathbf{G}^\top(\boldsymbol{x}) \in \mathbb{R}^{K\times K}$.

Finally, we obtain an equivalent formulation of (1):

$$\underset{\boldsymbol{x}}{\text{minimize}}\, \|\mathbf{G}^\top(\boldsymbol{x})\boldsymbol{\Gamma}^{-1}\boldsymbol{h}(\boldsymbol{x})\|_2^2 \tag{13}$$

In this formulation, the objective function $\mathcal{L}(\boldsymbol{x})$ is defined as the squared ℓ_2-norm of the optimal solution vector $\Delta\boldsymbol{p}_*(\boldsymbol{x})$, whose Jacobian can be derived analytically. Owing to these properties, it has been shown in [19] that the optimization problem can be efficiently solved using the Levenberg–Marquardt method.

B. Initialization of VP

As with other gradient-based methods, solving this problem requires an initial estimate of $\boldsymbol{x}$. In this formulation, $\boldsymbol{x}$ lies in the one-dimensional null space of the approximated Hankel matrix $\mathcal{H}_{r+1}(\hat{\boldsymbol{p}})$. However, since $\hat{\boldsymbol{p}}$ is not available at initialization, the right singular vector corresponding to the smallest singular value of $\mathcal{H}_{r+1}(\boldsymbol{p})$–the Hankel matrix constructed from the observed signal–is usually used [21]. This choice corresponds to an unstructured low-rank approximation of $\mathcal{H}_{r+1}(\boldsymbol{p})$ and provides a reasonable starting point for the optimization. Since the problem is non-convex and may have multiple local minima, the choice of initial $\boldsymbol{x}$ has a significant impact on the solution. Its proper selection is, therefore, critical to the overall performance.

Hereafter, we denote by $\boldsymbol{x}^{(\mathrm{VP})}(\boldsymbol{v})$ the stationary point obtained by applying the VP method to problem (13) with initial value $\boldsymbol{v}$. The corresponding objective function value is written

as $\mathcal{L}^{(\mathrm{VP})}(\boldsymbol{v}) := \mathcal{L}(\boldsymbol{x}^{(\mathrm{VP})}(\boldsymbol{v}))$, and the resulting signal estimate is denoted by $\boldsymbol{p}^{(\mathrm{VP})}(\boldsymbol{v}) := \boldsymbol{p} - \Delta\boldsymbol{p}_*(\boldsymbol{x}^{(\mathrm{VP})}(\boldsymbol{v}))$.

IV. PROPOSED METHOD

A. Limitations of Existing Initialization under Singular Value Switching

As noted in the previous section, a common initialization strategy is to use the right singular vector corresponding to the smallest singular value of the Hankel matrix $\boldsymbol{\mathcal{H}}_{r+1}(\boldsymbol{p})$. This approach implicitly assumes that the corresponding singular vector is close to the true singular vector $\bar{\boldsymbol{v}}_1$, to which the solution $\boldsymbol{x}$ is expected to converge. In this subsection, we analyze cases where this assumption may fail due to perturbations in $\boldsymbol{p}$, and show that such failures are closely related to a phenomenon we refer to as *singular value switching*.

Suppose the true signal is denoted by $\bar{\boldsymbol{p}}$, and let $\boldsymbol{p} = \bar{\boldsymbol{p}} + \delta\boldsymbol{p}$ be its perturbed version. The singular values of the Hankel matrices $\boldsymbol{\mathcal{H}}_{r+1}(\bar{\boldsymbol{p}})$ and $\boldsymbol{\mathcal{H}}_{r+1}(\boldsymbol{p})$ are given by $0 = \bar{\sigma}_1 < \bar{\sigma}_2 \leq \cdots \leq \bar{\sigma}_L$ and $\sigma_1 \leq \cdots \leq \sigma_L$, respectively, with corresponding right singular vectors denoted by $\bar{\boldsymbol{v}}_j$ and $\boldsymbol{v}_j$ for $j = 1, \ldots, L$. We denote the changes in $\bar{\sigma}_1$ and $\bar{\sigma}_2$ due to the perturbation by $\delta\bar{\sigma}_1$ and $\delta\bar{\sigma}_2$, respectively.

In certain algorithms, such as the initialization method described in section III-B, it is assumed that $\boldsymbol{v}_1$ is sufficiently close to $\bar{\boldsymbol{v}}_1$—in other words, that $\sigma_1 = \bar{\sigma}_1 + \delta\bar{\sigma}_1 < \bar{\sigma}_2 + \delta\bar{\sigma}_2 = \sigma_2$. However, this inequality is not guaranteed to hold, and when it fails, the ordering of the singular values may change.

In the presence of noise, the perturbations $\delta\bar{\sigma}_j$ may alter the singular value structure with realistic possibility; for example, $\delta\bar{\sigma}_1$ and $\delta\bar{\sigma}_2$ of comparable magnitude but opposite sign can result in a reordering of $\bar{\sigma}_1$ and $\bar{\sigma}_2$. In such cases, the singular vector $\boldsymbol{v}_j$ (for some $j \geq 2$) may resemble the true singular vector $\bar{\boldsymbol{v}}_1$ more closely than $\boldsymbol{v}_1$ does. We refer to this situation, where a singular vector $\boldsymbol{v}_j$ (with $j \geq 2$) exhibits greater similarity to $\bar{\boldsymbol{v}}_1$ than $\boldsymbol{v}_1$, as *singular value switching*.

This phenomenon pose a challenge to existing initialization methods. When singular value switching occurs, relying solely on the right singular vector corresponding to the smallest singular value of $\boldsymbol{\mathcal{H}}_{r+1}(\boldsymbol{p})$ may lead to poor alignment with the true solution, thereby degrading the performance of the overall optimization.

B. Proposed Method

Method 1 (M1): To address the issue of singular value switching, we aim to identify a more suitable initial vector for $\boldsymbol{x}$ from among the right singular vectors beyond just $\boldsymbol{v}_1$. Inspired by the multi-start method commonly used in local optimization [24], we evaluate the objective function (13) for each of the α candidate initial vectors $\boldsymbol{v}_i$ $(i = 1, \ldots, \alpha)$, and select the one that yields the smallest value. Specifically, we define

$$i_1 := \arg\min_{i=1,\ldots,\alpha} \mathcal{L}^{(\mathrm{VP})}(\boldsymbol{v}_i),$$

and use $\boldsymbol{v}_{i_1}$ as the initial point for the VP method.

Method 2 (M2): In practice, perturbations affect not only the singular values but also the singular vectors. Depending on the magnitude and direction of these perturbations, it is possible that none of the right singular vectors $\boldsymbol{v}_i$ $(i = 1, \ldots, L)$ of $\boldsymbol{\mathcal{H}}_{r+1}(\boldsymbol{p})$ are sufficiently aligned with the true singular vector $\bar{\boldsymbol{v}}_1$. In such cases, using a single singular vector as the initial estimate for $\boldsymbol{x}$ may lead to suboptimal performance. To mitigate the potential misalignment caused by such perturbations, we extend the candidate set to include random linear combinations of the singular vectors. Specifically, we construct $\gamma := \alpha + \beta$ initial vectors of the form

$$\boldsymbol{x}^{(i)} = \begin{cases} \boldsymbol{v}_i & \text{for} \quad i = 1, \ldots, \alpha \\ \mathbf{B}\boldsymbol{c}^{(i-\alpha)} & \text{for} \quad i = \alpha+1, \ldots, \gamma, \end{cases} \tag{14}$$

where $\mathbf{B} := [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_\alpha] \in \mathbb{R}^{L\times\alpha}$, and $\boldsymbol{c}^{(i)} \in \mathbb{R}^\alpha$ is a random coefficient vector with entries sampled independently from the uniform distribution on $[-1, 1]$. Each candidate $\boldsymbol{x}^{(i)}$ $(i = 1, \ldots, \gamma)$ is evaluated using the VP objective $\mathcal{L}^{(\mathrm{VP})}(\boldsymbol{x}^{(i)})$, and the one with the smallest value is selected:

$$i_2 := \arg\min_{i=1,\ldots,\gamma} \mathcal{L}^{(\mathrm{VP})}(\boldsymbol{x}^{(i)}). \tag{15}$$

V. NUMERICAL EXPERIMENTS

To evaluate the effectiveness of the proposed initialization strategies, we conducted a denoising experiment using synthetic signals generated from the low-rank signal model (3). In this experiment, we set $g_j(t) = 1$, reducing the model to a sum of exponentially decaying sinusoids. The parameters were set to $\alpha = 3$ and $\beta = 15$, and the weighting matrix $\mathbf{W}$ was chosen as the identity matrix. MATLAB [25] was employed for the numerical experiments. The implementation provided in [21] was used with its default parameters to solve the optimization problem in (13).

We generated a clean signal $\bar{\boldsymbol{p}} \in \mathbb{R}^N$ according to this model with $R = 5$ and $N = 40$, using randomly sampled parameters: decay rates d_j from the uniform distribution over $[0.9, 1.1]$, frequencies ω_j from $[10^{-6}, 1 - 10^{-6}]$, and phases ϕ_j from $[0, \pi]$. The observed signal $\boldsymbol{p}$ was then constructed by adding Gaussian noise:

$$\boldsymbol{p} = \bar{\boldsymbol{p}} + \boldsymbol{\xi},$$

where the noise term $\boldsymbol{\xi}$ follows a white Gaussian distribution with standard deviation η. We compared the performance of the baseline VP method (denoted as C) and the proposed methods (M1 and M2) in terms of the following two metrics:

1) Estimation error: $\|\boldsymbol{p}^{(\mathrm{VP})}(\boldsymbol{x}^{(i)}) - \bar{\boldsymbol{p}}\|$,
2) Normalized objective value: $\mathcal{L}^{(\mathrm{VP})}(\boldsymbol{x}^{(i)})/\|\boldsymbol{p}\|$.

These metrics respectively quantify the accuracy of signal recovery and the quality of the solution in terms of the VP objective function. Fig. 2 shows box plots of the estimation errors and normalized objective values over 100 trials for C, M1 and M2. It can be observed that the proposed methods significantly reduce the occurrence of extreme estimation error. In particular, the spread of errors is narrower for M1 and even more so for M2, indicating improved robustness against poor initializations. To further quantify this improvement, we count

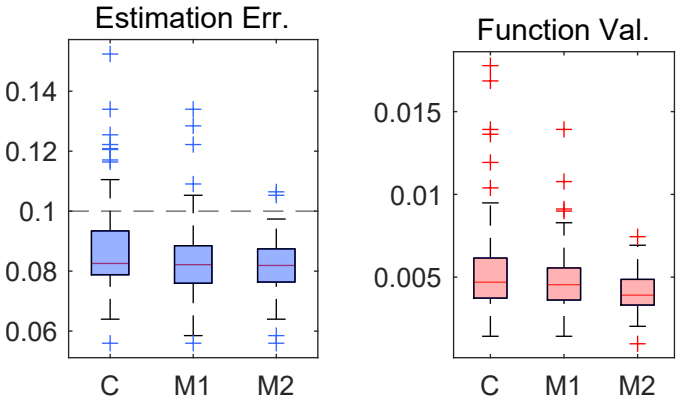

Fig. 2: Performance comparison of C, M1, and M2 at noise level $\eta = 0.1$. In the left pannel, the dashed horizontal line indicates the noise standard deviation.

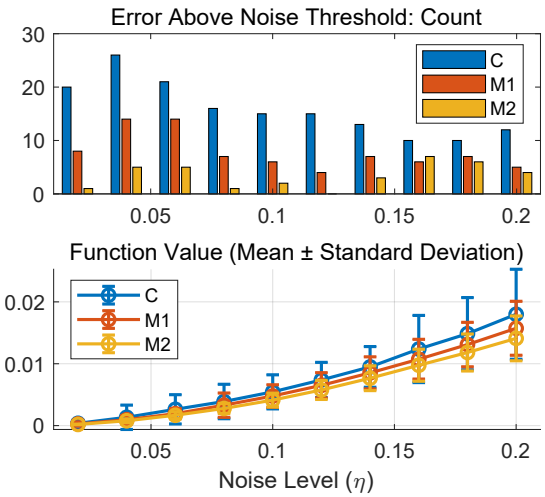

Fig. 3: Performance comparison of C, M1, and M2 under varying noise levels $\eta = 0.02, 0.04, \ldots, 0.2$. Upper: Number of trials where the reconstruction error exceeds the noise standard deviation. Lower: Normalized VP objective function value.

the number of trials in which the reconstruction error exceeds the standard deviation of the added noise—that is, cases where the reconstructed signal is farther from the ground truth than the noisy observation itself. The number of such cases was 15 for C, 6 for M1, and 2 for M2. These results clearly demonstrate the effectiveness of the proposed initialization strategies in improving the reliability of the VP method under noisy conditions.

To further assess the robustness of the proposed methods, we evaluated their performance across a range of noise levels $\eta = 0.02, 0.04, \ldots, 0.2$. The evaluation focused on two metrics introduced earlier: the normalized objective function value and the estimation error. Since the proposed methods are designed to mitigate severe estimation failures, we also counted the number of trials in which the estimation error exceeded the noise standard deviation. This count serves as an indicator of reconstruction reliability under noisy conditions. Fig. 3 summarizes the results based on these two metrics. Its upper panel shows the number of trials in which the reconstruction error surpasses the noise level, thus illustrating cases where the reconstructed signal is less accurate than the original noisy observation. The counts decrease substantially from C to M1 to M2, demonstrating that the proposed methods effectively suppress large estimation errors[1]. The lower panel shows the normalized VP objective function values. Again, M1 and M2 outperform the baseline C, with M2 achieving the lowest values across all noise levels. These results demonstrate that the proposed initialization strategies enhance robustness under noisy conditions.

To examine reconstruction performance under varying signal lengths, we fixed the noise level at $\eta = 0.02$ and used a synthetic signal composed of five sine waves to avoid numerical issues from exponential decay. For each signal length $N = 50, 100, 200, 400, 800$, we randomly generated the signal and noise 100 times and evaluated the two metrics used in the previous experiment: mean reconstruction error and the number of trials where the error exceeds the noise standard deviation. Table I summarizes the results. Both M1 and M2 outperformed the baseline method C, with M2 consistently achieving the lowest errors and near-zero exceedance counts across all signal lengths.

TABLE I: Reconstruction performance across varying signal lengths ($\eta = 0.02$).

N	50	100	200	400	800
Mean Estimation Error					
C	0.031	0.043	0.039	0.049	0.091
M1	0.025	0.016	0.011	0.014	0.036
M2	0.013	0.009	0.006	0.005	0.003
Error Above Noise Threshold: Count					
C	14	16	10	10	18
M1	9	8	1	3	9
M2	0	0	0	1	0

VI. Conclusions

This paper has presented two novel initialization strategies for solving the Hankel structured low-rank approximation problem via the variable projection method. Through an analysis of the impact of singular value switching, we demonstrated that conventional initialization using the smallest singular vector can be unreliable under noisy conditions. The proposed methods, M1 and M2, address this issue by evaluating multiple candidate vectors and incorporating random linear combinations, respectively. Experimental results on synthetic decaying sinusoidal signals confirmed that both methods outperform the baseline in terms of estimation error and objective function value, particularly under moderate to high noise levels. These findings highlight the importance of robust initialization in structured low-rank optimization. Future work includes a more detailed analysis of the conditions under which singular value

[1]In addition to reducing the number of large errors, we also confirmed that both M1 and M2 consistently achieve average estimation errors that are comparable to or slightly lower than those of C.

switching is likely to occur.

REFERENCES

[1] I. Markovsky, *Low Rank Approximation: Algorithms, Implementation, Applications*, 2nd. Springer London, 2018.

[2] G. Heinig and K. Rost, *Algebraic methods for Toeplitz-lie matrices and operators*. Akademie-Verlag, 1984.

[3] J. Gillard and K. Usevich, "Hankel low-rank approximation and completion in time series analysis and forecasting: A brief review," *Statistics and Its Interface*, vol. 16, pp. 287–303, 2 2023.

[4] J. Razavilar, Y. Li, and K. J. R. Liu, "Spectral estimation based on structured low rank matrix pencil," in *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*, vol. 5, 1996, 2503–2506 vol. 5.

[5] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Prentice Hall, 2005.

[6] I. Markovsky and P. L. Dragotti, "Using hankel structured low-rank approximation for sparse signal recovery," in *Latent Variable Analysis and Signal Separation*, Y. Deville, S. Gannot, R. Mason, M. D. Plumbley, and D. Ward, Eds., Springer International Publishing, 2018, pp. 479–487.

[7] L. Condat and A. Hirabayashi, "Cadzow denoising upgraded: A new projection method for the recovery of dirac pulses from noisy linear measurements," *Sampling Theory in Signal and Image Processing*, vol. 14, pp. 17–47, 1 2015.

[8] T. Lobos, J. Rezmer, and J. Schegner, "Parameter estimation of distorted signals using prony method," in *2003 IEEE Bologna Power Tech Conference Proceedings*, vol. 4, 2003, p. 5. DOI: 10.1109/PTC.2003.1304801.

[9] R. Zhang and G. Plonka, "Optimal approximation with exponential sums by a maximum likelihood modification of prony's method," *Advances in Computational Mathematics*, vol. 45, pp. 1657–1687, 3 2019.

[10] P. J. Shin, P. E. Z. Larson, M. A. Ohliger, *et al.*, "Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion," *Magnetic Resonance in Medicine*, vol. 72, pp. 959–970, 4 Oct. 2014.

[11] X. Zhang, H. Lu, D. Guo, *et al.*, "Accelerated mri reconstruction with separable and enhanced low-rank hankel regularization," *IEEE Transactions on Medical Imaging*, vol. 41, pp. 2486–2498, 9 2022.

[12] V. Oropeza and M. Sacchi, "Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis," *GEOPHYSICS*, vol. 76, pp. V25–V32, 3 2011.

[13] J. He, Y. Wang, and Y. Zhou, "Three dimensional seismic data reconstruction based on truncated nuclear norm," *Journal of Applied Geophysics*, vol. 214, p. 105 049, 2023.

[14] G. Ottaviani, P.-J. Spaenlehauer, and B. Sturmfels, "Exact solutions in structured low-rank approximation," *SIAM Journal on Matrix Analysis and Applications*, vol. 35, pp. 1521–1542, 4 2014.

[15] J. A. Cadzow, "Signal enhancement-a composite property mapping algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, pp. 49–62, 1 1988.

[16] J. R. Winkler and M. Hasan, "An improved non-linear method for the computation of a structured low rank approximation of the sylvester resultant matrix," *Journal of Computational and Applied Mathematics*, vol. 237, pp. 253–268, 1 Jan. 2013.

[17] K. Nagasaka, "Relaxed newtonslra for approximate gcd," *Computer Algebra in Scientific Computing*, pp. 272–292, 2021.

[18] A. Fazzi, N. Guglielmi, and I. Markovsky, "A gradient system approach for hankel structured low-rank approximation," *Linear Algebra and its Applications*, vol. 623, pp. 236–257, 2021.

[19] K. Usevich and I. Markovsky, "Variable projection for affinely structured low-rank approximation in weighted 2-norms," *Journal of Computational and Applied Mathematics*, vol. 272, pp. 430–448, 2014.

[20] G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM Journal on numerical analysis*, vol. 10, pp. 413–432, 2 1973.

[21] I. Markovsky and K. Usevich, "Software for weighted structured low-rank approximation," *Journal of Computational and Applied Mathematics*, vol. 256, pp. 278–292, 2014.

[22] N. Zvonarev and N. Golyandina, "Fast and stable modification of the gauss-newton method for low-rank signal estimation," *Numerical Linear Algebra with Applications*, vol. 29, 4 2022.

[23] A. T. Natsuki Yoshino, "Methods for constructing initial estimates for quotient polynomials in approximate greatest common divisor computation via variable projection," *un-published*, 2025.

[24] B. Peng, F. Yang, C. Yan, X. Zeng, and D. Zhou, "Efficient multiple starting point optimization for automated analog circuit optimization via recycling simulation data," in *2016 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2016, pp. 1417–1422.

[25] T. M. Inc., *Matlab version: 24.1.0 (r2024a)*, Natick, Massachusetts, United States, 2024. [Online]. Available: https://www.mathworks.com.