On the Local Convergence of an Iterative Approach for Inverse Singular Value Problems

Zheng-jian Bai* Benedetta Morini† Shu-fang Xu‡

Abstract

The purpose of this paper is to provide the convergence theory for the iterative approach given by Chu [*SIAM J. Numer. Anal.,29 (1992), pp. 885–903*] in the context of solving inverse singular value problems. We give a detailed convergence analysis and investigate the ultimate rate of convergence. Numerical results which confirm our theory are presented.

Keywords. inverse problems, singular values, nonlinear equations, Newton method.

AMS subject classifications. 15A29, 15A18, 65F15.

1 Introduction

Inverse problems arise in many practical situations such as medical imaging, exploration geophysics, and non-destructive evaluation where some general properties, for instance matrices, are to be determined from known data, e.g. eigenvalues, singular values, some prescribed entries. We refer to Chu and Golub [4] and Xu [11] for a comprehensive survey on structured and unstructured inverse eigenvalue and inverse singular value problems.

In this paper we consider the inverse singular value problem which is formally defined as follows.

Problem ISVP *Given n real $m \times n$ matrices $\{A_i\}_{i=1}^n$, $m \geq n$ and n nonnegative real numbers $\sigma_1^* \geq \sigma_2^* \geq \cdots \geq \sigma_n^*$, find $\mathbf{c} \in \mathbb{R}^n$ such that the singular values of the matrix*

$$A(\mathbf{c}) \equiv c_1 A_1 + c_2 A_2 + \cdots + c_n A_n \tag{1}$$

are precisely $\sigma_1^*, \ldots, \sigma_n^*$.

This problem first proposed by Chu [1], where two numerical methods for solving Problem ISVP are presented. We restrict our attention to the second method of [1] which generalizes an effective iterative process proposed originally by Friedland, Nocedal, and Overton [6] for solving inverse eigenvalue problems. In [1] it is shown that the iterative approach is a variation of the Newton method and some convergence theory is provided.

*Department of Mathematics, Chinese University of Hong Kong, Shatin, NT, Hong Kong, China. The research was partially supported by the Hong Kong Research Grant Council Grant CUHK4243/01P and CUHK DAG 2060220.(zjbai@math.cuhk.edu.hk)

†Dipartimento di Energetica 'S. Stecco' Università di Firenze, Via C. Lombroso 6/17, 50134 Firenze. Research was partially supported by GNCS and MIUR, Rome, Italy, through "Cofinanziamenti Programmi di Ricerca Scientifica di Interesse Nazionale".(benedetta.morini@unifi.it)

‡School of Mathematical Sciences, Peking University, Beijing 100871, People's Republic of China.(xsf@math.pku.edu.cn)

However, several theoretical issues raised in [1] deserve further attention. Here we show that the proof of local quadratic convergence in the quotient sense given in [1, Theorem 4.2] is incorrect. In addition, it seems to us that it is not clear how to derive the locally quadratic convergence of the iterative method proceeding as in [1, Theorem 4.2]. Our purpose is to fill this gap by laying down a detailed convergence analysis of the iterative approach. Our analysis reveals that the iterative method converges at least quadratically in the root sense. This is a weaker notion of convergence than quadratic convergence in the quotient sense. Thus it does not contradict the claim of [1]. In fact, proving the stronger result as stated in [1] remains an open issue.

In §2 we review the formulation and theory of the iterative method given in [1]. In §3 we present our convergence analysis and in §4 we show that our results are confirmed by numerical experiments.

In what follows for any vector $\mathbf{c} = (c_1, c_2, \ldots, c_n)^T \in \mathbb{R}^n$ we use $\{\sigma_i(\mathbf{c})\}_{i=1}^n$ to denote the singular values of $A(\mathbf{c})$ defined by (1), where $\sigma_1(\mathbf{c}) \geq \sigma_2(\mathbf{c}) \geq \cdots \geq \sigma_n(\mathbf{c}) \geq 0$. Assume that all the given singular values $\{\sigma_i^*\}_{i=1}^n$ are positive and distinct, and let $\Sigma_* = \mathrm{diag}(\sigma_1^*, \cdots, \sigma_n^*) \in \mathbb{R}^{m\times n}$, and $\mathcal{O}(n)$ denote the set of all orthogonal matrices in $\mathbb{R}^{n\times n}$.

2 The iterative approach

In this section, we briefly recall the second method given in [1]. Define the affine subspace $\mathcal{A} \equiv \{A(\mathbf{c}) \mid \mathbf{c} \in \mathbb{R}^n\}$ and the surface $\mathcal{M}_s(\Sigma_*) \equiv \{U\Sigma_* V^T \mid U \in \mathcal{O}(m),\ V \in \mathcal{O}(n)\}$, i.e. the set of all matrices in $\mathbb{R}^{m\times n}$ with singular values $\sigma_1^* > \sigma_2^* \cdots > \sigma_n^* > 0$. Thus, solving Problem ISVP is equivalent to finding an intersection of $\mathcal{M}_s(\Sigma_*)$ and $\mathcal{A}$. The second method of [1] can be viewed as a variation of the Newton method where each iteration is composed of two major steps.

Let $\mathbf{c}^k$ be the current iterate and X_k a "lift" of $A(\mathbf{c}^k)$ from the affine subspace $\mathcal{A}$ to the surface $\mathcal{M}_s(\Sigma_*)$. In the first step, the new iterate $\mathbf{c}^{k+1}$ is computed so that $A(\mathbf{c}^{k+1})$ is an $\mathcal{A}$-intercept of a line that is tangent to the manifold $\mathcal{M}_s(\Sigma_*)$ at X_k. This amounts to finding two skew-symmetric matrices $F_{k+1} \in \mathbb{R}^{m\times m}$, $T_{k+1} \in \mathbb{R}^{n\times n}$ and a vector $\mathbf{c}^{k+1} \in \mathbb{R}^n$ such that

$$X_k + F_{k+1}X_k - X_kT_{k+1} = A(\mathbf{c}^{k+1}). \tag{2}$$

Notice that $X_k \in \mathcal{M}_s(\Sigma_*)$ implies that there exist $U_k \in \mathcal{O}(m)$ and $V_k \in \mathcal{O}(n)$ such that $U_k^T X_k V_k = \Sigma_*$. It follows from (2) that

$$\Sigma_* + H_{k+1}\Sigma_* - \Sigma_* K_{k+1} = U_k^T A(\mathbf{c}^{k+1})V_k, \tag{3}$$

where $H_{k+1} = U_k^T F_{k+1} U_k \in \mathbb{R}^{m\times m}$ and $K_{k+1} = V_k^T T_{k+1} V_k \in \mathbb{R}^{n\times n}$ are skew-symmetric matrices.

In the second step, the matrix $A(\mathbf{c}^{k+1}) \in \mathcal{A}$ is lifted up to a new point $X_{k+1} \in \mathcal{M}_s(\Sigma_*)$ which is defined as

$$X_{k+1} \equiv U_{k+1}\Sigma_* V_{k+1}^T,$$

where U_{k+1} and V_{k+1} are two orthogonal matrices defined by

$$U_{k+1} = U_k R_{k+1} \quad \text{and} \quad V_{k+1} = V_k S_{k+1}. \tag{4}$$

Here, R_{k+1} and S_{k+1} are the Cayley transforms

$$R_{k+1} \equiv (I + \frac{1}{2}H_{k+1})(I - \frac{1}{2}H_{k+1})^{-1} \quad \text{and} \quad S_{k+1} \equiv (I + \frac{1}{2}K_{k+1})(I - \frac{1}{2}K_{k+1})^{-1}. \tag{5}$$

Overall we have

Iterative Algorithm

1. *Given* $\mathbf{c}^0$*, compute the singular value* $\{\sigma_i(\mathbf{c}^0)\}_{i=1}^n$ *and the normalized left singular vectors* $\{\mathbf{u}_i(\mathbf{c}^0)\}_{i=1}^m$ *and the normalized right singular vectors* $\{\mathbf{v}_i(\mathbf{c}^0)\}_{i=1}^n$ *of* $A(\mathbf{c}^0)$ *respectively. Let* $U_0 = [\mathbf{u}_1^0, \dots, \mathbf{u}_m^0] = [\mathbf{u}_1(\mathbf{c}^0), \dots, \mathbf{u}_m(\mathbf{c}^0)]$*,* $V_0 = [\mathbf{v}_1^0, \dots, \mathbf{v}_n^0] = [\mathbf{v}_1(\mathbf{c}^0), \dots, \mathbf{v}_n(\mathbf{c}^0)]$*, and*

$$\boldsymbol{\sigma}^0 = (\sigma_1(\mathbf{c}^0), \dots, \sigma_n(\mathbf{c}^0))^T.$$

2. *For* $k = 0, 1, 2, \dots$*, until convergence, do:*

   (a) *Form the approximate Jacobian matrix* J_k *by*

   $$[J_k]_{ij} \equiv (\mathbf{u}_i^k)^T A_j \mathbf{v}_i^k, \quad 1 \le i, j \le n. \tag{6}$$

   (b) *Solve* $\mathbf{c}^{k+1}$ *from the approximate Jacobian equation*

   $$J_k \mathbf{c}^{k+1} = \boldsymbol{\sigma}^*, \quad \boldsymbol{\sigma}^* = (\sigma_1^*, \dots, \sigma_n^*)^T. \tag{7}$$

   (c) *Form the matrix* $A(\mathbf{c}^{k+1})$ *by (1).*

   (d) *Form the matrix* $W_k \equiv U_k^T A(\mathbf{c}^{k+1}) V_k$*.*

   (e) *Compute the skew-symmetric matrices* H_{k+1} *and* K_{k+1} *by*

   $$\begin{aligned} &[H_{k+1}]_{ij} = 0 \quad \text{for } n+1 \le i \ne j \le m, \\ &[H_{k+1}]_{ij} = -[H_{k+1}]_{ji} = \frac{[W_k]_{ij}}{\sigma_j^*}, \quad \text{for } n+1 \le i \le m, 1 \le j \le n, \\ &[H_{k+1}]_{ij} = -[H_{k+1}]_{ji} = \frac{\sigma_i^*[W_k]_{ji} + \sigma_j^*[W_k]_{ij}}{(\sigma_j^*)^2 - (\sigma_i^*)^2}, \quad \text{for } 1 \le i < j \le n, \\ &[K_{k+1}]_{ij} = -[K_{k+1}]_{ji} = \frac{\sigma_i^*[W_k]_{ij} + \sigma_j^*[W_k]_{ji}}{(\sigma_j^*)^2 - (\sigma_i^*)^2}, \quad \text{for } 1 \le i < j \le n. \end{aligned} \tag{8}$$

   (f) *Compute* $U_{k+1} = [\mathbf{u}_1^{k+1}, \dots, \mathbf{u}_m^{k+1}]$ *and* $V_{k+1} = [\mathbf{v}_1^{k+1}, \dots, \mathbf{v}_n^{k+1}]$ *by solving*

   $$\begin{aligned} \left(I + \frac{1}{2}H_{k+1}\right) U_{k+1}^T &= \left(I - \frac{1}{2}H_{k+1}\right) U_k^T, \\ \left(I + \frac{1}{2}K_{k+1}\right) V_{k+1}^T &= \left(I - \frac{1}{2}K_{k+1}\right) V_k^T. \end{aligned}$$

Clearly, equating the "diagonal" equations of (3) gives rise to (7). The skew-symmetric matrices H_{k+1} and K_{k+1} are obtained by the "off-diagonal" equations in (3). The $((m-n)(m-n-1))/2$ unknowns located at the lower-right corner of H_{k+1} are set identically zeros.

The convergence behaviour of this iterative method was studied in [1]. Suppose that the ISVP has a solution $\mathbf{c}^*$ and that $A(\mathbf{c}^*) = U_* \Sigma_* V_*^T$ with $U_* \in \mathcal{O}(m)$ and $V_* \in \mathcal{O}(n)$. Let $E_k \equiv (E_1^k, E_2^k) = (U_k - U_*, V_k - V_*)$ denote the error matrix at the kth iteration, $\|\cdot\|$ denote the Euclidean vector norm or its corresponding induced matrix norm, and $\|\cdot\|_F$ denote the Frobenius norm or the induced Frobenius norm in $\mathbb{R}^{m\times m} \times \mathbb{R}^{n\times n}$. Then the following result states that the method is locally quadratically convergent.

Theorem 1 [1, Theorem 4.2] *Suppose that all singular values $\sigma_1^*, \ldots, \sigma_n^*$ are positive and distinct. Suppose also that the matrix $J^{(k)}$ defined in (6) is nonsingular. Then we have*

$$\|E_{k+1}\|_F = O(\|E_k\|_F^2) \quad \textit{and} \quad \|\mathbf{c}^{k+1} - \mathbf{c}^*\| = O(\|E_k\|_F^2).$$

In [1], this theorem was proved as follows. Let

$$U_k^T A(\mathbf{c}^*) V_k \equiv e^{\hat{M}_k} \Sigma_* e^{-\hat{N}_k} \tag{9}$$

where $e^{\hat{M}_k} = U_k^T U_*$ and $e^{\hat{N}_k} = V_k^T V_*$. By [1, Lemma 4.1]

$$\|(\hat{M}_k, \hat{N}_k)\|_F = O(\|E_k\|_F). \tag{10}$$

Together with (9), it follows that

$$U_k^T A(\mathbf{c}^*) V_k = \Sigma_* + \hat{M}_k \Sigma_* - \Sigma_* \hat{N}_k + O(\|E_k\|_F^2). \tag{11}$$

By taking the difference between (11) and (3), we get

$$U_k^T (A(\mathbf{c}^*) - A(\mathbf{c}^{k+1})) V_k = (\hat{M}_k - H_{k+1})\Sigma_* - (\hat{N}_k - K_{k+1})\Sigma_* + O(\|E_k\|_F^2). \tag{12}$$

The diagonal equations of (12) yields

$$J^{(k)}(\mathbf{c}^* - \mathbf{c}^{k+1}) = O(\|E_k\|_F^2),$$

and from the nonsingularity of $J^{(k)}$, we have

$$\|\mathbf{c}^* - \mathbf{c}^{k+1}\| = O(\|E_k\|_F^2).$$

Similarly, from the off-diagonal equations of (12) the following estimates are derived

$$\begin{aligned} \|\hat{M}_k - H_{k+1}\|_F &= O(\|E_k\|_F^2), \\ \|\hat{N}_k - K_{k+1}\|_F &= O(\|E_k\|_F^2). \end{aligned} \tag{13}$$

Because of (10), it must be that

$$\|(H_{k+1}, K_{k+1})\|_F = O(\|E_k\|_F). \tag{14}$$

We observe that

$$\begin{aligned} E_1^{k+1} &\equiv U_{k+1} - U_* = U_k R_{k+1} - U_k e^{\hat{M}_k} \\ &= U_k \left[(I + \frac{1}{2} H_{k+1}) - (I + \hat{M}_k + O(\|\hat{M}_k\|^2))(I - \frac{1}{2} H_{k+1}) \right] (I - \frac{1}{2} H_{k+1})^{-1} \\ &= U_k \left[H_{k+1} - \hat{M}_k + O(\|\hat{M}_k H_{k+1}\|) + O(\|\hat{M}_k\|^2) \right] (I - \frac{1}{2} H_{k+1})^{-1}. \end{aligned}$$

Thus it is clear now that

$$\|E_1^{k+1}\| = O(\|E_k\|_F^2).$$

A similar argument works for E_2^{k+1}. Therefore, the proof is completed.

We note that the estimate of $\|\hat{M}_k - H_{k+1}\|_F$ in (13) is incorrect. The reason is as follows. Since the system (12) shows that the $((m-n)(m-n-1))/2$ unknowns located at the lower-right corner of the matrix $\hat{M}_k - H_{k+1}$ are not bound to any equations at all, we can not ensure that (13) holds. In fact, by (10) and (8), we have only

$$|[\hat{M}_k - H_{k+1}]_{ij}| = |[\hat{M}_k]_{ij}| = O(\|E_k\|_F), \quad n+1 \le i \ne j \le m.$$

Thus as a whole, $\|\hat{M}_k - H_{k+1}\|_F = O(\|E_k\|_F)$. Therefore the quadratic convergence of the second method is not guaranteed when $m > n+1$.

In the next section, we develop the convergence analysis for the vector iterates $\{\mathbf{c}^k\}$ and the approximate singular vectors $\{U^{(k)}\}$, $\{V^{(k)}\}$. What we are more concerned is the convergence of the iterates $\{\mathbf{c}^k\}$.

3 Convergence Analysis

In what follows, we assume that $\mathbf{c}^*$ is a solution of the ISVP and let $\mathbf{c}^k$ be the kth iterate produced by the iterative algorithm.

3.1 Preliminary Lemmas

In this subsection, we give some preliminary lemmas, which are necessary for the convergence analysis. We first give the perturbation bound for singular values.

Lemma 1 [7, Corollary 8.6.2] *If B and $B+E$ are in $\mathbb{R}^{m\times n}$ with $m \ge n$, then, for any $1 \le k \le n$,*

$$|\sigma_k(B+E) - \sigma_k(B)| \le \|E\|,$$

where $\sigma_k(B)$ denotes the kth largest singular value of B.

Then, we provide an approximation to the Cayley transform.

Lemma 2 *If $E \in \mathbb{R}^{n\times n}$ and $\|E\| < 1$, then $I - \frac{1}{2}E$ is nonsingular and*

$$\left\| \left(I + \frac{1}{2}E\right)\left(I - \frac{1}{2}E\right)^{-1} - (I+E) \right\| \le \|E\|^2. \tag{15}$$

Proof: It is obvious that $I - \frac{1}{2}E$ is nonsingular. In the following, we will show that (15) holds. It is easy to verify that

$$\left(I - \frac{1}{2}E\right)^{-1} = I + \frac{1}{2}E + \frac{1}{4}E^2\left(I - \frac{1}{2}E\right)^{-1},$$

and so it follows that

$$\left(I + \frac{1}{2}E\right)\left(I - \frac{1}{2}E\right)^{-1} = I + E + \frac{1}{4}E^2\left(I + \left(I + \frac{1}{2}E\right)\left(I - \frac{1}{2}E\right)^{-1}\right). \tag{16}$$

Since $\|E\| < 1$ we have that

$$\left\|\left(I - \frac{1}{2}E\right)^{-1}\right\| \le \frac{1}{1 - \frac{1}{2}\|E\|} < 2,$$

and by (16) the inequality (15) follows. □

Next, we give the following useful result.

Lemma 3 *Let* $B \in \mathbb{R}^{n\times n}$ *and* $\Sigma = \mathrm{diag}(\sigma_1, \dots, \sigma_n) \in \mathbb{R}^{n\times n}$ *with* $\sigma_1 > \sigma_2 > \cdots > \sigma_n > 0$. *If the skew-symmetric matrices* H *and* K *satisfy that*

$$H\Sigma - \Sigma K = B, \tag{17}$$

then we have

$$H = Q \circ \left(B\Sigma + \Sigma B^T\right), \qquad \|H\| \le \frac{2n\sigma_1}{d}\|B\|, \tag{18}$$

$$K = Q \circ \left(\Sigma B + B^T\Sigma\right), \qquad \|K\| \le \frac{2n\sigma_1}{d}\|B\|, \tag{19}$$

where $\circ$ *denotes the Hadamard product,* $d = \min_{i\ne j} |\sigma_i^2 - \sigma_j^2|$, *and* $Q = [q_{ij}]$ *with*

$$q_{ij} = \begin{cases} 0, & \text{if } i = j, \\ \dfrac{1}{\sigma_j^2 - \sigma_i^2}, & \text{otherwise.} \end{cases}$$

Proof: Since $H^T = -H$ and $K^T = -K$, from (17) we have

$$-\Sigma H + K\Sigma = B^T. \tag{20}$$

Eliminating the matrix H in (17) and (20) gives rise to

$$K\Sigma^2 - \Sigma^2 K = \Sigma B + B^T\Sigma.$$

Equating the off-diagonal elements yields the expression of the matrix K as given in (19). Using the expression we have

$$\begin{aligned} \|K\|_\infty &\le \frac{1}{d}\left\|\Sigma B + B^T\Sigma\right\|_\infty \le \frac{1}{d}\left(\|\Sigma\|_\infty\|B\|_\infty + \|B^T\|_\infty\|\Sigma\|_\infty\right) \\ &\le \frac{\sigma_1}{d}\left(\|B\|_\infty + \|B^T\|_\infty\right), \end{aligned} \tag{21}$$

where $\|\cdot\|_\infty$ denotes the row sum norm. Recalling that $\|K\| \le \sqrt{n}\|K\|_\infty$, $\|B\|_\infty \le \sqrt{n}\|B\|$, $\|B^T\|_\infty \le \sqrt{n}\|B^T\| = \sqrt{n}\|B\|$, it follows from (21) that the inequality in (19) holds. Similarly, we can prove (18). □

In the following lemma, we give a perturbation bound for $A(\mathbf{c})$ defined in (1).

Lemma 4 *For any* $\mathbf{c}, \bar{\mathbf{c}} \in \mathbb{R}^n$, *we have*

$$\|A(\mathbf{c}) - A(\bar{\mathbf{c}})\| \le \left(\sum_{i=1}^n \|A_i\|^2\right)^{1/2} \|\mathbf{c} - \bar{\mathbf{c}}\|, \tag{22}$$

where $A(\mathbf{c})$ *is defined in (1).*

Proof: It follows from (1) that

$$\|A(\mathbf{c}) - A(\bar{\mathbf{c}})\| \le \sum_{i=1}^n |c_i - \bar{c}_i|\|A_i\| \le \left(\sum_{i=1}^n \|A_i\|^2\right)^{1/2} \left(\sum_{i=1}^n |c_i - \bar{c}_i|^2\right)^{1/2},$$

which is just the inequality (22). $\square$

Now, let the singular value decomposition of $A(\mathbf{c}^*)$ be

$$A(\mathbf{c}^*) = U_*\Sigma_* V_*^T, \tag{23}$$

and define

$$J_* = [(J_*)_{ij}] \quad \text{with} \quad (J_*)_{ij} = (\mathbf{u}_i^*)^T A_j \mathbf{v}_i^*, \quad 1 \le i, j \le n,$$

where $\mathbf{u}_i^*$ and $\mathbf{v}_i^*$ are the ith column of U_* and V_* respectively. In what follows we always assume that J_* is nonsingular. Thus, letting $U_* = [U_{*1}, U_{*2}]$ with $U_{*1} \in \mathbb{R}^{m\times n}$, by the continuity of the matrix inverse there exist positive numbers δ and C such that if $\mathbf{u}_i \in \mathbb{R}^m$, $\mathbf{v}_i \in \mathbb{R}^n$ satisfy

$$\max\{\|[\mathbf{u}_1, \ldots, \mathbf{u}_n] - U_{*1}\|, \|[\mathbf{v}_1, \ldots, \mathbf{v}_n] - V_*\|\} \le \delta, \tag{24}$$

then the matrix $J = [\mathbf{u}_i^T A_j \mathbf{v}_i]$ is nonsingular and

$$\|J^{-1}\| \le C. \tag{25}$$

We then define

$$\begin{aligned} &\gamma_1 = \frac{23}{2}\sigma_1^*, && \gamma_2 = \sqrt{n}\,\gamma_1 C, && \gamma_3 = \sqrt{2}\,\eta\left(\beta\gamma_2 + \gamma_1\right), \\ &\gamma_4 = \sqrt{2n}\;\eta C\beta^2, && && \eta = \frac{2n\sigma_1^*}{d_*} + \frac{1}{\sigma_n^*}, \end{aligned} \tag{26}$$

where

$$\beta = \left(\sum_{i=1}^{n} \|A_i\|^2\right)^{1/2} \quad \text{and} \quad d_* = \min_{i\ne j} |\sigma_i^{*2} - \sigma_j^{*2}|. \tag{27}$$

Also, partition Σ_* as $\Sigma_* = \begin{bmatrix} \Sigma_{*1} \\ 0 \end{bmatrix}$ with $\Sigma_{*1} \in \mathbb{R}^{n\times n}$ and U_k as $U_k = [U_{k1}, U_{k2}]$ with $U_{k1} \in \mathbb{R}^{m\times n}$ for $k = 0, 1, \ldots$. Then we have the following lemma.

Lemma 5 *If*

$$\max\{\|U_{01} - U_{*1}\|, \|V_0 - V_*\|\} \le \frac{\delta}{4}, \tag{28}$$

$$\rho_1 \equiv \sqrt{\|H_1\|^2 + \|K_1\|^2} < \min\left\{1, \frac{1}{2\gamma_3}, \frac{3\delta}{8(\gamma_3+3)}\right\} \equiv \epsilon_0, \tag{29}$$

then for any $k \ge 1$ the iterates $\{\mathbf{c}^k\}$, $\{H_k\}$, $\{K_k\}$ and $\left\{U_k^T A(\mathbf{c}^k) V_k\right\}$ generated by the iterative algorithm satisfy

$$\left\|U_k^T A(\mathbf{c}^k)V_k - \Sigma_*\right\| \le \gamma_1\left(\|H_k\|^2 + \|K_k\|^2\right), \tag{30}$$

$$\|\mathbf{c}^{k+1} - \mathbf{c}^k\| \le \gamma_2\left(\|H_k\|^2 + \|K_k\|^2\right), \tag{31}$$

$$\sqrt{\|H_{k+1}\|^2 + \|K_{k+1}\|^2} \le \gamma_3\left(\|H_k\|^2 + \|K_k\|^2\right), \tag{32}$$

$$\|U_{k+1} - U_k\| \le 2\gamma_3\left(\|H_k\|^2 + \|K_k\|^2\right), \tag{33}$$

$$\|V_{k+1} - V_k\| \le 2\gamma_3\left(\|H_k\|^2 + \|K_k\|^2\right). \tag{34}$$

Proof: By (24) and (25), (28) implies that J_0 is nonsingular and $\|J_0^{-1}\| \leq C$. Thus $\mathbf{c}_1, U_1$ and V_1 are well defined.

First we show that (30)–(34) hold for $k = 1$. By (5) and Lemma 2, we have

$$\|U_1 - U_0\| = \|U_0(R_1 - I)\| = \|R_1 - I\| \leq 2\|H_1\| \leq 2\rho_1,$$
$$\|V_1 - V_0\| = \|V_0(S_1 - I)\| = \|S_1 - I\| \leq 2\|K_1\| \leq 2\rho_1,$$

since $\max\{\|H_1\|, \|K_1\|\} < 1$. Therefore, it follows that

$$\begin{aligned}\|U_{11} - U_{*1}\| &\leq \|U_{11} - U_{01}\| + \|U_{01} - U_{*1}\| \leq \|U_1 - U_0\| + \|U_{01} - U_{*1}\| \\ &\leq 2\rho_1 + \frac{\delta}{4} < \frac{3}{4}\delta + \frac{1}{4}\delta = \delta.\end{aligned}$$

Similarly, we have

$$\|V_1 - V_*\| \leq \delta.$$

Thus by (24) and (25), we know that J_1 is nonsingular and $\|J_1^{-1}\| \leq C$.

In order to prove (30), let

$$R_1 = I + H_1 + E_1 \quad \text{and} \quad S_1 = I + K_1 + F_1. \tag{35}$$

Hence, By Lemma 2, it follows from (5) that

$$\|E_1\| \leq \|H_1\|^2 \quad \text{and} \quad \|F_1\| \leq \|K_1\|^2. \tag{36}$$

Notice that it follows from (3) and (4) that

$$U_0^T A(\mathbf{c}^1) V_0 = \Sigma_* + H_1\Sigma_* - \Sigma_* K_1, \quad U_1 = U_0 R_1 \quad \text{and} \quad V_1 = V_0 S_1.$$

Then, by (35), a short calculation gives rise to

$$U_1^T A(\mathbf{c}^1) V_1 = \Sigma_* + G_1, \tag{37}$$

where

$$\begin{aligned}G_1 = {} & H_1\left(\Sigma_* - H_1\Sigma_* + \Sigma_* K_1\right) K_1 - H_1^2\Sigma_* - \Sigma_* K_1^2 \\ & + E_1^T\left(\Sigma_* + H_1\Sigma_* - \Sigma_* K_1\right)(I + K_1) \\ & + \left(I - H_1 + E_1^T\right)\left(\Sigma_* + H_1\Sigma_* - \Sigma_* K_1\right) F_1.\end{aligned}$$

Using (36) and the assumption $\max\{\|H_1\|, \|K_1\|\} < 1$ we have

$$\begin{aligned}\|G_1\| &\leq 3\sigma_1^*\|H_1\|\|K_1\| + \sigma_1^*\|H_1\|^2 + \sigma_1^*\|K_1\|^2 + 6\sigma_1^*\|H_1\|^2 + 9\sigma_1^*\|K_1\|^2 \\ &\leq \frac{3}{2}\sigma_1^*(\|H_1\|^2 + \|K_1\|^2) + 10\sigma_1^*(\|H_1\|^2 + \|K_1\|^2) \\ &= \gamma_1(\|H_1\|^2 + \|K_1\|^2),\end{aligned} \tag{38}$$

where γ_1 is defined in (26). This shows that (30) is true for $k = 1$.

Combining (37) with

$$U_1^T A(\mathbf{c}^2) V_1 = \Sigma_* + H_2\Sigma_* - \Sigma_* K_2$$

yields

$$U_1^T\left(A(\mathbf{c}^2) - A(\mathbf{c}^1)\right) V_1 = H_2\Sigma_* - \Sigma_* K_2 - G_1. \tag{39}$$

The diagonal equations of (39) give rise to

$$J_1(\mathbf{c}^2 - \mathbf{c}^1) = \mathbf{g}_1,$$

where $\mathbf{g}_1$ is the diagonal vector of the matrix $-G_1$, and so we have

$$\|\mathbf{c}^2 - \mathbf{c}^1\| \le C\|\mathbf{g}_1\| \le C\sqrt{n}\|G_1\| \le \gamma_2(\|H_1\|^2 + \|K_1\|^2), \tag{40}$$

where γ_2 is defined in (26). This shows that (31) holds for $k = 1$. Let

$$Z \equiv U_1^T\left(A(\mathbf{c}^2) - A(\mathbf{c}^1)\right)V_1 + G_1 = \begin{bmatrix} Z_{11} \\ Z_{21} \end{bmatrix}\begin{matrix} n \\ m-n \end{matrix}$$

with $Z_{11} \in \mathbb{R}^{n\times n}$ and $Z_{21} \in \mathbb{R}^{(m-n)\times n}$. Noting that H_2 has the form

$$H_2 = \begin{bmatrix} H_{11}^{(2)} & -H_{21}^{(2)T} \\ H_{21}^{(2)} & 0 \end{bmatrix}$$

with $H_{11}^{(2)} \in \mathbb{R}^{n\times n}$, from (39) we obtain

$$H_{11}^{(2)}\Sigma_{*1} - \Sigma_{*1}K_2 = Z_{11}, \tag{41}$$

and

$$H_{21}^{(2)}\Sigma_{*1} = Z_{21}. \tag{42}$$

By Lemma 3, it follows from (41) that

$$\|H_{11}^{(2)}\| \le \frac{2n\sigma_1^*}{d_*}\|Z_{11}\|, \tag{43}$$

$$\|K_2\| \le \frac{2n\sigma_1^*}{d_*}\|Z_{11}\|. \tag{44}$$

On the other hand, by (42) we have

$$\|H_{21}^{(2)}\| \le \frac{1}{\sigma_n^*}\|Z_{21}\|.$$

This, together with (43), yields

$$\begin{aligned}\|H_2\| &\le \left\|\begin{bmatrix} H_{11}^{(2)} & 0 \\ 0 & 0 \end{bmatrix}\right\| + \left\|\begin{bmatrix} 0 & -H_{21}^{(2)T} \\ H_{21}^{(2)} & 0 \end{bmatrix}\right\| \\ &= \|H_{11}^{(2)}\| + \|H_{21}^{(2)}\| \le \frac{2n\sigma_1^*}{d_*}\|Z_{11}\| + \frac{1}{\sigma_n^*}\|Z_{21}\| \le \eta\|Z\|,\end{aligned} \tag{45}$$

where η is defined in (26). By Lemma 4, from (38) and (40) we get

$$\|Z\| \le \beta\|\mathbf{c}^2 - \mathbf{c}^1\| + \|G_1\| \le (\beta\gamma_2 + \gamma_1)(\|H_1\|^2 + \|K_1\|^2), \tag{46}$$

where β is defined in (27). Combining (46) with (44) and (45) gives rise to

$$\sqrt{\|H_2\|^2 + \|K_2\|^2} \le \gamma_3(\|H_1\|^2 + \|K_1\|^2), \tag{47}$$

where γ_3 is defined in (26). This shows that (32) is true for $k = 1$. Moreover, (47) implies that

$$\sqrt{\|H_2\|^2 + \|K_2\|^2} \leq \sqrt{\|H_1\|^2 + \|K_1\|^2}, \tag{48}$$

since we have assumed that $\sqrt{\|H_1\|^2 + \|K_1\|^2} < \frac{1}{\gamma_3}$.

Thus by Lemma 2, it follows from (48), (47) and (29) that

$$\|U_2 - U_1\| = \|U_1(R_2 - I)\| \leq 2\|H_2\| \leq 2\gamma_3\rho_1^2,$$
$$\|V_2 - V_1\| = \|V_1(S_2 - I)\| \leq 2\|K_2\| \leq 2\gamma_3\rho_1^2,$$

which shows that (33) and (34) are true for $k = 1$.

Now we show that the inequalities (30)–(34) hold for the integer k, assuming that they are true for all positive integer less than or equal to $k - 1$. From (47) and the induction assumption, we can easily derive that

$$\sqrt{\|H_k\|^2 + \|K_k\|^2} \leq \sqrt{\|H_1\|^2 + \|K_1\|^2}. \tag{49}$$

Similarly to the proof of (37), we can show that

$$U_k^T A(\mathbf{c}^k) V_k = \Sigma_* + G_k, \tag{50}$$

where

$$\|G_k\| \leq \gamma_1(\|H_k\|^2 + \|K_k\|^2). \tag{51}$$

By the induction assumptions we know that, for $j = 2, 3 \ldots, k$,

$$\|U_{j1} - U_{j-1,1}\| \leq \|U_j - U_{j-1}\| \leq 2\gamma_3\rho_{j-1}^2, \quad \|V_j - V_{j-1}\| \leq 2\gamma_3\rho_{j-1}^2, \quad \rho_j \leq \gamma_3\rho_{j-1}^2,$$

where $\rho_j = \sqrt{\|H_j\|^2 + \|K_j\|^2}$. By (29), we get $\gamma_3\rho_1 < 1/2$ and also $\gamma_3\rho_1 < 3/8\delta$. Thus we have

$$\begin{aligned}
\|V_k - V_*\| &\leq \sum_{j=1}^{k} \|V_j - V_{j-1}\| + \|V_0 - V_*\| \leq \sum_{j=2}^{k} 2\gamma_3\rho_{j-1}^2 + 2\rho_1 + \frac{\delta}{4} \\
&\leq \sum_{j=2}^{k} 2(\gamma_3\rho_1)^{2^{j-1}} + 2\rho_1 + \frac{\delta}{4} \leq 2 \cdot \frac{(\gamma_3\rho_1)^2}{1 - (\gamma_3\rho_1)^2} + 2\rho_1 + \frac{\delta}{4} \\
&\leq 2 \cdot \frac{2}{3}\gamma_3\rho_1 + 2\rho_1 + \frac{\delta}{4} \leq \frac{\delta}{2} + \frac{\delta}{4} + \frac{\delta}{4} = \delta.
\end{aligned}$$

Similarly, we can prove that

$$\|U_{k1} - U_{*1}\| \leq \delta.$$

Thus it follows from (24) and (25) that J_k is nonsingular and $\|J_k^{-1}\| \leq C$.

Combining (50) with

$$U_k^T A(\mathbf{c}^{k+1}) V_k = \Sigma_* + H_{k+1}\Sigma_* - \Sigma_* K_{k+1},$$

we have

$$U_k^T \left(A(\mathbf{c}^{k+1}) - A(\mathbf{c}^k) \right) V_k = H_{k+1}\Sigma_* - \Sigma_* K_{k+1} + G_k. \tag{52}$$

From (52), completely similar to the proofs of (40) and (47), we can derive that

$$\left\| \mathbf{c}^{k+1} - \mathbf{c}^k \right\| \leq \gamma_2(\|H_k\|^2 + \|K_k\|^2) \tag{53}$$

and

$$\sqrt{\|H_{k+1}\|^2 + \|K_{k+1}\|^2} \le \gamma_3(\|H_k\|^2 + \|K_k\|^2).$$

This, together with (49), gives rise to

$$\sqrt{\|H_{k+1}\|^2 + \|K_{k+1}\|^2} \le \sqrt{\|H_k\|^2 + \|K_k\|^2} \le \sqrt{\|H_1\|^2 + \|K_1\|^2},$$

which implies that

$$\|H_{k+1}\| \le \sqrt{\|H_{k+1}\|^2 + \|K_{k+1}\|^2} \le \sqrt{\|H_1\|^2 + \|K_1\|^2} < 1.$$

Thus, by Lemma 2 and (53), we get

$$\|U_{k+1} - U_k\| = \|U_k(R_{k+1} - I)\| = \|R_{k+1} - I\| \le 2\|H_{k+1}\| \le 2\gamma_3(\|H_k\|^2 + \|K_k\|^2).$$

Similarly, we can prove that (34) holds.

Therefore, by mathematical induction principle, we have showed that the inequalities (30)–(34) hold for all positive integers. □

Finally, we estimate the errors in $\{\mathbf{u}_i(\mathbf{c}^k)\}_{i=1}^n$ and $\{\mathbf{v}_i(\mathbf{c}^k)\}_{i=1}^n$ in terms of $\|\mathbf{c}^k - \mathbf{c}^*\|$.

Lemma 6 *Let the given singular values $\{\sigma_i^*\}_{i=1}^n$ be positive and distinct, and U_* and V_* denote associated matrices of the normalized left and normalized right singular vectors of $A(\mathbf{c}^*)$ respectively. Let the vectors $\mathbf{u}_i(\mathbf{c}^k)$ and $\mathbf{v}_i(\mathbf{c}^k)$ stand for the unit left and unit right singular vectors of $A(\mathbf{c}^k)$ respectively. Then there exist positive numbers ϵ_1 and κ such that, if $\|\mathbf{c}^k - \mathbf{c}^*\| \le \epsilon_1$, we have*

$$\|[\mathbf{u}_1(\mathbf{c}^k), \ldots, \mathbf{u}_n(\mathbf{c}^k)] - U_{*1}\| \le \kappa\|\mathbf{c}^k - \mathbf{c}^*\|, \tag{54}$$

$$\|[\mathbf{v}_1(\mathbf{c}^k), \ldots, \mathbf{v}_n(\mathbf{c}^k)] - V_*\| \le \kappa\|\mathbf{c}^k - \mathbf{c}^*\|. \tag{55}$$

Proof: It follows from the analyticity of a simple singular value and its corresponding left and right singular vectors. The proof of this lemma is similar to [11, p. 249]. Therefore we omit the proof here. □

3.2 R-Convergence Rate

In this subsection, we will show that the three sequences of the iterates $\{\mathbf{c}^k\}$, $\{U_k\}$ and $\{V_k\}$ generated by the iterative method are all at least quadratically convergent in the root sense. Here, we recall the definition of root-convergence, see [8, Chap. 9].

Definition 1 *Let $\{\mathbf{x}^k\}$ be a sequence with limit $\mathbf{x}^*$. Then the numbers*

$$R_p\{\mathbf{x}^k\} = \begin{cases} \limsup_{k\to\infty} \left\|\mathbf{x}^k - \mathbf{x}^*\right\|^{1/k}, & \text{if } p = 1, \\ \limsup_{k\to\infty} \left\|\mathbf{x}^k - \mathbf{x}^*\right\|^{1/p^k}, & \text{if } p > 1, \end{cases} \tag{56}$$

are the root-convergence factors of $\{\mathbf{x}^k\}$. The quantity

$$O_R(\mathbf{x}^*) = \begin{cases} \infty, & \text{if } R_p\{\mathbf{x}^k\} = 0, \forall p \in [1,\infty), \\ \inf\left\{p \in [1,\infty) | R_p\{\mathbf{x}^k\} = 1\right\}, & \text{otherwise}, \end{cases} \tag{57}$$

is called the root-convergence rate of $\{\mathbf{x}^k\}$.

Next, we prove the main result of this paper.

Theorem 2 *Let the given singular values* $\{\sigma_i^*\}_{i=1}^n$ *be positive and distinct. Then there exist* $\epsilon > 0, \tilde{\mathbf{c}} \in \mathbb{R}^n, \tilde{U} \in \mathcal{O}(m)$ *and* $\tilde{V} \in \mathcal{O}(m)$ *such that if* $\|\mathbf{c}^0 - \mathbf{c}^*\| \le \epsilon$*, the iterates* $\{\mathbf{c}^k\}, \{U_k\}, \{V_k\}$*, and* $\left\{U_k^T A(\mathbf{c}^k) V_k\right\}$ *generated by the iterative algorithm converge to* $\tilde{\mathbf{c}}, \tilde{U}, \tilde{V}$*, and* $\tilde{U}^T A(\tilde{\mathbf{c}})\tilde{V} = \Sigma_*$*, respectively.*

Proof: By Lemmas 1 and 4, we have

$$\max_i \left|\sigma_i(\mathbf{c}^0) - \sigma_i^*\right| \le \left\|A(\mathbf{c}^0) - A(\mathbf{c}^*)\right\| \le \beta \left\|\mathbf{c}^0 - \mathbf{c}^*\right\|. \tag{58}$$

By Lemma 6, if $\left\|\mathbf{c}^0 - \mathbf{c}^*\right\| \le \min\{\epsilon_1, \delta/(4\kappa)\}$, then

$$\max\{\|U_{01} - U_{*1}\|, \|V_0 - V_*\|\} \le \kappa \|\mathbf{c}^0 - \mathbf{c}^*\| \le \delta/4, \tag{59}$$

where δ is given in (24). Thus by (24) and (25), we know that J_0 is nonsingular and $\|J_0^{-1}\| \le C$. Note that

$$U_0^T A(\mathbf{c}^0) V_0 = \Sigma_0 = \operatorname{diag}(\sigma_1(\mathbf{c}^0), \ldots, \sigma_n(\mathbf{c}^0)), \tag{60}$$

$$U_0^T A(\mathbf{c}^1) V_0 = \Sigma_* + H_1 \Sigma_* - \Sigma_* K_1. \tag{61}$$

Taking the difference between (60) and (61) yields

$$U_0^T \left(A(\mathbf{c}^1) - A(\mathbf{c}^0)\right) V_0 = \Sigma_* - \Sigma_0 + H_1 \Sigma_* - \Sigma_* K_1. \tag{62}$$

The diagonal equations of (62) give rise to

$$J_0(\mathbf{c}^1 - \mathbf{c}^0) = \boldsymbol{\sigma}^* - \boldsymbol{\sigma}^0,$$

and so, by (58), we have

$$\left\|\mathbf{c}^1 - \mathbf{c}^0\right\| \le C\sqrt{n} \max_i |\sigma_i(\mathbf{c}^0) - \sigma_i^*| \le \sqrt{n} C \beta \left\|\mathbf{c}^0 - \mathbf{c}^*\right\|. \tag{63}$$

Similarly to the proofs of (44) and (45), from (62), we can derive that

$$\max\{\|H_1\|, \|K_1\|\} \le \eta \left\|A(\mathbf{c}^1) - A(\mathbf{c}^0)\right\|, \tag{64}$$

where η is defined in (26). By Lemma 4, it follows from (63) that

$$\left\|A(\mathbf{c}^1) - A(\mathbf{c}^0)\right\| \le \beta \left\|\mathbf{c}^1 - \mathbf{c}^0\right\| \le \sqrt{n} C \beta^2 \left\|\mathbf{c}^0 - \mathbf{c}^*\right\|. \tag{65}$$

Then, from (64) and (65), we get

$$\max\{\|H_1\|, \|K_1\|\} \le \eta \sqrt{n} C \beta^2 \left\|\mathbf{c}^0 - \mathbf{c}^*\right\|,$$

and so we have

$$\sqrt{\|H_1\|^2 + \|K_1\|^2} \le \eta \sqrt{2n} C \beta^2 \left\|\mathbf{c}^0 - \mathbf{c}^*\right\| = \gamma_4 \left\|\mathbf{c}^0 - \mathbf{c}^*\right\|, \tag{66}$$

where γ_4 is defined in (26).

Now we let $\gamma = \max\{\gamma_1, \gamma_2, \gamma_3\}$. If $\left\|\mathbf{c}^0 - \mathbf{c}^*\right\| \le \epsilon$, where

$$\epsilon < \min\left\{1, \epsilon_1, \frac{\delta}{4\kappa}, \frac{\epsilon_0}{\gamma_4}, \frac{1}{\gamma_4\gamma}\right\}, \tag{67}$$

then, from (59) and (66), we obtain

$$\max\{\|U_{01} - U_{*1}\|, \|V_0 - V_*\|\} \le \frac{\delta}{4} \quad \text{and} \quad \rho_1 \equiv \sqrt{\|H_1\|^2 + \|K_1\|^2} < \epsilon_0.$$

Thus by Lemma 5, we have, for any $k \ge 1$,

$$\begin{array}{lll} \left\|\mathbf{c}^{k+1} - \mathbf{c}^k\right\| \le \gamma\rho_k^2, & \rho_{k+1} \le \gamma\rho_k^2, & \left\|U_k^T A(\mathbf{c}^k) V_k - \Sigma_*\right\| \le \gamma\rho_k^2, \\ \|U_{k+1} - U_k\| \le 2\gamma\rho_k^2, & & \|V_{k+1} - V_k\| \le 2\gamma\rho_k^2. \end{array} \tag{68}$$

where $\rho_k \equiv \sqrt{\|H_k\|^2 + \|K_k\|^2}$.

Let $\rho = \gamma\rho_1$, by (67) and (66), we know that $\rho < 1$. From (68) we have for each $k \ge 2$,

$$\begin{aligned} \left\|\mathbf{c}^k - \mathbf{c}^{k-1}\right\| &\le \gamma\rho_{k-1}^2 \le \gamma\left(\gamma\rho_{k-2}^2\right)^2 = \gamma^{1+2}\rho_{k-2}^{2^2} \\ &\le \cdots \le \gamma^{1+2+2^2+\cdots+2^{k-2}}\rho_1^{2^{k-1}} \\ &\le \left(\gamma^{\frac{1+2+2^2+\cdots+2^{k-2}}{2^{k-1}}}\rho_1\right)^{2^{k-1}} \\ &\le (\gamma\rho_1)^{2^{k-1}} \le \rho^{2^{k-1}}. \end{aligned}$$

Thus for any integer $m \ge 1$,

$$\begin{aligned} \left\|\mathbf{c}^{k+m} - \mathbf{c}^k\right\| &\le \sum_{l=1}^m \left\|\mathbf{c}^{k+l} - \mathbf{c}^{k+l-1}\right\| \le \sum_{l=1}^m \rho^{2^{k+l-1}} = \sum_{l=1}^m \left(\rho^{2^{k-1}}\right)^{2^l} \\ &\le \sum_{l=1}^m \left(\rho^{2^{k-1}}\right)^l = \frac{\rho^{2^{k-1}} - \left(\rho^{2^{k-1}}\right)^{m+1}}{1 - \rho^{2^{k-1}}}. \end{aligned} \tag{69}$$

This shows that $\{\mathbf{c}^k\}$ is a Cauchy sequence since $\rho < 1$. Therefore, there exists a $\tilde{\mathbf{c}} \in \mathbb{R}^n$ such that $\{\mathbf{c}^k\}$ converge to $\tilde{\mathbf{c}}$.

Similarly, from (68) we have, for any integer $m > 1$,

$$\max\{\|U_{k+m} - U_k\|, \|V_{k+m} - V_k\|\} \le 2\frac{\rho^{2^{k-1}} - \left(\rho^{2^{k-1}}\right)^{m+1}}{1 - \rho^{2^{k-1}}}. \tag{70}$$

It shows that $\{U_k\}$ and $\{V_k\}$ are both Cauchy sequences. Thus there exist two matrices $\tilde{U} \in \mathcal{O}(m)$ and $\tilde{V} \in \mathcal{O}(n)$ such that $\{U_k\}$ and $\{V_k\}$ converge to $\tilde{U}$ and $\tilde{V}$, respectively. Finally, from (68), we have that $\left\{U_k^T A(\mathbf{c}^k) V_k\right\}$ converge to $\tilde{U}^T A(\tilde{\mathbf{c}})\tilde{V} = \Sigma_*$. □

Remark. *It is worthwhile to point out that* $\tilde{\mathbf{c}}$ *may not equal to the solution* $\mathbf{c}^*$. *We can observe this from the numerical tests in §4.*

We end this section by establishing quadratic convergence of our method in the root sense.

Theorem 3 *Under the same conditions as in Theorem 2, the three sequences of iterates* $\{\mathbf{c}^k\}$*,* $\{U_k\}$ *and* $\{V_k\}$ *generated by the iterative algorithm are all locally convergent with root-convergence rate at least equal to 2.*

Proof: By Theorem 2, we know that $\{\mathbf{c}^k\}$ is locally convergent with

$$\lim_{k\to\infty} \mathbf{c}^k = \tilde{\mathbf{c}}.$$

Since $\rho < 1$, let $m \to \infty$ in (69), we have, for each $k \geq 1$,

$$\left\|\tilde{\mathbf{c}} - \mathbf{c}^k\right\| \leq \frac{\rho^{2^{k-1}}}{1-\rho^{2^{k-1}}} \leq \xi\rho^{2^{k-1}},$$

where $\xi = \frac{1}{1-\rho} > 1$.

Next, We estimate the root-convergence factors of $\{\mathbf{c}^k\}$ defined in (56) for different values of p:

1. If $p = 1$, then

$$R_1\{\mathbf{c}^k\} = \limsup_{k\to\infty} \left\|\mathbf{c}^k - \tilde{\mathbf{c}}\right\|^{1/k} \leq \limsup_{k\to\infty} \xi^{1/k}\left(\rho^{1/2}\right)^{2^k/k} = 0.$$

2. If $1 < p < 2$, then

$$R_p\{\mathbf{c}^k\} = \limsup_{k\to\infty} \left\|\mathbf{c}^k - \tilde{\mathbf{c}}\right\|^{1/p^k} \leq \limsup_{k\to\infty} \xi^{1/p^k}\left(\rho^{1/2}\right)^{(2/p)^k} = 0.$$

3. If $p = 2$, then

$$R_p\{\mathbf{c}^k\} = \limsup_{k\to\infty} \left\|\mathbf{c}^k - \tilde{\mathbf{c}}\right\|^{1/2^k} \leq \limsup_{k\to\infty} \xi^{1/p^k}\rho^{1/2} = \rho^{1/2} < 1.$$

4. If $p > 2$, then

$$R_p\{\mathbf{c}^k\} = \limsup_{k\to\infty} \left\|\mathbf{c}^k - \tilde{\mathbf{c}}\right\|^{1/p^k} \leq \limsup_{k\to\infty} \xi^{1/p^k}\left(\rho^{1/2}\right)^{(2/p)^k} = 1.$$

Therefore, $R_p\{\mathbf{c}^k\} = 0$ for any $p \in [1, 2)$ and $R_p\{\mathbf{c}^k\} \leq 1$ for any $p \in [2, \infty)$. Thus according to (57), $O_R(\mathbf{c}^*) \geq 2$.

By similar arguments, we can prove that $\{U_k\}$ and $\{V_k\}$ converge to their limits $\tilde{U}$ and $\tilde{V}$ quadratically in the root sense. □

4 Numerical Experiments

In this section, we report some numerical experiments to show the performance of the iterative algorithm. For demonstration purpose, we consider the case when $m = 7$ and $n = 4$. The tests were performed using `Matlab` 6.1 with machine precision 2.2×10^{-16}. All the basis matrices were generated randomly by `Matlab` from a normal distribution with mean 0.0 and variance 1.0. To make sure that the ISVP under testing does have a solution, we first randomly generate a vector $\mathbf{c}^* \in \mathbb{R}^4$. Then singular values of the corresponding matrix $A(\mathbf{c}^*)$ are used as the prescribed singular values. We perturb each entry of the vector $\mathbf{c}^*$ by a uniform distribution between -1 and 1 and use the perturbed vector as the initial guess $\mathbf{c}^0$ for the iteration. In our experiments, the iterations are stopped when

$$\|U_k^T A(\mathbf{c}^k)V_k - \Sigma_*\|_F \leq 10^{-13}.$$

Table 1 includes $\mathbf{c}^*$, the initial guess $\mathbf{c}^0$ and the corresponding limit point $\tilde{\mathbf{c}}$ for three cases. Table 2 lists the errors between $\mathbf{c}^0, \mathbf{c}^*$ and $\tilde{\mathbf{c}}$. The number of performed iterations is 9, 5, 10 respectively. We can see from Table 2 that for Case (a) and Case (c), the limit point $\tilde{\mathbf{c}}$ of the iteration is not equal to the original vector $\mathbf{c}^*$ to which $\mathbf{c}^0$ is reasonably close. In particular, in Case (c), $\mathbf{c}^0$ is nearer to $\tilde{\mathbf{c}}$ than to $\mathbf{c}^*$ while Case (a) is the opposite. We point out that this occurrence is in accordance with the proved convergence results and with the convergence features of iterative processes based on Newton method.

	Case (a)	Case (b)	Case (c)
c_1^*	$2.4467e+00$	$1.0987e+00$	$2.1995e+00$
c_2^*	$9.9836e-01$	$4.5028e-01$	$7.2577e-01$
c_3^*	$1.4491e+00$	$1.0739e+00$	$-7.0029e-01$
c_4^*	$-1.0565e+00$	$-7.5681e-01$	$2.0901e+00$
c_1^0	$2.6943e+00$	$1.6498e+00$	$2.4376e+00$
c_2^0	$1.4342e+00$	$7.6865e-01$	$7.6459e-01$
c_3^0	$2.4266e+00$	$1.6628e+00$	$2.3170e-01$
c_4^0	$-1.9655e-01$	$-5.1570e-01$	$2.5964e+00$
$\tilde{c}_1$	$3.1675e+00$	$1.0987e+00$	$1.9495e+00$
$\tilde{c}_2$	$1.5874e+00$	$4.5028e-01$	$1.0043e+00$
$\tilde{c}_3$	$2.8446e-01$	$1.0739e+00$	$-2.7139e-01$
$\tilde{c}_4$	$-3.5270e-01$	$-7.5681e-01$	$2.2095e+00$

Table 1: Initial and final values of $\mathbf{c}^k$.

	$\|\mathbf{c}^0 - \mathbf{c}^*\|$	$\|\mathbf{c}^0 - \tilde{\mathbf{c}}\|$	$\|\mathbf{c}^* - \tilde{\mathbf{c}}\|$
Case (a)	$1.3951e+00$	$2.2047e+00$	$1.6487e+00$
Case (b)	$8.9994e-01$	$8.9994e-01$	$6.1294e-15$
Case (c)	$1.0877e+00$	$8.3577e-01$	$5.8163e-01$

Table 2: Errors between $\mathbf{c}^0, \mathbf{c}^*$ and $\tilde{\mathbf{c}}$.

Clearly, the singular values of $A(\tilde{\mathbf{c}})$ do agree with those of $A(\mathbf{c}^*)$. Table 3 indicates how the singular values of $A(\mathbf{c}^0)$ differ from those of $A(\tilde{\mathbf{c}})$.

In order to further illustrate our theoretical results, in Table 4, we give the convergence history of the three sequences $\{\mathbf{c}^k\}$, $\{U_k\}$ and $\{V_k\}$ for Case (a). Here, the limits $\tilde{\mathbf{c}}$, $\widetilde{U}$ and $\widetilde{V}$ are computed up to full precision. From Table 4, we can observe that the three sequences converge fast.

Finally, Table 5 displays the distance between $\boldsymbol{\sigma}(\mathbf{c}^k) = (\sigma_1(\mathbf{c}^k), \ldots, \sigma_n(\mathbf{c}^k))$ and $\boldsymbol{\sigma}^*$ measured in the 2-norm. Also, we can see that $\{\boldsymbol{\sigma}(\mathbf{c}^k)\}$ converges fast. All these numerical observation agrees with our prediction.

$\boldsymbol{\sigma}^k$	Case (a)	Case (b)	Case (c)
σ_1^0	$2.0092e+01$	$1.2835e+01$	$1.0998e+01$
σ_2^0	$1.1860e+01$	$8.0329e+00$	$1.0673e+01$
σ_3^0	$7.5045e+00$	$4.8781e+00$	$7.3750e+00$
σ_4^0	$3.3831e+00$	$2.7887e+00$	$3.9632e+00$
σ_1^*	$1.5364e+01$	$8.6619e+00$	$9.8737e+00$
σ_2^*	$1.0882e+01$	$6.0069e+00$	$7.8588e+00$
σ_3^*	$5.8869e+00$	$3.5380e+00$	$6.7673e+00$
σ_4^*	$2.6861e+00$	$2.1041e+00$	$3.3554e+00$

Table 3: Singular values of $A(\mathbf{c}^k)$.

Iterations	$\|\mathbf{c}^k-\tilde{\mathbf{c}}\|$	$\|U_k-\tilde{U}\|$	$\|V_k-\tilde{V}\|$
0	$2.2047e+00$	$2.5985e+00$	$9.5168e-01$
1	$2.3630e+00$	$2.1818e+00$	$1.3436e+00$
2	$1.7445e+00$	$1.2882e+00$	$6.8597e-01$
3	$6.9574e-01$	$8.9750e-01$	$2.8418e-01$
4	$1.5478e-01$	$1.9838e-01$	$1.4848e-01$
5	$6.0905e-02$	$7.2257e-02$	$3.7492e-02$
6	$2.7500e-03$	$3.2571e-03$	$2.0557e-03$
7	$1.3086e-05$	$1.4488e-05$	$8.9536e-06$
8	$2.2557e-10$	$2.5370e-10$	$1.5630e-10$
9	$5.3705e-15$	$5.1940e-15$	$3.7348e-15$

Table 4: Convergence history of $\{\mathbf{c}^k\}$, $\{U_k\}$ and $\{V_k\}$ for Case (a).

Acknowledgments

We would like to thank Prof. Raymond Chan for his valuable suggestions when we began this paper. We are also grateful to Prof. Moody Chu for the useful discussion and comments on this work.

References

[1] M. T. Chu, *Numerical Methods for Inverse Singular Value Problems*, SIAM J. Numer. Anal., 29 (1992), pp. 885–903.

[2] M. T. Chu, *Inverse Eigenvalue Problems*, SIAM Rev., 40 (1998), pp. 1–39.

[3] M. T. Chu, *Matrix Differential Equations: A Continuous Realization Process for Linear Algebra Problems*, Nonlinear Anal. TMA, 18 (1992), pp. 1125–1146.

[4] M. T. Chu and G. H. Golub, *Structured Inverse Eigenvalue Problems*, Acta Numerica, 11(2002), pp. 1–71.

[5] D. Fokkema, G. Sleijen, and H. van der Vorst, *Accelerated Inexact Newton Schemes for Large Systems of Nonlinear Equations*, SIAM J. Sci. Comput., 19 (1998), 657–674.

Iterations	Case (a)	Case (b)	Case (c)
0	$5.1399e+00$	$4.8769e+00$	$3.1501e+00$
1	$1.8645e+00$	$2.2626e-01$	$3.0338e-01$
2	$5.5573e+00$	$3.2821e-02$	$5.3643e-01$
3	$2.1745e+00$	$4.2234e-04$	$1.4622e-01$
4	$4.9857e-01$	$1.8717e-07$	$7.6712e-01$
5	$2.7352e-02$	$1.5397e-14$	$4.4488e-01$
6	$2.2679e-03$		$3.3527e-02$
7	$6.6781e-06$		$3.0239e-03$
8	$1.2317e-10$		$9.3265e-06$
9	$4.7830e-15$		$2.7301e-10$
10			$6.2804e-15$

Table 5: Errors of singular values.

[6] S. Friedland, J. Nocedal, and M. L. Overton, *The Formulation and Analysis of Numerical Methods for Inverse Eigenvalue Problems*, SIAM J. Numer. Anal., 24 (1987), pp. 634–667.

[7] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, 1996.

[8] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, 1970.

[9] T. Politi, *A Discrete Approach for the Inverse Singular Value Problem in Some Quadratic Group*, Lecture Notes in Computer Science, 2658 (2003), 121–130.

[10] D. F. Sun and J. Sun, *Nonsmooth Matrix Value Functions Defined by Singular Values*, Optimization Online Digest–January 2003.

[11] S. F. Xu, *An Introduction to Inverse Algebraic Eigenvalue Problems*, Peking University Press and Vieweg Publishing, 1998.