Convergence of the Newton-type methods for the square inverse singular value problems with multiple and zero singular values

Weiping Shen* Yaohua Hu† Chong Li‡ Jen-Chih Yao§

Abstract. In this paper, we study the convergence of the Newton-type methods for solving the square inverse singular value problem with possible multiple and zero singular values. Comparing with other known results, positivity assumption of the given singular values is removed. Under the nonsingularity assumption in terms of the (relative) generalized Jacobian matrices, quadratic/suplinear convergence properties (in the root-convergence sense) are proved. Moreover, numerical experiments are given in the last section to demonstrate our theoretic results.

Keywords. Inverse singular value problem, Newton's method, Newton-type method

AMS subject classification. 65F18, 65F10, 15A18

1 Introduction

The inverse singular value problem (ISVP) arises in different applications such as the determination of mass distributions, orbital mechanics, irrigation theory, computed tomography, circuit theory, etc. [13, 15, 17, 18, 19, 21, 23, 29, 30, 33]. In the present paper, we consider the following special kind of ISVP. Let p and q be two positive integers. Let $\mathbb{R}^p$ denote the p-dimensional Euclidean space and $\mathbb{R}^{p\times q}$ be the set of all real $p\times q$ matrices. Let m and n be two positive integers such that $m \geq n$. Let

*Department of Mathematics, Zhejiang Normal University, Jinhua 321004, P. R. China (shenweiping@zjnu.cn). This author is supported in part by Zhejiang Provincial Natural Science Foundation of China (grant LY17A010006)

†Corresponding author. College of Mathematics and Statistics, Shenzhen Key Laboratory of Advanced Machine Learning and Applications, Shenzhen University, Shenzhen 518060, P. R. China (mayhhu@szu.edu.cn). This author's work was supported in part by the National Natural Science Foundation of China (11601343, 11871347), Natural Science Foundation of Guangdong (2016A030310038), Natural Science Foundation of Shenzhen (JCYJ20170817100950436, JCYJ20170818091621856) and Interdisciplinary Innovation Team of Shenzhen University.

‡Department of Mathematics, Zhejiang University, Hangzhou 310027, P. R. China (cli@zju.edu.cn). This author is supported in part by the National Natural Science Foundation of China (grants 11571308, 11371325).

§Research Center for Interneural Computing, China Medical University Hospital, China Medical University, Taichung, Taiwan (yaojc@mail.cmu.edu.tw). This author is supported in part by the Grant MOST MOST 106-2923-E-039-001-MY3.

$\{A_i\}_{i=0}^n \subset \mathbb{R}^{m\times n}$. Given $\mathbf{c} = (c_1, c_2, \ldots, c_n)^T \in \mathbb{R}^n$, we define

$$A(\mathbf{c}) := A_0 + \sum_{i=1}^{n} c_i A_i \tag{1.1}$$

and denote its singular values by $\{\sigma_i(\mathbf{c})\}_{i=1}^n$ with the order $\sigma_1(\mathbf{c}) \geq \sigma_2(\mathbf{c}) \geq \cdots \geq \sigma_n(\mathbf{c}) \geq 0$. The ISVP considered here is, for n given real numbers $\{\sigma_i^*\}_{i=1}^n$ ordering with

$$\sigma_1^* \geq \sigma_2^* \geq \cdots \geq \sigma_n^* \geq 0,$$

to find a vector $\mathbf{c}^* \in \mathbb{R}^n$ such that $\{\sigma_i^*\}_{i=1}^n$ are exactly the singular values of $A(\mathbf{c}^*)$, i.e.,

$$\sigma_i(\mathbf{c}^*) = \sigma_i^*, \quad \text{for each } i = 1, 2, \ldots, n. \tag{1.2}$$

The vector $\mathbf{c}^*$ is called a solution of the ISVP (1.2). This type of ISVP was originally proposed by Chu [5] in 1992 and was further studied in [2, 3, 5, 16, 27, 31]. In the case when $m = n$, we call the problem is square. Obviously, if $\{A_i\}_{i=0}^n$ are symmetric, the square inverse eigenvalue problem is reduced to the inverse eigenvalue problem (IEP) which arises in a variety of applications and was studied extensively in [1, 4, 6, 7, 11, 25, 32].

Even though the solvability issue for the ISVP (1.2) is very complicated, some numerical algorithms for solving (1.2) (and so the square inverse singular value problems) have still been developed [2, 3, 5, 16, 27, 31]. In general, these numerical methods can be distinguished into two classes. One is the continuous method which consists of solving an ordinary differential obtained from an explicit calculation of the projected gradient of a certain objective function (cf.[5]). The other kind of method that we are interested in below is the iterative methods. Define the function $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\mathbf{f}(\mathbf{c}) := (\sigma_1(\mathbf{c}) - \sigma_1^*,\ \sigma_2(\mathbf{c}) - \sigma_2^*,\ \ldots,\ \sigma_n(\mathbf{c}) - \sigma_n^*)^T, \quad \text{for any } \mathbf{c} \in \mathbb{R}^n. \tag{1.3}$$

Then, as noted in [2, 3, 5, 16, 27, 31], solving the ISVP (1.2) is equivalent to finding a solution $\mathbf{c}^* \in \mathbb{R}^n$ of the nonlinear equation $\mathbf{f}(\mathbf{c}) = \mathbf{0}$. Note that, in the case when the given singular values are distinct and positive, i.e.,

$$\sigma_1^* > \sigma_2^* > \cdots > \sigma_n^* > 0, \tag{1.4}$$

the singular vectors $\{\mathbf{u}_i(\mathbf{c})\}_{i=1}^n$ and $\{\mathbf{v}_i(\mathbf{c})\}_{i=1}^n$ are continuous with respect to $\mathbf{c}$ around $\mathbf{c}^*$. Moreover, there exists a neighborhood of $\mathbf{c}^*$ where the function $\mathbf{f}$ is differentiable around the solution $\mathbf{c}^*$ (cf. [3]). Thus, in this case, one can certainly apply Newton's method for solving the nonlinear equation $\mathbf{f}(\mathbf{c}) = \mathbf{0}$ to produce Newton's method for solving ISVP (1.2). However, Newton's method for the ISVP (1.2) requires solving a complete singular value problem for the matrix $A(\mathbf{c})$ at each outer iteration. This sometimes makes it inefficient from the viewpoint of practical calculations especially when the problem size is large. Therefore, Chu designed in [5] a Newton-type method for solving the ISVP (1.2) which requires computing approximate singular vectors instead of singular vectors at each iteration. Under the assumption that the given singular values $\{\sigma_i^*\}_{i=1}^n$ are distinct and positive, the Newton-type method was proved in [2] to be quadratically convergent (in the root-convergence sense). To alleviate the over-solving problem, Bai *et al* designed in [3] an inexact version of the Newton-type method for the distinct and positive case where the approximate Jacobian equation was solved inexactly by adopting a suitable stopping criteria. Also under the assumption (1.4), a convergence analysis for the inexact Newton-type method was presented in that paper and the superlinear convergence was proved. On the other hand, motivated by the Ulm-like Cayley transform method introduced in [26] for solving the IEP, Vong, Bai, and Jin presented in [31] a Ulm-like method for the ISVP. Again under the assumption (1.4), they showed that the proposed method converged at least quadratically.

As mentioned above, there are some works on the formation and convergence analysis of numerical methods for solving the ISVP (1.2). However, to our knowledge, distinct and positive assumption is always assumed for the given singular values among those works[1]. In the case when multiple and/or zero singular values are present, that is, without loss of generality, $\{\sigma_i^*\}_{i=1}^n$ satisfies that

$$\sigma_1^* = \cdots = \sigma_s^* > \sigma_{s+1}^* \cdots > \sigma_{n-t}^* > \sigma_{n-t+1}^* = \ldots = \sigma_n^* = 0, \tag{1.5}$$

solving the ISVP (1.2) becomes harder and the techniques for the distinct and positive case could no longer work as the function $\mathbf{f}$ may be not analytic around $\mathbf{c}^*$ and the singular vectors may be not continuous around $\mathbf{c}^*$. As mentioned in [5], the appearance of zero singular values is especially challenging since zero singular values indicate rank deficiency and, to find a lower rank matrix in the generic affine subspace:

$$\mathcal{A}(\mathbf{c}) := \{A(\mathbf{c}) | \mathbf{c} \in \mathbb{R}^n\}$$

is intuitively a quite difficult problem.

The purpose of this paper is trying to studying the numerical solutions of the square inverse singular value problem (i.e., the ISVP (1.2) in the case when $m = n$) with multiple and/or zero singular values given by (1.5). By modifying the Newton-type method in [2], a (exact) Newton-type method for solving the square inverse singular value problem with assumption (1.5) is proposed, and the convergence issue of the proposed Newton-type method is studied. Under the nonsingularity assumption on the (relative) generalized Jacobian matrices at the solution $\mathbf{c}^*$ (which is used by Shen *et al* in [26]), we show that the sequence $\{\mathbf{c}^k\}$ generated by the proposed method converges quadratically to $\mathbf{c}^*$ even when multiple and/or zero singular values are presented. Moreover, to alleviate the over-solving problem, an inexact version of the proposed method is also designed here where the approximate Jacobian equation is solved inexactly by adopting a suitable stopping criteria. In particular, the main results of this paper improve/extend partially the corresponding ones of [2] and [3] in the case when $m = n$. Finally, some numerical experiments are presented to illustrate the theoretical results in the last section.

2 Preliminaries

Let $\mathbf{B}(\mathbf{x}, \delta)$ be the open ball in $\mathbb{R}^p$ with center $\mathbf{x} \in \mathbb{R}^p$ and radius $\delta > 0$. Let $\mathcal{S}(p)$ and $\mathcal{O}(p)$ denote respectively the sets of symmetric and orthogonal matrices in $\mathbb{R}^{p\times p}$. Let $\mathcal{D}(n)$ be the set of diagonal matrices in $\mathbb{R}^{p\times p}$ with increasing diagonal entries. Let I denote an identity matrix. Let $\|\cdot\|$ be the Euclidean vector norm or its induced matrix norm, and let $\|\cdot\|_F$ denote the Frobenius norm. Then,

$$\|A\| \le \|A\|_F \le \sqrt{q}\,\|A\|, \quad \text{for each } A \in \mathbb{R}^{p\times q}. \tag{2.1}$$

The symbol $\mathrm{Diag}(a_1, \ldots, a_n)$ denotes a diagonal matrix with $a_1, \ldots, a_n$ being its diagonal elements and $\mathrm{diag}(M) := (m_{11}, \ldots, m_{nn})^T$ denotes a vector containing the diagonal elements of an $n \times n$

[1] An exact/inexact Newton-type method was proposed in [27] for solving the ISVP (1.2) with the multiple but positive singular values, and it's quadratic/superlinear convergence was claimed there. However, as explained in the conclusion section, there is a fatal gap in the proof for the main theorem there (i.e.,[27, Theorem 3.1].

matrix $M := (m_{ij})$. Let $\{\sigma_i^*\}_{i=1}^n$ be the given singular values satisfying (1.5). Write

$$\boldsymbol{\sigma}^* := (\sigma_1^*, \ \ldots, \ \sigma_n^*)^T \quad \text{and} \quad \Sigma^* := \mathrm{Diag}(\sigma_1^*, \ \ldots, \ \sigma_n^*) \in \mathbb{R}^{n\times n}. \tag{2.2}$$

Let $\mathbf{c} \in \mathbb{R}^n$ and $A(\mathbf{c})$ be defined by (1.1). Let $\{\sigma_i(\mathbf{c})\}_{i=1}^n$ stand for the singular values of $A(\mathbf{c})$ with the order $\sigma_1(\mathbf{c}) \geq \sigma_2(\mathbf{c}) \geq \cdots \geq \sigma_n(\mathbf{c}) \geq 0$. Write

$$\Sigma(\mathbf{c}) := \mathrm{Diag}(\sigma_1(\mathbf{c}), \ \ldots, \ \sigma_n(\mathbf{c})) \in \mathbb{R}^{n\times n}.$$

Define

$$\mathcal{W}(\mathbf{c}) := \{[U(\mathbf{c}), \ V(\mathbf{c})] \mid U(\mathbf{c})^T A(\mathbf{c}) V(\mathbf{c}) = \Sigma(\mathbf{c}), \ U(\mathbf{c}) \text{ and } V(\mathbf{c}) \in \mathcal{O}(n)\}.$$

As in [31], we ignore the choice of possible sign for $[U(\mathbf{c}), \ V(\mathbf{c})]$. For each $[U(\mathbf{c}), \ V(\mathbf{c})] \in \mathcal{W}(\mathbf{c})$, we write

$$U(\mathbf{c}) := [U^{(1)}(\mathbf{c}), \ U^{(2)}(\mathbf{c}), \ U^{(3)}(\mathbf{c})] \quad \text{and} \quad V(\mathbf{c}) := [V^{(1)}(\mathbf{c}), \ V^{(2)}(\mathbf{c}), \ V^{(3)}(\mathbf{c})]$$

where $U^{(1)}(\mathbf{c}), \ V^{(1)}(\mathbf{c}) \in \mathbb{R}^{n\times s}$, and $U^{(3)}(\mathbf{c}), \ V^{(3)}(\mathbf{c}) \in \mathbb{R}^{n\times t}$. Throughout this paper, we suppose that $\mathbf{c}^*$ is a solution of the square inverse singular value problem. For $i = 1$ and $i = 3$, define

$$\Pi_{U,i} = U^{(i)}(\mathbf{c}^*) U^{(i)}(\mathbf{c}^*)^T \quad \text{and} \quad \Pi_{V,i} = V^{(i)}(\mathbf{c}^*) V^{(i)}(\mathbf{c}^*)^T. \tag{2.3}$$

We first present some auxiliary lemmas. In particular, Lemma 2.1 gives a perturbation bound for the inverse which is known in [12, pp.58–59]; Lemma 2.2 is a direct consequence of the Cholesky factorization (cf. [11, Lemma 3.1]); while Lemmas 2.3 and 2.4 have been presented respectively in [2, Lemma 2] and [25, Lemma 4.1].

Lemma 2.1. *Let $A, \ B \in \mathbb{R}^{p\times p}$. Assume that B is nonsingular and $\|B^{-1}\| \cdot \|A - B\| < 1$. Then A is nonsingular and moreover*

$$\|A^{-1}\| \leq \frac{\|B^{-1}\|}{1 - \|B^{-1}\| \cdot \|A - B\|}.$$

Lemma 2.2. *Let $M \in \mathbb{R}^{p\times q}$ where $p \geq q$. Let $W = (w_{ij})$ be a $q \times q$ nonsingular upper triangle matrix such that $w_{11} > 0$ and $W^T W = I - M^T M$. Then there exist two numbers $\epsilon \in (0, \ 1)$ and $\alpha \in (0, +\infty)$ such that the following implication holds:*

$$\|M\| \leq \epsilon \Longrightarrow \|I - W\| \leq \alpha \|M\|^2.$$

Lemma 2.3. *There exists a constant $\alpha \in (0, +\infty)$ such that for any $\mathbf{c}, \ \bar{\mathbf{c}} \in \mathbb{R}^n$,*

$$\|A(\mathbf{c}) - A(\bar{\mathbf{c}})\| \leq \alpha \|\mathbf{c} - \bar{\mathbf{c}}\|.$$

Lemma 2.4. *Suppose that $\hat{A} \in \mathcal{S}(n)$. Then there exist positive constants β and κ such that*

$$\min_{\hat{Q}\in\mathcal{O}(n), \ \hat{Q}^T \hat{A} \hat{Q} \in \mathcal{D}(n)} \|Q - \hat{Q}\| \leq \beta \|A - \hat{A}\|, \ \text{whenever } A \in \mathcal{S}(n), \ Q \in \mathcal{O}(n), \ Q^T A Q \in \mathcal{D}(n), \ \|A - \hat{A}\| \leq \kappa.$$

The following lemma given in [24, Lemma 2.5] is also needed.

Lemma 2.5. *Let $Z \in \mathbb{R}^{n\times n}$. Suppose that the skew-symmetric matrices $H, \ K \in \mathbb{R}^{n\times n}$ satisfy*

$$H\Sigma^* - \Sigma^* K = Z.$$

Then we have

$$[H]_{ij} = \frac{[Z]_{ij}}{\sigma_j^*}, \quad n-t+1 \le i \le n, \quad 1 \le j \le n-t,$$

$$[K]_{ij} = -\frac{[Z]_{ij}}{\sigma_i^*}, \quad 1 \le i \le n-t, \quad n-t+1 \le j \le n,$$

$$[H]_{ij} = \frac{1}{(\sigma_j^*)^2 - (\sigma_i^*)^2}\left(\sigma_j^*[Z]_{ij} + \sigma_i^*[Z]_{ji}\right), \quad s+1 \le i \le n-t, \quad 1 \le j \le n-t, \quad i > j,$$

$$[K]_{ij} = \frac{1}{(\sigma_j^*)^2 - (\sigma_i^*)^2}\left(\sigma_j^*[Z]_{ji} + \sigma_i^*[Z]_{ij}\right), \quad s+1 \le i \le n-t, \quad 1 \le j \le n-t, \quad i > j.$$

The following two lemmas can also been found in [24, Lemma 2.6] and [24, Lemma 2.7]. Recall that $\Pi_{U,i}$ and $\Pi_{V,i}$ are defined by (2.3) for $i = 1,\ 3$. Let $U := [U^{(1)},\ U^{(2)},\ U^{(3)}]$ and $V := [V^{(1)},\ V^{(2)},\ V^{(3)}] \in \mathcal{O}(n)$ where $U^{(1)},\ V^{(1)} \in \mathbb{R}^{n\times s}$ and $U^{(3)},\ U^{(3)} \in \mathbb{R}^{n\times t}$. We form the QR factorization of $\Pi_{U,i}U^{(i)}$ and $\Pi_{V,i}V^{(i)}$ for $i = 1$ and $i = 3$:

$$\Pi_{U,i}U^{(i)} = \tilde{U}^{(i)}(\mathbf{c}^*)R_U^{(i)} \quad \text{and} \quad \Pi_{V,i}V^{(i)} = \tilde{V}^{(i)}(\mathbf{c}^*)R_V^{(i)},$$

where $R_U^{(i)}$, $R_V^{(i)}$ are nonsingular upper triangular matrices, and $\tilde{U}^{(i)}(\mathbf{c}^*)$, $\tilde{V}^{(i)}(\mathbf{c}^*)$ are matrices whose columns are orthonormal. Let

$$\tilde{U}(\mathbf{c}^*) := [\tilde{U}^{(1)}(\mathbf{c}^*),\ U^{(2)}(\mathbf{c}^*),\ \tilde{U}^{(3)}(\mathbf{c}^*)] \quad \text{and} \quad \tilde{V}(\mathbf{c}^*) := [\tilde{V}^{(1)}(\mathbf{c}^*),\ V^{(2)}(\mathbf{c}^*),\ \tilde{V}^{(3)}(\mathbf{c}^*)].$$

Clearly, $[\tilde{U}(\mathbf{c}^*)\ \tilde{V}(\mathbf{c}^*)] \in \mathcal{W}(\mathbf{c}^*)$. Suppose that the skew-symmetric matrices $\tilde{X}$, $\tilde{Y} \in \mathbb{R}^{n\times n}$ satisfy

$$e^{\tilde{X}} = U^T\tilde{U}(\mathbf{c}^*) \quad \text{and} \quad e^{\tilde{Y}} = V^T\tilde{V}(\mathbf{c}^*). \tag{2.4}$$

Define the error matrices E_U and E_V:

$$E_U := [E_U^{(1)},\ E_U^{(2)},\ E_U^{(3)}] \quad \text{and} \quad E_V := [E_V^{(1)},\ E_V^{(2)},\ E_V^{(3)}], \tag{2.5}$$

where

$$E_U^{(i)} := (I - \Pi_{U,i})U^{(i)} \quad \text{and} \quad E_V^{(i)} := (I - \Pi_{V,i})V^{(i)}, \quad i = 1,\ 3,$$

and

$$E_U^{(2)} := U^{(2)} - U^{(2)}(\mathbf{c}^*) \quad \text{and} \quad E_V^{(2)} := V^{(2)} - V^{(2)}(\mathbf{c}^*). \tag{2.6}$$

Finally, for any matrix $M \in \mathbb{R}^{n\times n}$, we use $M^{[s]_U}$ and $M^{[t]_L}$ to denote respectively the $s\times s$ upper left and $t\times t$ lower right blocks of the matrix M.

Lemma 2.6. *There exist two numbers $\delta \in (0,\ 1)$ and $\gamma \in [1, +\infty)$ such that for any $\mathbf{c} \in \mathbf{B}(\mathbf{c}^*,\ \delta)$ and $[U(\mathbf{c}),\ V(\mathbf{c})] \in \mathcal{W}(\mathbf{c})$, the following assertions hold:*

(i) $\|U^{(2)}(\mathbf{c}) - U^{(2)}(\mathbf{c}^*)\| \le \gamma\|\mathbf{c} - \mathbf{c}^*\|$ *and* $\|(I - \Pi_{U,i})U^{(i)}(\mathbf{c})\| \le \gamma\|\mathbf{c} - \mathbf{c}^*\|$ *for* $i = 1,\ 3$*;*

(ii) $\|V^{(2)}(\mathbf{c}) - V^{(2)}(\mathbf{c}^*)\| \le \gamma\|\mathbf{c} - \mathbf{c}^*\|$ *and* $\|(I - \Pi_{V,i})V^{(i)}(\mathbf{c})\| \le \gamma\|\mathbf{c} - \mathbf{c}^*\|$ *for* $i = 1,\ 3$.

Lemma 2.7. *There exist two numbers $\delta \in (0,\ 1)$ and $\gamma \in [1, +\infty)$ such that*

(i) *for any matrix $U \in \mathcal{O}(n)$ with $\|E_U\| < \delta$, the skew-symmetric matrix $\tilde{X}$ defined by (2.4) satisfies*

$$\|\tilde{X}\|_F \le \gamma\|E_U\|, \quad \|\tilde{X}^{[s]_U}\|_F \le \gamma\|E_U\|^2 \quad \textit{and} \quad \|\tilde{X}^{[t]_L}\|_F \le \gamma\|E_U\|^2$$

(ii) *for any matrix* $V \in \mathcal{O}(n)$ *with* $\|E_V\| < \delta$, *the skew-symmetric matrix* $\tilde{Y}$ *defined by* (2.4) *satisfies*

$$\|\tilde{Y}\|_F \leq \gamma\|E_V\|, \quad \|\tilde{Y}^{[s]_U}\|_F \leq \gamma\|E_V\|^2 \quad \text{and} \quad \|\tilde{Y}^{[t]_L}\|_F \leq \gamma\|E_V\|^2.$$

Now we present the definitions and some properties of the B-differential Jacobian, the generalized Jacobian and the relative generalized Jacobian. For this, let $\mathbf{g} : \mathbb{R}^p \to \mathbb{R}^q$ be a locally Lipschitz continuous function. Let $\mathbf{g}'$ be the Fréchet derivative of $\mathbf{g}$ whenever it exists and $D_{\mathbf{g}}$ be the set of differentiable points of $\mathbf{g}$. Recall from [8, 22] that the B-differential Jacobian of $\mathbf{g}$ at $\mathbf{x} \in \mathbb{R}^p$ is defined by

$$\partial_B \mathbf{g}(\mathbf{x}) := \{J \in \mathbb{R}^{q\times p} \mid J = \lim_{\mathbf{x}_k \to \mathbf{x}} \mathbf{g}'(\mathbf{x}_k), \ \mathbf{x}_k \in D_{\mathbf{g}}\}.$$

Consider the composite nonsmooth function:

$$\mathbf{g} := \varphi \circ \psi, \tag{2.7}$$

where $\varphi : \mathbb{R}^l \to \mathbb{R}^q$ is nonsmooth but of special structure and $\psi : \mathbb{R}^p \to \mathbb{R}^l$ is continuously differentiable. Let $\mathcal{S}$ be a subset of $\mathbb{R}^n$ and cl$S$ denote the closure of $\mathcal{S}$. The generalized Jacobian $\partial_Q \mathbf{g}(\cdot)$ and relative generalized Jacobian $\partial_{Q|S}\mathbf{g}(\cdot)$ at $\mathbf{x} \in \mathbb{R}^n$, which were introduced respectively in [20] and [28], are defined as follows:

$$\partial_Q \mathbf{g}(\mathbf{x}) := \partial_B(\varphi(\psi(\mathbf{x})))\psi'(\mathbf{x});$$

$$\partial_{Q|\mathcal{S}}\mathbf{g}(\mathbf{x}) := \{J \mid J \text{ is a limit of } G_k \in \partial_Q \mathbf{g}(\mathbf{y}_k), \ \mathbf{y}_k \in \mathcal{S}, \ \mathbf{y}_k \to \mathbf{x}\}.$$

The following lemma is known in [27, Proposition 2.1].

Lemma 2.8. *Let* $\bar{\mathbf{x}} \in \mathbb{R}^p$ *and let* $\mathcal{S}$ *be a subset of* $\mathbb{R}^p$. *Let* g *be defined by* (2.7). *Then* $\partial_B \mathbf{g}(\bar{\mathbf{x}})$ *and* $\partial_Q \mathbf{g}(\bar{\mathbf{x}})$ *are nonempty and compact, and so is* $\partial_{Q|\mathcal{S}}\mathbf{g}(\bar{\mathbf{x}})$ *if* $\bar{\mathbf{x}} \in \text{cl}\mathcal{S}$.

In the remainder of the present paper, let

$$\mathcal{S} := \{\mathbf{c} \in \mathbb{R}^n \mid A(\mathbf{c}) \text{ has positive and distinct singular values}\}.$$

For any matrix $M \in \mathbb{R}^{n\times n}$, we use $\{\sigma_i(M)\}_{i=1}^n$ to denote the singular values of M with $\sigma_1(M) \geq \ldots \geq \sigma_n(M) \geq 0$. Define the operator $\boldsymbol{\sigma} : \mathbb{R}^{n\times n} \longrightarrow \mathbb{R}^n$ by

$$\boldsymbol{\sigma}(M) := (\sigma_1(M), \ldots, \sigma_n(M))^T, \quad \text{for any} \quad M \in \mathbb{R}^{n\times n}. \tag{2.8}$$

Recall that the operators A and $\mathbf{f}$ are defined by (1.1) and (1.3) respectively. Then

$$\mathbf{f} = \boldsymbol{\sigma} \circ A - \boldsymbol{\sigma}^*.$$

Thus we have the following two lemmas. Lemma 2.9, which has been proved in [27], gives the B-differential Jacobian, generalized Jacobian, and relative generalized Jacobian of $\mathbf{f}$ at $\mathbf{c}$. While Lemma 2.10, which is a direct consequence of [27, Lemma 2.1], gives a perturbation bound for the inverses of B-differential Jacobian, the generalized Jacobian and the relative generalized Jacobian.

Lemma 2.9. *Let* $\mathbf{f}$ *be defined by* (1.3). *Then we have the following assertions:*

(i) *If* $\mathbf{c} \in \mathbb{R}^n$ *such that* $\sigma_n(\mathbf{c}) > 0$, *then* $\partial_Q \mathbf{f}(\mathbf{c}) = \{J \mid [J]_{ij} = \mathbf{u}_i(\mathbf{c})^T A_j \mathbf{v}_i(\mathbf{c}), \ [U(\mathbf{c}), \ V(\mathbf{c})] \in \mathcal{W}(\mathbf{c})\}$.

(ii) *If* $\mathbf{c} \in \mathcal{S}$, *then* $\mathbf{f}$ *is continuously differentiable at* $\mathbf{c}$ *and moreover* $\partial_B \mathbf{f}(\mathbf{c}) = \partial_Q \mathbf{f}(\mathbf{c}) = \{\mathbf{f}'(\mathbf{c})\}$;

(iii) *If* $\mathbf{c} \in \mathrm{cl}\mathcal{S}$*, then* $\partial_{Q|S}\mathbf{f}(\mathbf{c}) = \{J \mid J = \lim_{k\to+\infty} f'(\mathbf{y}^k)$ *with* $\{\mathbf{y}^k\} \subset S$ *and* $\mathbf{y}^k \to \mathbf{c}\}$.

Lemma 2.10. *Let* $\mathbf{c}^* \in \mathbb{R}^n$ *such that the matrix* $A(\mathbf{c}^*)$ *has singular values given by* (1.5)*. Suppose that each* $J \in \partial_{Q|S}\mathbf{f}(\mathbf{c}^*)$ *(resp. each* $J \in \partial_B\mathbf{f}(\mathbf{c}^*)$*, each* $J \in \partial_Q\mathbf{f}(\mathbf{c}^*)$*) is nonsingular. Then there exist two numbers* $\delta \in (0,\ 1)$ *and* $\gamma \in [1,\ +\infty)$ *such that for any* $\mathbf{c} \in \mathbf{B}(\mathbf{c}^*, \delta)$*,*

$$\sup_{J\in\partial_{Q|S}\mathbf{f}(\mathbf{c})} \|J^{-1}\| \Big(\textit{resp.} \sup_{J\in\partial_B\mathbf{f}(\mathbf{c})} \|J^{-1}\|,\ \sup_{J\in\partial_Q\mathbf{f}(\mathbf{c})} \|J^{-1}\|\Big) \le \gamma, \tag{2.9}$$

where we adopt the convention that $\sup\varnothing = -\infty$.

3 The Newton-type method and convergence analysis

In this section, we begin with the (exact) Newton-type method for solving the square inverse singular value problems with the singular values given by (1.5). For the original idea of the Newton-type method, one may refer to [2, 5]. Clearly, in the case when $t = 0$, the method presented below is reduced to the Newton-type method proposed in [27] (with $m = n$) for the multiple but positive case.

Algorithm 1. the Newton-type method

1. Given $\mathbf{c}^0 \in \mathbb{R}^n$, compute the singular values $\{\sigma_i(\mathbf{c}^0)\}_{i=1}^n$, the orthonormal left singular vectors $\{\mathbf{u}_i(\mathbf{c}^0)\}_{i=1}^m$ and right singular vectors $\{\mathbf{v}_i(\mathbf{c}^0)\}_{i=1}^n$ of $A(\mathbf{c}^0)$. Write

$$U_0 := [\mathbf{u}_1^0, \ldots, \mathbf{u}_m^0] = [\mathbf{u}_1(\mathbf{c}^0), \ldots, \mathbf{u}_m(\mathbf{c}^0)],$$

$$V_0 := [\mathbf{v}_1^0, \ldots, \mathbf{v}_n^0] = [\mathbf{v}_1(\mathbf{c}^0), \ldots, \mathbf{v}_n(\mathbf{c}^0)].$$

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

   (a) Form the approximate Jacobian matrix J_k and the vector $\mathbf{b}^k$:

$$[J_k]_{ij} := (\mathbf{u}_i^k)^T A_j \mathbf{v}_i^k, \quad 1 \le i,\ j \le n; \tag{3.1}$$

$$[\mathbf{b}^k]_i := (\mathbf{u}_i^k)^T A_0 \mathbf{v}_i^k, \quad 1 \le i \le n. \tag{3.2}$$

   (b) Compute the vector $\mathbf{c}^{k+1}$ by

$$J_k\mathbf{c}^{k+1} = \boldsymbol{\sigma}^* - \mathbf{b}^k. \tag{3.3}$$

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

(d) Calculate the skew-symmetric matrices X_k and Y_k:

$$[X_k]_{ij} := 0, \quad 1 \le i,\ j \le s \quad \text{or} \quad n-t+1 \le i,\ j \le n,$$

$$[X_k]_{ij} := -[X_k]_{ji} = \frac{[W_k]_{ij}}{\sigma_j^*}, \quad n-t+1 \le i \le n,\ 1 \le j \le n-t,$$

$$[X_k]_{ij} := -[X_k]_{ji} = \frac{\sigma_i^*[W_k]_{ji} + \sigma_j^*[W_k]_{ij}}{(\sigma_j^*)^2 - (\sigma_i^*)^2}, \quad s+1 \le i \le n-t,\ 1 \le j \le n-t,\ i > j,$$

$$[Y_k]_{ij} := 0, \quad n-t+1 \le i,\ j \le n,$$

$$[Y_k]_{ij} := -[Y_k]_{ji} = -\frac{[W_k]_{ij}}{\sigma_i^*}, \quad 1 \le i,\ j \le s,\ i > j,$$

$$[Y_k]_{ij} := -[Y_k]_{ji} = -\frac{[W_k]_{ij}}{\sigma_i^*}, \quad n-t+1 \le j \le n, \quad 1 \le i \le n-t,$$

$$[Y_k]_{ij} := -[Y_k]_{ji} = \frac{\sigma_i^*[W_k]_{ij} + \sigma_j^*[W_k]_{ji}}{(\sigma_j^*)^2 - (\sigma_i^*)^2}, \quad s+1 \le i \le n-t,\ 1 \le j \le n-t,\ i > j.$$

(e) Compute $U_{k+1} := [\mathbf{u}_1^{k+1}, \ldots, \mathbf{u}_n^{k+1}]$ and $V_{k+1} := [\mathbf{v}_1^{k+1}, \ldots, \mathbf{v}_n^{k+1}]$ by solving

$$\left(I + \frac{1}{2}X_k\right) U_{k+1}^T = \left(I - \frac{1}{2}X_k\right) U_k^T \tag{3.4}$$

and

$$\left(I + \frac{1}{2}Y_k\right) V_{k+1}^T = \left(I - \frac{1}{2}Y_k\right) V_k^T. \tag{3.5}$$

Now we present a convergence analysis for the Newton-type method. Recall that we have assumed that the given singular values satisfy (1.5). There is no difficulty in generalizing all our results to an arbitrary set of given positive singular values. Let $\{\mathbf{c}^k\}$, $\{U_k\}$, $\{V_k\}$, $\{X_k\}$, $\{Y_k\}$, and $\{J_k\}$ be generated by the Newton-type method with initial point $\mathbf{c}^0$. Let E_{U_k} and E_{V_k} be defined by (2.5)-(2.6) with $\{U = U_k\}$ and $\{V = V_k\}$ respectively. Then we have the following lemma.

Lemma 3.1. *There exist two numbers $\delta \in (0,\ 1)$ and $\gamma \in [1,\ +\infty)$ such that for any $k \ge 0$ and $[U(\mathbf{c}^*)\ V(\mathbf{c}^*)] \in \mathcal{W}(\mathbf{c}^*)$ with $\max\{\|E_{U_k}\|,\ \|E_{V_k}\} < \delta$, the following assertions hold:*

(i) $\|\mathbf{c}^{k+1} - \mathbf{c}^*\| \le \gamma \|J_k^{-1}\|(\|E_{U_k}\|^2 + \|E_{V_k}\|^2)$, *if J_k^{-1} exists;*

(ii) $\max\{\|U_{k+1} - U_k\|,\ \|V_{k+1} - V_k\|\} \le \gamma(\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + \|E_{U_k}\| + \|E_{V_k}\|)$, *if $\mathbf{c}^{k+1} \in \mathbf{B}(\mathbf{c}^*, \delta)$;*

(iii) $\max\{\|E_{U_{k+1}}\|,\ \|E_{V_{k+1}}\|\} \le \gamma[\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + (\|E_{U_k}\| + \|E_{V_k}\|)^2]$, *if $\mathbf{c}^{k+1} \in \mathbf{B}(\mathbf{c}^*, \delta)$.*

Proof. Let $\{\tilde{X}_k\}$ and $\{\tilde{Y}_k\}$ be the skew-symmetric matrices defined by (2.4) with $\{U = U_k\}$ and $\{V = V_k\}$ respectively. By Lemma 2.7, let $\delta_1 \in (0,\ 1)$ and $\gamma_1 \in [1,\ +\infty)$ be such that for any $k \ge 0$,

$$\|\tilde{X}_k\|_F \le \gamma_1 \|E_{U_k}\|, \quad \|\tilde{X}_k^{[s]_U}\|_F \le \gamma_1 \|E_{U_k}\|^2, \quad \|\tilde{X}_k^{[t]_L}\|_F \le \gamma_1 \|E_{U_k}\|^2, \tag{3.6}$$

$$\|\tilde{Y}_k\|_F \le \gamma_1 \|E_{V_k}\|, \quad \|\tilde{Y}_k^{[s]_U}\|_F \le \gamma_1 \|E_{V_k}\|^2, \quad \|\tilde{Y}_k^{[t]_L}\|_F \le \gamma_1 \|E_{V_k}\|^2, \tag{3.7}$$

when $\max\{\|E_{U_k}\|,\ \|E_{V_k}\} < \delta_1$. Let α the positive number determined in Lemma 2.3. Write

$$\eta_1 := (n^2 - s^2 - t^2) \max_{s \le i < n-t} \left\{ \frac{1}{\sigma_{i+1}^* - \sigma_i^*},\ \frac{1}{\sigma_i^*} \right\}, \quad \eta_2 := \max\{2\eta_1\alpha,\ 4 + 2\gamma_1 + 8\gamma_1\eta_1\|\boldsymbol{\sigma}^*\|\}. \tag{3.8}$$

Set

$$\gamma := \max\left\{4\gamma_1^2\|\boldsymbol{\sigma}^*\|,\ \frac{9}{2}\sqrt{n}\gamma_1\eta_2\right\} \quad \text{and} \quad \delta := \min\left\{\delta_1,\ \frac{2}{3\gamma}\right\}. \tag{3.9}$$

Clearly, $\delta \in (0,\ 1)$ and $\gamma \in [1,\ +\infty)$. Below we prove that δ and γ are as desired. For this purpose, we assume that $[U(\mathbf{c}^*),\ V(\mathbf{c}^*)] \in \mathcal{W}(\mathbf{c}^*)$. Let $k \ge 0$ be such that $\max\{\|E_{U_k}\|,\ \|E_{V_k}\|\} < \delta$. Then one has by (3.6), (3.7), and (3.9) that

$$\|\tilde{X}_k\|_F \le \gamma_1\|E_{U_k}\| < \gamma\delta < 1 \quad \text{and} \quad \|\tilde{Y}_k\|_F \le \gamma_1\|E_{V_k}\| < \gamma\delta < 1.$$

Thus, by direct computations, we have

$$\left\| \sum_{m=2}^{\infty} \frac{\tilde{X}_k^{m-2}}{m!} \right\|_F < \sum_{m=2}^{\infty} \frac{1}{m!} \le \sum_{m=2}^{\infty} \frac{1}{m(m-1)} = 1. \tag{3.10}$$

Similarly,

$$\left\| \sum_{m=2}^{\infty} \frac{(-\tilde{Y}_k)^{m-2}}{m!} \right\|_F < 1, \qquad \left\| \sum_{m=1}^{\infty} \frac{(-\tilde{Y}_k)^{m-1}}{m!} \right\|_F < 2, \qquad \left\| \sum_{m=0}^{\infty} \frac{(-\tilde{Y}_k)^{m}}{m!} \right\|_F < 3.$$

Write

$$R_k := -\tilde{X}_k^2\Big(\sum_{m=2}^{\infty} \frac{\tilde{X}_k^{m-2}}{m!}\Big)\Sigma^*\Big(\sum_{m=0}^{\infty} \frac{(-\tilde{Y}_k)^{m}}{m!}\Big) - \Sigma^*\tilde{Y}_k^2 \sum_{m=2}^{\infty} \frac{(-\tilde{Y}_k)^{m-2}}{m!} + \tilde{X}_k\Sigma^*\tilde{Y}_k \sum_{m=1}^{\infty} \frac{(-\tilde{Y}_k)^{m-1}}{m!}. \tag{3.11}$$

Hence, one has by (3.10)–(3.11) that

$$\|R_k\|_F \le (3\|\tilde{X}_k\|_F^2 + \|\tilde{Y}_k\|_F^2 + 2\|\tilde{X}_k\|_F \cdot \|\tilde{Y}_k\|_F) \cdot \|\Sigma^*\|_F \le 4(\|\tilde{X}_k\|_F^2 + \|\tilde{Y}_k\|_F^2) \cdot \|\Sigma^*\|_F. \tag{3.12}$$

It follows from (2.2), (3.6), (3.7), and (3.12) that

$$\|R_k\|_F \le 4\gamma_1^2\|\boldsymbol{\sigma}^*\|(\|E_{U_k}\|^2 + \|E_{V_k}\|^2) \le \gamma(\|E_{U_k}\|^2 + \|E_{V_k}\|^2), \tag{3.13}$$

where the last inequality holds because of the definition of γ in (3.9). On the other hand, noting that $e^{\tilde{X}_k} := U_k^T\tilde{U}(\mathbf{c}^*)$, $e^{\tilde{Y}_k} := V_k^T\tilde{V}(\mathbf{c}^*)$, and $[\tilde{U}(\mathbf{c}^*),\ \tilde{V}(\mathbf{c}^*)] \in \mathcal{W}(\mathbf{c}^*)$, we derive

$$e^{\tilde{X}_k}\Sigma^*e^{-\tilde{Y}_k} = U_k^T A(\mathbf{c}^*)V_k. \tag{3.14}$$

Thus, by (3.11) and the fact of $e^X = \sum_{m=0}^{\infty} \frac{X^m}{m!}$, we can write (3.14) into the form

$$\Sigma^* + \tilde{X}_k\Sigma^* - \Sigma^*\tilde{Y}_k = U_k^T A(\mathbf{c}^*)V_k + R_k, \tag{3.15}$$

The diagonal equalities of (3.15) are

$$(\mathbf{u}_i^k)^T A(\mathbf{c}^*)\mathbf{v}_i^k - \sigma_i^* = [R_k]_{ii}, \quad 1 \le i \le n. \tag{3.16}$$

Hence, by the definitions of $\boldsymbol{\sigma}^*$, J_k, $\mathbf{b}^k$, and $A(\mathbf{c}^*)$, we get from (3.16) that

$$J_k\mathbf{c}^* + \mathbf{b}^k - \boldsymbol{\sigma}^* = \mathrm{diag}(R_k)$$

Substituting this from (3.3), one has

$$J_k(\mathbf{c}^{k+1} - \mathbf{c}^*) = \mathrm{diag}(R_k).$$

Therefore, assertion (i) is seen to hold by (3.13).

For the proof of assertions (ii) and (iii), we assume further that $\mathbf{c}^{k+1} \in \mathbf{B}(\mathbf{c}^*, \delta)$ (and so $\|\mathbf{c}^{k+1} - \mathbf{c}^*\| < \delta$). The estimates of $\|\tilde{X}_k - X_k\|$, $\|X_k\|$, and $\left\|\left(I - \frac{1}{2}X_k\right)^{-1}\right\|$ are needed first. Indeed, using (3.15) and applying Lemma 2.5 (to $\tilde{X}_k$, $\tilde{Y}_k$, $U_k^TA(\mathbf{c}^*)V_k + R_k - \Sigma^*$ in place of H, K and Z), one has that

$$[\tilde{X}_k]_{ij} = \frac{(\mathbf{u}_i^k)^TA(\mathbf{c}^*)\mathbf{v}_j^k + [R_k]_{ij}}{\sigma_j^*}, \quad n-t+1 \le i \le n,\ 1 \le j \le n-t$$

and

$$[\tilde{X}_k]_{ij} = \frac{\sigma_j^*[(\mathbf{u}_i^k)^TA(\mathbf{c}^*)\mathbf{v}_j^k + [R_k]_{ij}] + \sigma_i^*[(\mathbf{u}_j^k)^TA(\mathbf{c}^*)\mathbf{v}_i^k + [R_k]_{ji}]}{(\sigma_j^*)^2 - (\sigma_i^*)^2}, \ s{+}1 \le i \le n{-}t,\ 1 \le j \le n{-}t, i > j.$$

This together with the formulation of X_k in the Newton-type method yields that

$$[\tilde{X}_k]_{ij} - [X_k]_{ij} = \frac{(\mathbf{u}_i^k)^T\Delta_{k+1}\mathbf{v}_j^k + [R_k]_{ij}}{\sigma_j^*}, \quad n-t+1 \le i \le n,\ 1 \le j \le n-t, \tag{3.17}$$

and

$$\begin{aligned}&[\tilde{X}_k]_{ij} - [X_k]_{ij}\\ =\ &\frac{\sigma_j^*(\mathbf{u}_i^k)^T\Delta_{k+1}\mathbf{v}_j^k + \sigma_i^*(\mathbf{u}_j^k)^T\Delta_{k+1}\mathbf{v}_i^k + \sigma_j^*[R_k]_{ij} + \sigma_i^*[R_k]_{ji}}{(\sigma_j^*)^2 - (\sigma_i^*)^2}, \ s+1 \le i \le n-t,\ 1 \le j \le n-t, i > j.\end{aligned} \tag{3.18}$$

where and in sequel $\Delta_{k+1} := A(\mathbf{c}^*) - A(\mathbf{c}^{k+1})$. Note that $\{\mathbf{u}_i^k\}_{i=1}^n$ and $\{\mathbf{v}_i^k\}_{i=1}^n$ are orthonormal and that, by Lemma 2.3,

$$\|\Delta_{k+1}\| \le \alpha\|\mathbf{c}^{k+1} - \mathbf{c}^*\|.$$

One has by (3.17) and (3.18) that

$$|[\tilde{X}_k]_{ij} - [X_k]_{ij}| \le \frac{1}{\sigma_j^*}\left(\alpha\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + \|R_k\|_F\right), \quad n-t+1 \le i \le n,\ 1 \le j \le n-t, \tag{3.19}$$

$$|[\tilde{X}_k]_{ij} - [X_k]_{ij}| \le \frac{1}{\sigma_j^* - \sigma_i^*}\left(\alpha\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + \|R_k\|_F\right), \quad s+1 \le i \le n-t,\ 1 \le j \le n-t,\ i > j. \tag{3.20}$$

Since $[X_k]_{ij} = 0$ for each $1 \le i,\ j \le s$ or $n-t+1 \le i,\ j \le n$, we have by (2.1), (3.19), (3.20) and the definition of η_1 that

$$\|\tilde{X}_k - X_k\| \le \|\tilde{X}_k - X_k\|_F \le \|\tilde{X}_k^{[s]_U}\|_F + \|\tilde{X}_k^{[t]_L}\|_F + \eta_1\left(\alpha\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + \|R_k\|_F\right),$$

Combining this with (3.6) and (3.13), we further derive that

$$\|\tilde{X}_k - X_k\| \le 2\gamma_1\|E_{U_k}\|^2 + \eta_1\left[\alpha\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + 4\gamma_1^2\|\boldsymbol{\sigma}^*\|(\|E_{U_k}\|^2 + \|E_{V_k}\|^2)\right], \tag{3.21}$$

$$\|X_k\| \le \gamma_1 \|E_{U_k}\| + 2\gamma_1 \|E_{U_k}\|^2 + \eta_1 \left[\alpha\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + 4\gamma_1^2 \|\boldsymbol{\sigma}^*\|(\|E_{U_k}\|^2 + \|E_{V_k}\|^2)\right].$$

Thus, by the fact of $\gamma_1 \max\{\|E_{U_k}\|,\ \|E_{V_k}\|\} \le \gamma_1\delta < 1$, one has

$$\|X_k\| \le \eta_1\alpha\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + (2 + \gamma_1 + 4\gamma_1\eta_1\|\boldsymbol{\sigma}^*\|)\|E_{U_k}\| + 4\gamma_1\eta_1\|\boldsymbol{\sigma}^*\| \cdot \|E_{V_k}\|; \tag{3.22}$$

hence,

$$\|X_k\| \le \frac{\eta_2}{2}\left(\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + \|E_{U_k}\| + \|E_{V_k}\|\right) \le \frac{\gamma}{2}\left(\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + \|E_{U_k}\| + \|E_{V_k}\|\right) \tag{3.23}$$

(noting that $\gamma \ge \eta_2 = \max\{2\eta_1\alpha,\ 4 + 2\gamma_1 + 8\gamma_1\eta_1\|\boldsymbol{\sigma}^*\|\}$). Since $\max\{\|E_{U_k}\|,\ \|E_{V_k}\|\} \le \delta$ and $\|\mathbf{c}^{k+1} - \mathbf{c}^*\| \le \delta$, we derive further by (3.9) and (3.22) that $\|X_k\| \le 1$. Therefore, applying Lemma 2.1 (for $A = I - \frac{1}{2}X_k$ and $B = I$), one has

$$\left\|\left(I - \tfrac{1}{2}X_k\right)^{-1}\right\| \le \frac{1}{1 - \frac{1}{2}\|X_k\|} \le 2. \tag{3.24}$$

Consequently, the estimates of $\|\tilde{X}_k - X_k\|$, $\|X_k\|$, and $\left\|\left(I - \frac{1}{2}X_k\right)^{-1}\right\|$ are complete. By a similar argument, we can have the following estimates:

$$\|\tilde{Y}_k - Y_k\| \le 2\gamma_1\|E_{V_k}\|^2 + \eta_1\left[\alpha\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + 4\gamma_1^2\|\boldsymbol{\sigma}^*\|(\|E_{U_k}\|^2 + \|E_{V_k}\|^2)\right],$$

$$\|Y_k\| \le \frac{\eta_2}{2}\left(\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + \|E_{U_k}\| + \|E_{V_k}\|\right) \le \frac{\gamma}{2}\left(\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + \|E_{U_k}\| + \|E_{V_k}\|\right), \tag{3.25}$$

and

$$\left\|\left(I - \tfrac{1}{2}Y_k\right)^{-1}\right\| \le 2. \tag{3.26}$$

Now we offer the estimates of $\|U_{k+1} - U_k\|$, $\|V_{k+1} - V_k\|$, $\|E_{U_{k+1}}\|$, and $\|E_{V_{k+1}}\|$. Note by (3.4) that

$$U_{k+1} - U_k = U_k[(I + \tfrac{1}{2}X_k) - (I - \tfrac{1}{2}X_k)](I - \tfrac{1}{2}X_k)^{-1} = U_kX_k(I - \tfrac{1}{2}X_k)^{-1}.$$

This together with (3.23), (3.24), and the orthonormal property of U_k gives rise to

$$\|U_{k+1} - U_k\| \le \gamma\left(\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + \|E_{U_k}\| + \|E_{V_k}\|\right).$$

Similarly, using (3.5), (3.25), and (3.26), we obtain

$$\|V_{k+1} - V_k\| \le \gamma\left(\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + \|E_{U_k}\| + \|E_{V_k}\|\right).$$

Thus, assertion (ii) holds. It remains to prove assertion (iii). The arguments for the estimates of $E_{U_{k+1}}$ and $E_{V_{k+1}}$ are similar and so we only provide the proof for the estimate of $E_{U_{k+1}}$. For this, note by (3.4) and the definition of $\tilde{X}_k$ that

$$U_{k+1} - \tilde{U}(\mathbf{c}^*) = U_k[(I + \tfrac{1}{2}X_k)(I - \tfrac{1}{2}X_k)^{-1} - e^{\tilde{X}_k}] = U_k[(I + \tfrac{1}{2}X_k) - e^{\tilde{X}_k}(I - \tfrac{1}{2}X_k)](I - \tfrac{1}{2}X_k)^{-1}.$$

Then, using the equality $e^{\tilde{X}_k} = \sum\limits_{m=0}^{\infty} \frac{\tilde{X}_k^m}{m!}$, it is easy to check that

$$\begin{aligned} U_{k+1} - \tilde{U}(\mathbf{c}^*) &= U_k\Big[X_k - \tilde{X}_k + \tfrac{1}{2}\tilde{X}_kX_k - \Big(\tilde{X}_k^2\sum_{m=2}^{\infty}\frac{\tilde{X}_k^{m-2}}{m!}\Big)\Big(I - \frac{1}{2}X_k\Big)\Big]\Big(I - \frac{1}{2}X_k\Big)^{-1} \\ &= U_k(X_k - \tilde{X}_k)(I - \tfrac{1}{2}X_k)^{-1} + \tfrac{1}{2}U_k\tilde{X}_kX_k(I - \tfrac{1}{2}X_k)^{-1} - U_k\tilde{X}_k^2\sum_{m=2}^{\infty}\frac{\tilde{X}_k^{m-2}}{m!}. \end{aligned} \tag{3.27}$$

Noting that U_k is orthonormal, we deduce from (3.27), (3.10), (3.24), and (2.1) that

$$\|U_{k+1}-\tilde{U}(\mathbf{c}^*)\| \le 2\|X_k-\tilde{X}_k\|+\|\tilde{X}_k\|\cdot\|X_k\|+\|\tilde{X}_k\|^2 \le 2\|X_k-\tilde{X}_k\|+\|\tilde{X}_k\|_F\cdot\|X_k\|+\|\tilde{X}_k\|_F^2.$$

Thus, one has by (3.6), (3.21), and (3.23) that

$$\begin{array}{rcl}\|U_{k+1}-\tilde{U}(\mathbf{c}^*)\| & \le & (4\gamma_1+\gamma_1^2+8\gamma_1^2\eta_1\|\boldsymbol{\sigma}^*\|+\frac{1}{2}\gamma_1\eta_2)(\|E_{U_k}\|+\|E_{V_k}\|)^2 \\ & & +(2\eta_1\alpha+\frac{1}{2}\eta_2)\|\mathbf{c}^{k+1}-\mathbf{c}^*\|.\end{array} \tag{3.28}$$

Recall from (3.8) that $\eta_2=\max\{2\eta_1\alpha,\ 4+2\gamma_1+8\gamma_1\eta_1\|\boldsymbol{\sigma}^*\|\}$. We then have by (3.28) that

$$\|U_{k+1}-\tilde{U}(\mathbf{c}^*)\| \le \frac{3}{2}\gamma_1\eta_2[(\|E_{U_k}\|+\|E_{V_k}\|)^2+\|\mathbf{c}^{k+1}-\mathbf{c}^*\|] \tag{3.29}$$

(noting that $\gamma_1 \ge 1$). To proceed, write $U_{k+1} := [U_{k+1}^{(1)},\ U_{k+1}^{(2)},\ U_{k+1}^{(3)}]$ where $U_{k+1}^{(1)} \in \mathbb{R}^{n\times s}$ and $U_{k+1}^{(3)} \in \mathbb{R}^{n\times t}$. Since $(I-\Pi_{U,i})\tilde{U}^{(i)}(\mathbf{c}^*)=\mathbf{0}$ and $\|I-\Pi_{U,i}\|\le 1$ hold for $i=1,\ 3$, one has

$$\|(I-\Pi_{U,i})U_{k+1}^{(i)}\| = \|(I-\Pi_{U,i})(U_{k+1}^{(i)}-\tilde{U}^{(i)}(\mathbf{c}^*))\| \le \|U_{k+1}-\tilde{U}(\mathbf{c}^*)\|, \quad i=1,\ 3. \tag{3.30}$$

Noting that $E_{U_{k+1}}=[(I-\Pi_{U,1})U_{k+1}^{(1)},\ U_{k+1}^{(2)}-U^{(2)}(\mathbf{c}^*),\ (I-\Pi_{U,3})U_{k+1}^{(3)}]$, we obtain from (2.1), (3.29), and (3.30) that

$$\begin{array}{rcl}\|E_{U_{k+1}}\| & \le & \|(I-\Pi_{U,1})U_{k+1}^{(1)}\|_F+\|U_{k+1}^{(2)}-U^{(2)}(\mathbf{c}^*)\|_F+\|(I-\Pi_{U,3})U_{k+1}^{(3)}\|_F \\ & \le & 3\sqrt{n}\|U_{k+1}-\tilde{U}(\mathbf{c}^*)\|.\end{array} \tag{3.31}$$

Therefore, thanks to (3.29) and (3.31), one sees that

$$\|E_{U_{k+1}}\| \le \gamma[\|\mathbf{c}^{k+1}-\mathbf{c}^*\|+(\|E_{U_k}\|+\|E_{V_k}\|)^2];$$

hence, assertion (iii) holds. The proof is complete. □

Now we present a convergence result of the Newton-type method which shows that the sequence $\{\mathbf{c}^k\}$ generated by the Newton-type method converges quadratically to a solution of the ISVP (1.2). For this purpose, we require a basic assumption: $\partial_{Q|\mathcal{S}}\mathbf{f}(\mathbf{c}^*)$ is nonempty, which is guaranteed by Lemma 2.8 if $\mathbf{c}^*\in \mathrm{cl}\mathcal{S}$ (actually, the nonemptiness of $\partial_{Q|\mathcal{S}}\mathbf{f}(\mathbf{c}^*)$ is equivalent that $\mathbf{c}^*\in\mathrm{cl}\mathcal{S}$). Furthermore, we also require the assumption that the initial point $\mathbf{c}^0\in\mathcal{S}$, so that each J_k generated by the Newton-type method is close to $\partial_{Q|\mathcal{S}}\mathbf{f}(\mathbf{c}^*)$ (see the proof for Theorem 3.1 below).

Theorem 3.1. *Let $\mathbf{c}^*\in\mathrm{cl}\mathcal{S}$ such that the matrix $A(\mathbf{c}^*)$ has singular values given by (1.5). Suppose that each $J\in\partial_{Q|\mathcal{S}}\mathbf{f}(\mathbf{c}^*)$ is nonsingular. Then there exist $\delta\in(0,\ 1)$ such that for each $\mathbf{c}^0\in\mathbf{B}(\mathbf{c}^*,\delta)\cap\mathcal{S}$, the sequence $\{\mathbf{c}^k\}$ generated by the Newton-type method with initial point $\mathbf{c}^0$ converges quadratically to $\mathbf{c}^*$.*

Proof. By Lemma 3.1, let $\delta_1\in(0,\ 1)$ and $\gamma\in[1,\ +\infty)$ such that for any $k\ge 0$ and $[U(\mathbf{c}^*)\ V(\mathbf{c}^*)]\in\mathcal{W}(\mathbf{c}^*)$, if $\max\{\|E_{U_k}\|,\ \|E_{V_k}\}<\delta_1$, the assertions (i)–(iii) in Lemma 3.1 hold with $\delta=\delta_1$. Moreover, thanks to Lemmas 2.6 and 2.10, we assume without loss of generality that for any $\mathbf{c}\in\mathbf{B}(\mathbf{c}^*,\delta_1)$, (2.9) and the assertions (i)–(iv) in Lemma 2.6 hold. Write

$$q := 6\sqrt{n}\gamma.$$

Take δ such that

$$0 < \delta < \min\left\{\frac{\delta_1}{q},\ \frac{1}{4q\gamma(1+\gamma^2)},\ \frac{1}{6n\gamma^2 q\cdot\max\limits_j\|A_j\|}\right\}. \tag{3.32}$$

Clearly, $\delta \in (0,\ 1)$. Below we shall show that δ is as desired. For this purpose, let $\mathbf{c}^0 \in \mathbf{B}(\mathbf{c}^*, \delta)\bigcap\mathcal{S}$. Then, thanks to Lemma 2.9 and the definition of J_0, one has that $\partial_{Q|\mathcal{S}}\mathbf{f}(\mathbf{c}^0) = \{\mathbf{f}'(\mathbf{c}^0)\} = \{J_0\}$. In addition, by Lemma 2.10 (as $\delta < \frac{2\delta_1}{q} \le \delta_1$), we have

$$\|J_0^{-1}\| \le \gamma. \tag{3.33}$$

It suffices to prove that for any $k = 0,\ 1,\ \ldots,$

$$\|\mathbf{c}^k - \mathbf{c}^*\| \le q\delta\left(\frac{1}{2}\right)^{2^k} \tag{3.34}$$

and

$$\max\{\|E_{U_k}\|,\ \|E_{V_k}\|\} \le q\delta\left(\frac{1}{2}\right)^{2^k}, \tag{3.35}$$

We proceed by mathematical induction. Since $\|\mathbf{c}^0 - \mathbf{c}^*\| < \delta$ and $q \ge 2$, (3.34) is trivial for $k = 0$. Noting that $E_{U_0} = [(I - \Pi_{U,1})U_0^{(1)},\ U_0^{(2)} - U^{(2)}(\mathbf{c}^*),\ (I - \Pi_{U,3})U_0^{(3)}]$, one has by (2.1) and Lemma 2.6 that

$$\|E_{U_0}\| \le \|E_{U_0}\|_F \le \|(I - \Pi_{U,1})U_0^{(1)}\|_F + \|U_0^{(2)} - U^{(2)}(\mathbf{c}^*)\|_F + \|(I - \Pi_{U,3})U_0^{(3)}\|_F \le 3\sqrt{n}\gamma\delta = \frac{1}{2}q\delta,$$

where the equality holds because of the definition of q. Similarly, one can prove that $\|E_{V_0}\| \le \frac{1}{2}q\delta$; hence, (3.35) is shown for $k = 0$. Assume that (3.34) and (3.35) hold for all $k \le l$. Then, by (3.32),

$$\|\mathbf{c}^k - \mathbf{c}^*\| \le \frac{1}{2}q\delta < \delta_1 \quad \text{and} \quad \max\{\|E_{U_k}\|,\ \|E_{V_k}\| \le \frac{1}{2}q\delta < \delta_1, \quad \text{for each } k \le l. \tag{3.36}$$

Thus, applying Lemma 3.1, we get the following two assertions:

$$J_k^{-1} \text{ exists } \Rightarrow \|\mathbf{c}^{k+1} - \mathbf{c}^*\| \le \gamma\|J_k^{-1}\|(\|E_{U_k}\|^2 + \|E_{V_k}\|^2), \quad \text{for each } k \le l, \tag{3.37}$$

and

$$\max\{\|U_{k+1} - U_k\|,\ \|V_{k+1} - V_k\|\} \le \gamma(\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + \|E_{U_k}\| + \|E_{V_k}\|), \quad \text{for each } k \le l-1. \tag{3.38}$$

Noting that

$$\|U_l - U_0\| \le \sum_{k=0}^{l-1}\|U_{k+1} - U_k\| \quad \text{and} \quad \|V_l - V_0\| \le \sum_{k=0}^{l-1}\|V_{k+1} - V_k\|,$$

we have by (3.38), (3.34) (with $k \le l-1$), and (3.35) (with $k \le l-1$) that

$$\max\{\|U_l - U_0\|,\ \|V_l - V_0\|\} \le 3\gamma q\delta\left[\frac{1}{2} + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^{2^2} + \cdots + \left(\frac{1}{2}\right)^{2^{l-1}}\right].$$

Since $2^n \ge n+1$ for each $n \ge 0$, it follows that

$$\max\{\|U_l - U_0\|,\ \|V_l - V_0\|\} \le 3\gamma q\delta\left[\frac{1}{2} + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 + \cdots + \left(\frac{1}{2}\right)^l\right] \le 3\gamma q\delta.$$

This together with the definition of δ yields

$$2n\gamma \cdot \max_j \|A_j\| \cdot \max\{\|U_l - U_0\|,\ \|V_l - V_0\|\} \leq 6n\gamma^2 q\delta \cdot \max_j \|A_j\| < \frac{1}{2}. \tag{3.39}$$

Moreover, by the definition of $[J_l]_{ij}$ (cf. (3.2)), one has

$$\big|[J_l]_{ij} - [J_0]_{ij}\big| = \big|(\mathbf{u}_i^l - \mathbf{u}_i^0)^T A_j \mathbf{v}_i^l + (\mathbf{u}_i^0)^T A_j(\mathbf{v}_i^l - \mathbf{v}_i^0)\big| \leq 2\|A_j\| \cdot \max\{\|\mathbf{u}_i^l - \mathbf{u}_i^0\|,\ \|\mathbf{v}_i^l - \mathbf{v}_i^0\|\}.$$

Then, thanks to (2.1), we get that

$$\|J_l - J_0\| \leq \|J_l - J_0\|_F \leq 2n \max_j \|A_j\| \cdot \max\{\|U_l - U_0\|,\ \|V_l - V_0\|\}.$$

It follows from (3.33) and (3.39) that

$$\|J_0^{-1}\| \cdot \|J_l - J_0\| \leq 2n\gamma \cdot \max_j \|A_j\| \cdot \max\{\|U_l - U_0\|,\ \|V_l - V_0\|\} < \frac{1}{2}.$$

Thus, applying Lemma 2.1 (for $A = J_l$ and $B = J_0$) and using (3.33) again, we obtain

$$\|J_l^{-1}\| \leq \frac{\|J_0^{-1}\|}{1 - \|J_0^{-1}\| \cdot \|J_l - J_0\|} < 2\gamma.$$

Then, by implication (3.37) and the inductive assumption (3.35) with $k = l$, we obtain

$$\|\mathbf{c}^{l+1} - \mathbf{c}^*\| \leq 2\gamma^2(\|E_{U_l}\|^2 + \|E_{V_l}\|^2) \leq 4\gamma^2(q\delta)^2 \left(\frac{1}{2}\right)^{2^{l+1}}. \tag{3.40}$$

Thus, (3.34) holds for $k = l + 1$ and moreover $\|\mathbf{c}^{l+1} - \mathbf{c}^*\| \leq \delta_1$ as $q\delta \leq \min\{\delta_1,\ 1/(4\gamma^2)\}$ by (3.32) and the fact of $\gamma \geq 1$. Hence, noting (3.36), Lemma 3.1 (ii) and (iii) (with $k = l$) are applicable to concluding that

$$\max\{\|U_{l+1} - U_l\|,\ \|V_{l+1} - V_l\|\} \leq \gamma(\|\mathbf{c}^{l+1} - \mathbf{c}^*\| + \|E_{U_l}\| + \|E_{V_l}\|)$$

and

$$\max\{\|E_{U_{l+1}}\|,\ \|E_{V_{l+1}}\|\} \leq \gamma[\|\mathbf{c}^{l+1} - \mathbf{c}^*\| + (\|E_{U_l}\| + \|E_{V_l}\|)^2]$$

Therefore, we derive from (3.40) and the inductive assumption (3.35) (with $k = l$) that

$$\max\{\|E_{U_{l+1}}\|,\ \|E_{V_{l+1}}\|\} \leq 4\gamma(1 + \gamma^2)(q\delta)^2 \left(\frac{1}{2}\right)^{2^{l+1}}.$$

Thus, thanks to (3.32), (3.35) holds for $k = l + 1$ and the proof is complete.

□

Theorems 3.2 and 3.3 below, the proofs of which are similar to that of Theorems 3.1, show that the condition $\mathbf{c}^* \in \mathrm{cl}\mathcal{S}$ is not required if the nonsingularity assumption for each $J \in \partial_{Q|\mathcal{S}}\mathbf{f}(\mathbf{c}^*)$ is replaced by the nonsingularity assumption for each $J \in \partial_B \mathbf{f}(\mathbf{c}^*)$ or each $J \in \partial_Q \mathbf{f}(\mathbf{c}^*)$.

Theorem 3.2. *Let $\mathbf{c}^* \in \mathbb{R}^n$ be such that the matrix $A(\mathbf{c}^*)$ has singular values given by* (1.5)*. Suppose that each $J \in \partial_B \mathbf{f}(\mathbf{c}^*)$ is nonsingular. Then there exist $\delta \in (0,\ 1)$ such that for each $\mathbf{c}^0 \in \mathbf{B}(\mathbf{c}^*, \delta) \cap \mathcal{S}$, the sequence $\{\mathbf{c}^k\}$ generated by the Newton-type method with initial point $\mathbf{c}^0$ converges quadratically to $\mathbf{c}^*$.*

Theorem 3.3. *Let $\mathbf{c}^* \in \mathbb{R}^n$ be such that the matrix $A(\mathbf{c}^*)$ has singular values given by* (1.5)*. Suppose that each $J \in \partial_Q \mathbf{f}(\mathbf{c}^*)$ is nonsingular. Then there exist $\delta \in (0,\ 1)$ such that for each $\mathbf{c}^0 \in \mathbf{B}(\mathbf{c}^*, \delta) \bigcap \mathcal{S}$, the sequence $\{\mathbf{c}^k\}$ generated by the Newton-type method with initial point $\mathbf{c}^0$ converges quadratically to $\mathbf{c}^*$.*

4 An inexact version

In this section, we design an inexact version of the Newton-type method proposed in Section 3 where the linear system (3.3) is solved inexactly, and establish some convergence results for the inexact version. In the remainder of the paper, let $\beta \in (1,\ 2]$.

Algorithm 2. the inexact Newton-type method

1. Given $\mathbf{c}^0 \in \mathbb{R}^n$, compute the singular values $\{\sigma_i(\mathbf{c}^0)\}_{i=1}^n$, the orthonormal left singular vectors $\{\mathbf{u}_i(\mathbf{c}^0)\}_{i=1}^m$ and right singular vectors $\{\mathbf{v}_i(\mathbf{c}^0)\}_{i=1}^n$ of $A(\mathbf{c}^0)$. Write

$$U_0 := [\mathbf{u}_1^0, \ldots, \mathbf{u}_m^0] = [\mathbf{u}_1(\mathbf{c}^0), \ldots, \mathbf{u}_m(\mathbf{c}^0)],$$

$$V_0 := [\mathbf{v}_1^0, \ldots, \mathbf{v}_n^0] = [\mathbf{v}_1(\mathbf{c}^0), \ldots, \mathbf{v}_n(\mathbf{c}^0)],$$

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

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

   (a) Same as (a) in Algorithm 1.

   (b) Solve (3.3) to find $\mathbf{c}^{k+1}$ such that the residual $\mathbf{r}^k$ defined by

$$\mathbf{r}^k := J_k \mathbf{c}^{k+1} + \mathbf{b}^k - \boldsymbol{\sigma}^* \tag{4.1}$$

   satisfies that

$$\|\mathbf{r}^k\| \le \frac{\|\boldsymbol{\sigma}^k - \boldsymbol{\sigma}^*\|^\beta}{\|\boldsymbol{\sigma}^*\|^\beta}, \quad \beta \in (1,\ 2]. \tag{4.2}$$

   (c) Same as (c) in Algorithm 1.

   (d) Same as (d) in Algorithm 1.

   (e) Same as (e) in Algorithm 1.

   (f) Compute $\boldsymbol{\sigma}^{k+1} := (\boldsymbol{\sigma}_1^{k+1}, \ldots, \boldsymbol{\sigma}_n^{k+1})^T$ by

$$\boldsymbol{\sigma}_i^{k+1} = (\mathbf{u}_i^{k+1})^T A(\mathbf{c}^{k+1}) \mathbf{v}_i^{k+1}, \quad 1 \le i \le n. \tag{4.3}$$

To establish the convergence results of the inexact Newton-type method, we also need the following lemma. For this purpose, let $\{\mathbf{c}^k\}$, $\{U_k\}$, $\{V_k\}$, $\{X_k\}$, $\{Y_k\}$, and $\{J_k\}$ be generated by the inexact Newton-type method with initial point $\mathbf{c}^0$.

Lemma 4.1. *There exist two numbers $\delta \in (0,\ 1)$ and $\gamma \in [1,\ +\infty)$ such that for any $k \ge 0$ and $[U(\mathbf{c}^*)\ V(\mathbf{c}^*)] \in \mathcal{W}(\mathbf{c}^*)$ with $\max\{\|E_{U_k}\|,\ \|E_{V_k}\} < \delta$, the following assertions hold:*

(i) $\|\mathbf{c}^{k+1} - \mathbf{c}^*\| \le \gamma \|J_k^{-1}\| (\|E_{U_k}\|^2 + \|E_{V_k}\|^2 + \|\boldsymbol{\sigma}^k - \boldsymbol{\sigma}^*\|^\beta)$, *if J_k^{-1} exists;*

(ii) $\max\{\|E_{U_{k+1}}\|,\ \|E_{V_{k+1}}\|\} \le \gamma[\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + (\|E_{U_k}\| + \|E_{V_k}\|)^2]$, *if $\mathbf{c}^{k+1} \in \mathbf{B}(\mathbf{c}^*, \delta)$;*

(iii) $\max\{\|U_{k+1} - U_k\|,\ \|V_{k+1} - V_k\|\} \le \gamma(\|\mathbf{c}^{k+1} - \mathbf{c}^*\| + \|E_{U_k}\| + \|E_{V_k}\|)$, *if $\mathbf{c}^{k+1} \in \mathbf{B}(\mathbf{c}^*, \delta)$;*

(iv) $\|\boldsymbol{\sigma}^k - \boldsymbol{\sigma}^*\| \le \gamma(\|\mathbf{c}^k - \mathbf{c}^*\| + \|E_{U_k}\|^2 + \|E_{V_k}\|^2)$.

Proof. The proof for assertions (i)-(iii) is very similar to that for Lemma 3.1. Below we show assertion (iv). To do this, as in the proof for Lemma 3.1, let $\{\tilde{X}_k\}$ and $\{\tilde{Y}_k\}$ be the skew-symmetric matrices defined by (2.4) with $\{U = U_k\}$ and $\{V = V_k\}$ respectively, and let $\alpha > 0$, $\delta_1 \in (0,\ 1)$, $\gamma_1 \in [1,\ +\infty)$ be the numbers determined by Lemmas 2.3 and 2.7, respectively. Let R_k be defined by (3.11). Then, one can check as in the proof for Lemma 3.1 that (3.13) and (3.16) also hold. Thus, by Lemma 2.3 and using the orthogonality of $\{\mathbf{u}_i^k\}_{i=1}^n$ and $\{\mathbf{v}_i^k\}_{i=1}^n$, one has that

$$\begin{aligned} |(\mathbf{u}_i^k)^T A(\mathbf{c}^k)\mathbf{v}_i^k - \sigma_i^*| &= \|(\mathbf{u}_i^k)^T (A(\mathbf{c}^k) - A(\mathbf{c}^*))\mathbf{v}_i^k + (\mathbf{u}_i^k)^T A(\mathbf{c}^*)\mathbf{v}_i^k - \sigma_i^*\| \\ &\le \alpha\|\mathbf{c}^k - \mathbf{c}^*\| + 4\gamma_1^2\|\boldsymbol{\sigma}^*\|(\|E_{U_k}\|^2 + \|E_{V_k}\|^2). \end{aligned}$$

This, together with the definitions of $\boldsymbol{\sigma}^k$, $\boldsymbol{\sigma}^*$ in (4.3) and (2.2), gives that

$$\|\boldsymbol{\sigma}^k - \boldsymbol{\sigma}^*\| \le \gamma(\|\mathbf{c}^k - \mathbf{c}^*\| + (\|E_{U_k}\|^2 + \|E_{V_k}\|^2)),$$

where $\gamma := \max\{\sqrt{n}\alpha, 4\sqrt{n}\gamma_1^2\|\boldsymbol{\sigma}^*\|\}$, and assertion (iv) is seen to hold. The proof is complete. □

Note the known Lipschitz continuity of $\boldsymbol{\sigma}$ defined by (2.8); see [12, Corollary 8.6.2]. Then, following the line for proving Theorem 3.1 and using Lemma 4.1 (where Lemma 3.1 is used), one can establish the following convergence result for the inexact Newton-type method .

Theorem 4.1. *Let* $\mathbf{c}^* \in \mathrm{cl}\mathcal{S}$ *such that the matrix* $A(\mathbf{c}^*)$ *has singular values given by* (1.5)*. Suppose that each* $J \in \partial_{Q|\mathcal{S}}\mathbf{f}(\mathbf{c}^*)$ *is nonsingular. Then there exist* $\delta \in (0,\ 1)$ *such that for each* $\mathbf{c}^0 \in \mathbf{B}(\mathbf{c}^*, \delta)\cap\mathcal{S}$*, the sequence* $\{\mathbf{c}^k\}$ *generated by the inexact Newton-type method with initial point* $\mathbf{c}^0$ *converges to* $\mathbf{c}^*$ *and the convergence rate is equivalent to* β*.*

Similarly, we can also present the following two results where the nonsingularity assumption for all $J \in \partial_{Q|\mathcal{S}}\mathbf{f}(\mathbf{c}^*)$ is replaced by the nonsingularity assumption for each $J \in \partial_B\mathbf{f}(\mathbf{c}^*)$ or each $J \in \partial_Q\mathbf{f}(\mathbf{c}^*)$.

Theorem 4.2. *Let* $\mathbf{c}^* \in \mathbb{R}^n$ *be such that the matrix* $A(\mathbf{c}^*)$ *has singular values given by* (1.5)*. Suppose that each* $J \in \partial_B\mathbf{f}(\mathbf{c}^*)$ *is nonsingular. Then there exist* $\delta \in (0,\ 1)$ *such that for each* $\mathbf{c}^0 \in \mathbf{B}(\mathbf{c}^*, \delta)\cap\mathcal{S}$*, the sequence* $\{\mathbf{c}^k\}$ *generated by the inexact Newton-type method with initial point* $\mathbf{c}^0$ *converges to* $\mathbf{c}^*$ *and the convergence rate is equivalent to* β*.*

Theorem 4.3. *Let* $\mathbf{c}^* \in \mathbb{R}^n$ *be such that the matrix* $A(\mathbf{c}^*)$ *has singular values given by* (1.5)*. Suppose that each* $J \in \partial_Q\mathbf{f}(\mathbf{c}^*)$ *is nonsingular. Then there exist* $\delta \in (0,\ 1)$ *such that for each* $\mathbf{c}^0 \in \mathbf{B}(\mathbf{c}^*, \delta)\bigcap\mathcal{S}$*, the sequence* $\{\mathbf{c}^k\}$ *generated by the inexact Newton-type method with initial point* $\mathbf{c}^0$ *converges to* $\mathbf{c}^*$ *and the convergence rate is equivalent to* β*.*

5 Numerical tests

In this section, we report some numerical tests to illustrate the convergence performance of the proposed Newton-type methods (including the Newton-type and inexact Newton-type methods). Our aim is, for the square inverse singular value problems with multiple and/or zero singular values, to illustrate the validity of the Newton-type methods. All the tests were implemented in MATLAB 7.0 on a Genuine Intel(R) PC with 1.6 GHz CPU.

Let $\{T_i\}_{i=1}^n$ be Toeplitz matrices given by

$$T_1 = I, \quad T_2 = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \ddots & \vdots \\ 0 & 1 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 & 1 \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix}, \quad \dots, \quad T_n = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ 1 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

Let $H_1,\ H_2 \subset \mathcal{O}(n)$ generated by Matlab-provided random function. Define $A_0 = \mathbf{0}$ and $\{A_i\}_{i=1}^n \subset \mathbb{R}^{n\times n}$ as follows

$$A_i = H_1 T_i H_2, \quad \text{for each } i = 1, 2, \dots, n,$$

where $\mathbf{0}$ is a zero matrix of appropriate size. Here we focus on the following three cases: (a) $m = n = 50$; (b) $m = n = 100$; (c) $m = n = 200$. To present multiple and zero singular values, we first generate in each test a vector $\hat{\mathbf{c}}^*$ randomly such that, for some integers p and q, $|\sigma_{p+1}(\tilde{\mathbf{c}}^*) - \sigma_p(\tilde{\mathbf{c}}^*)| < 5e-5$ and $\sigma_q(\tilde{\mathbf{c}}^*) < 1e-4$, where $\tilde{\mathbf{c}}^* := \hat{\mathbf{c}}^* * 10^{-4}$. Set

$$\sigma_i^* = \begin{cases} \sigma_p(\tilde{\mathbf{c}}^*), & i = p,\ p+1; \\ 0, & i = q; \\ \sigma_i(\tilde{\mathbf{c}}^*), & \text{otherwise}. \end{cases}$$

Then we choose $\{\sigma_i^*\}_{i=1}^n$ as the prescribed singular values.

Since both the Newton-type and inexact Newton-type methods are locally convergent, $\mathbf{c}^0$ is formed by chopping the components of $\tilde{\mathbf{c}}^*$ to five decimal places for the case (a), and to six decimal places for the cases (b), (c). For each case, ten test problems are constructed. In both algorithms, we need to solve three linear systems: the approximate Jacobian equation (3.3), systems (3.4) and (3.5). Note that, in the Newton-type method, the approximate Jacobian equation (3.3) is required to be solved exactly. Thus, one may solve it by the direct method or choose an iterative method and solve up it to machine precision eps. Recall that the matrices on the left-hand side of (3.4) and (3.5) approach the identity matrix in the limit (cf. (3.23) and (3.25)). Hence one can expect to solve them accurately by iterative methods in just a few iterations. While in the inexact Newton-type method, the approximate Jacobian equation (3.3) is only required to be solved inexactly and so iterative methods can be considered here to reduce the computational cost of solving (3.3). As in [3], we choose the QMR method [10] via the MATLAB QMR function as the iterative method, where the maximal number of iterations is set to be 1000. In particular, to speed up the convergence, we use Matlab-provided ILU (Incomplete LU factorization) preconditioner: LUINC(A,drop-tolerance) in solving (3.3) since the ILU preconditioner is one of the most versatile preconditioners for unstructured matrices (cf. [9], [14]). We use $\mathbf{c}^k$ and right-hand side vector as the initial guesses for the approximate Jacobian equation (3.3) and systems (3.4)–(3.5) respectively. The inner loop stopping tolerance for (3.3) in the inexact Newton-type method is given by (4.2). While systems (3.4), (3.5) and (3.3) in the Newton-type method are all solved up to machine precision eps. Finally, for both algorithms, the outer iteration is stopped when

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

We now report our experimental results. Table 1 illustrates the values of $d_k := \|U_k^T A(\mathbf{c}^k) V_k - \Sigma^*\|_F$ for one of the test in case (a), where the approximate Jacobian matrix (3.3) is solved by the QMR method mentioned above. While the convergence performance averaged over the ten test problems for all cases are illustrated in Tables 2 and 3, where "T", "N_o", "N_i" represent respectively the averaged CPU time, the averaged numbers of outer iterations and the total ones of inner iterations required for solving (3.3). It should be noted that, in Table 2, the direct method (denoted by "DIRECT") and the QMR method (denoted by "QMR") are both adopted for solving (3.3) of the Newton-type method. Moreover, in Table 3, "I" means no preconditioner is used and "P" means the MILU-preconditioner is used in the QMR iterative method. We use "$+\infty$" in the numbers of inner iterations to denote that the methods fails to converge (see Table 3). Furthermore, since it was shown in Theorem 4.1 that the inexact Newton-type method converges with convergence rate β, we present the convergence performance of the inexact Newton-type method with different values of β in all tables. To further illustrate the over-solving problem, we give the convergence history of the inexact Newton-type (with $\beta = 1.8$) and the Newton-type methods in Figure 1 for one test in case (a), in which the logarithm of the error $\|U_k^T A(\mathbf{c}^k)V_k - \Sigma\|$ versus the number of inner iterations for solving the approximate Jacobian equation (3.3) are depicted (each point denotes an outer iteration).

We can see from Table 1 that both the Newton-type and inexact Newton-type methods converge quadraticallly/superlinearly and that, for the inexact Newton-type method, convergence performance of larger β are better than that of smaller β on the whole. While from Table 2, we can see that the CPU time and the outer iteration numbers of the inexact Newton-type method with $\beta \geq 1.6$ are fewer than that of the Newton-type method. As shown in Table 3 and Figure 1, in terms of N_i, the inexact Newton-type method is more effective than the Newton-type method and the over-solving problem of the Newton-type method is much significant than the inexact Newton-type method especially for $\beta = 1.8$. We also note that for the same method or the same β in the inexact Newton-type method, the MILU-preconditioner is quite effective in speeding up the solving of the approximate Jacobian equation.

Table 1: Convergence performance of Algorithms 1 and 2

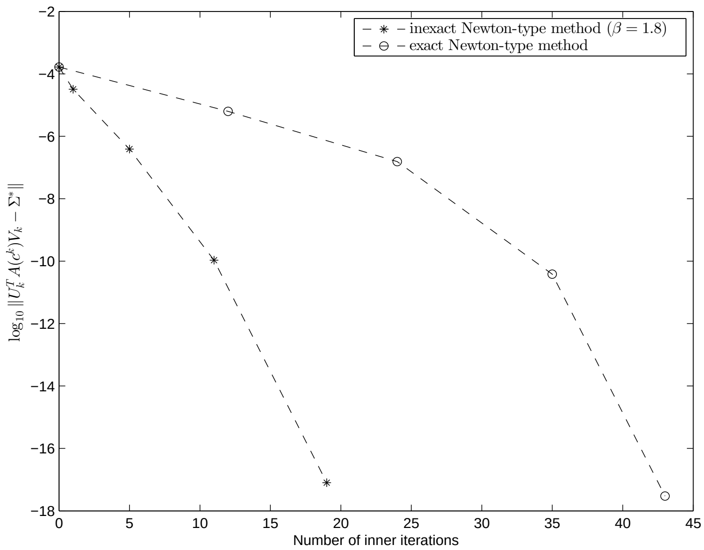
k	Algorithm 1	Algorithm2				
		$\beta = 1.2$	$\beta = 1.4$	$\beta = 1.6$	$\beta = 1.8$	$\beta = 2.0$
0	3.21e-04	3.21e-04	3.21e-04	3.21e-04	3.21e-04	3.21e-04
1	7.32e-06	8.67e-05	8.67e-05	8.23e-05	8.09e-05	7.73e-05
2	3.16e-08	5.92e-06	2.92e-06	2.58e-06	2.08e-06	1.93e-06
3	8.71e-12	1.83e-06	1.39e-08	2.91e-09	2.81e-09	2.77e-09
4	2.28e-16	8.96e-08	7.17e-11	1.51e-14	7.67e-15	4.37e-15
5	$\vdots$	1.28e-08	1.20e-16	$\vdots$	$\vdots$	$\vdots$
6	$\vdots$	1.07e-10	$\vdots$	$\vdots$	$\vdots$	$\vdots$
7	$\vdots$	1.10e-13	$\vdots$	$\vdots$	$\vdots$	$\vdots$
8	$\vdots$	5.29e-16	$\vdots$	$\vdots$	$\vdots$	$\vdots$
9	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table 2: Averaged total numbers of outer iterations N_o and CPU time T

		Algorithm1		Algorithm2				
		DIRECT	*QMR*	$\beta = 1.2$	$\beta = 1.4$	$\beta = 1.6$	$\beta = 1.8$	$\beta = 2.0$
(a)	N_o	4.40	4.40	7.40	5.40	4.60	4.40	4.40
	T	5.95	6.21	6.90	5.03	5.01	4.75	4.94
(b)	N_o	4.30	4.30	8.20	5.70	5.00	4.30	4.30
	T	$1.60e+01$	$1.49e+01$	$2.48e+01$	$1.70e+01$	$1.49e+01$	$1.13e+01$	$1.41e+01$
(c)	N_o	5.10	5.10	$1.04e+01$	7.50	5.50	5.10	5.10
	T	$4.66e+01$	$4.71e+01$	$9.32e+01$	$7.10e+01$	$5.03e+01$	$4.17e+1$	$4.38e+01$

Table 3: Averaged total numbers of inner iterations N_i

		Algorithm 1	Algorithm 2				
			$\beta = 1.2$	$\beta = 1.4$	$\beta = 1.6$	$\beta = 1.8$	$\beta = 2.0$
(a)	P	$4.69e+01$	$1.83e+01$	$1.62e+01$	$1.51e+01$	$1.46e+01$	$2.07e+01$
	I	$8.49e+02$	$7.08e+02$	$5.23e+02$	$4.76e+02$	$4.51e+02$	$4.91e+02$
(b)	P	$1.12e+02$	$5.43e+01$	$5.11e+01$	$4.80e+01$	$4.54e+01$	$5.53e+01$
	I	$4.04e+03$	$3.85e+03$	$3.20e+03$	$2.73e+03$	$2.48e+03$	$2.74e+03$
(c)	P	$1.09e+03$	$5.88e+02$	$5.34e+02$	$4.68e+02$	$4.62e+02$	$5.40e+02$
	I	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$



1: Convergence history of one test.

6 Conclusions

We proposed in this paper a Newton-type method (and its inexact version) for solving the ISVP (1.2) with multiple and/or zero singular values and established the corresponding quadratic/superlinear convergence results for the special case when $m = n$.

Recall that an exact/inexact Newton-type method was proposed in [27] for solving the ISVP (1.2) with the multiple but positive singular values, and its quadratic/superlinear convergence was claimed [27, Theorem 3.1] there. However, there is a fatal gap in the proof for Theorem 3.1 there. Indeed, the main idea for proving [27, Theorem 3.1] is to estimate $\|\mathbf{c}^{k+1} - \mathbf{c}^k\|$ to show that the generated sequence $\{\mathbf{c}^k\}$ is a Cauchy sequence, which used unavoidably the following equality:

$$\Sigma^* + X_k\Sigma^* - \Sigma^* Y_k = U_k^T A(\mathbf{c}^{k+1})V_k - D_k, \tag{6.1}$$

where $D_k := \mathrm{Diag}(r_1^k,\ r_2^k,\ \ldots,\ r_n^k) \in \mathbb{R}^{m\times n}$ and r_i^k $(1 \le i \le n)$ is the i-th component of the residual control vector $\mathbf{r}^k$ (see [27, Remark 3.1 and page 149, lines 20-21] for details). Unfortunately, equality (6.1) is not true in general if $s > 1$ even in the case when $m = n$ because, by the definitions of X_k and Y_k in the algorithm,

$$[\Sigma^* + X_k\Sigma^* - \Sigma^* Y_k]_{ij} = -[U_k^T A(\mathbf{c}^{k+1})V_k]_{ji} \neq [U_k^T A(\mathbf{c}^{k+1})V_k - D_k]_{ij}, \quad \text{for each} \quad 1 \le i < j \le s, \tag{6.2}$$

in general. Moreover, we remark that the technique used in [27] cannot be adopted to treat the distinct case with the zero singular values (i.e., (1.5) is satisfied with $s = 1$ and $t = 1$) as the same gap appears because

$$[\Sigma^* + X_k\Sigma^* - \Sigma^* Y_k]_{ij} = 0 \neq [U_k^T A(\mathbf{c}^{k+1})V_k - D_k]_{ij}, \quad \text{for each} \quad n - t + 1 \le i,\ j \le n. \tag{6.3}$$

Actually, we don't know whether [27, Theorem 3.1] is still true or not in the case when $m > n$.

The technique used in this paper for proving the convergence results is different from the one used in [27]. Indeed, we show the convergence of the sequence $\{\mathbf{c}^k\}$ in this paper for the case when $m = n$ via estimating directly $\|\mathbf{c}^k - \mathbf{c}^*\|$, rather than $\|\mathbf{c}^{k+1} - \mathbf{c}^k\|$, to avoid using the equality (6.1) but using the key Lemmas 2.5 and 2.7 to get the estimation for $\|\tilde{X}_k - X_k\|$. It seems that this technique does not work for the case when $m > n$ because, for each $n - t + 1 \le i,\ j \le m - t$, $|[\tilde{X}_k - X_k]_{ij}|$ can not be estimated as done for other components of $\tilde{X}_k - X_k$ (even though Lemmas 2.5 and 2.7 could be extended to the case when $m > n$).

Acknowledgments. We are grateful to the referees for their valuable comments which help us to improve the original presentation of the paper.

References

[1] Z. J. Bai, R. H. Chan, B. Morini, An inexact Cayley transform method for inverse eigenvalue problems, Inverse Problems 20 (2004) 1675–1689.

[2] Z. J. Bai, B. Morini, S. F. Xu, On the local convergence of an iterative approach for inverse singular value problems, J. Comput. Appl. Math. 198 (2007) 344–360.

[3] Z. J. Bai, S. F. Xu, An inexact Newton-type method for inverse singular value problems, Linear Algebra Appl. 429 (2008) 527–547.

[4] R. H. Chan, S. F. Xu, H. M. Zhou, On the convergence of a quasi-Newton method for inverse eigenvalue problem, SIAM J. Numer. Anal. 36 (1999) 436–441.

[5] M. T. Chu, Numerical methods for inverse singular value problems, SIAM J. Numer. Anal. 29 (1992) 885–903.

[6] M. T. Chu, G. H. Golub, Structured inverse eigenvalue problems, Acta Numer. 11 (2002) 1–71.

[7] M. T. Chu, G. H. Golub, Inverse Eigenvalue Problems: Theory, Algorithms, and Applications, Oxford University Press, Oxford, 2005.

[8] F. Clarke, Optimization and Nonsmooth Analysis, John Wiley and Sons, New York, 1983.

[9] T. Dupont, R.P. Kendall, H.H. JR. Rachford, An approximate factorization procedure for solving self-adjoint elliptic difference equations, SIAM J. Numer. Anal. 5 (1968) 559–573.

[10] R. W. Freund, N. M. Nachtigal, QMR: a quasi-minimal residual method for non-Hermitian linear systems, Numer. Math. 60 (1991) 315–339.

[11] S. Friedland, J. Nocedal, M. L. Overton, The formulation and analysis of numerical methods for inverse eigenvalue problems, SIAM J. Numer. Anal. 24 (1987) 634–667.

[12] G. H. Golub, C. F. Van Loan, Matrix Computations, 3rd ed., Johns Hopkins University Press, Baltimore, 1996.

[13] A.G. Gramm, Inverse Problems: Mathematical and Analytical Techniques with Applications to Engineering, Springer, 2005.

[14] I. Gustafsson, A class of first order factorizations, BIT 18 (1978) 142–156.

[15] L.P. Lebedev, I.I. Vorovich, G.M. Gladwell, Functional Analysis: Applications in Mechanics and Inverse Problems, Kluwer Academic Publishers, 2002.

[16] W. Ma, Z. J. Bai, A regularized directional derivative-based Newton method for inverse singular value problems, Inverse Problems 28 (2012) 125001.

[17] E. Montaño, M. Salas, R. L. Soto, Nonnegative matrices with prescribed extremal singular values, Comput. Math. Appl. 56 (2008) 30–42.

[18] E. Montaño, M. Salas, R. L. Soto, Positive matrices with prescribed singular values, Proyecciones (Antofagasta) 27 (2008) 289–305.

[19] T. Politi, A discrete approach for the inverse singular value problem in some quadratic group, Lecture Notes Comput. Sci. 2658 (2003) 121–130.

[20] F. A. Potra, L. Q. Qi, D. F. Sun, Secant methods for semismooth equations, Numer. Math. 80 (1998) 305–324.

[21] J. F. Queiró, An inverse problem for singular values and the Jacobian of the elementary symmetric functions, Linear Algebra Appl. 197–198 (1994) 277–282.

[22] L. Q. Qi, Convergence analysis of some algorithms for solving nonsmooth equations, Math. Oper. Res. 18 (1993) 227–244.

[23] C. S. Saunders, J. Hu, C. E. Christoffersen, M. B. Steer, Inverse singular value method for enforcing passivity in reduced-order models of distributed structures for transient and steady-state simulation, IEEE Trans. Microw. Theory Tech. 59 (2011) 837–847.

[24] W. P. Shen, Y. H. Hu, C. Li, J. C. Yao, Convergence of a Ulm-like method for square inverse singular value problems with multiple and zero singular values, Numer. Algor., 79 (2018) 375–398.

[25] W. P. Shen, C. Li, X. Q. Jin, An inexact Cayley transform method for inverse eigenvalue problems with multiple eigenvalues, Inverse Problems 31 (2015) 085007.

[26] W. P. Shen, C. Li, X. Q. Jin, A Ulm-like method for inverse eigenvalue problems, Appl. Numer. Math. 61 (2011) 356–367.

[27] W. P. Shen, C. Li, X. Q. Jin, J. C. Yao, Newton-type methods for inverse singular value problems with multiple singular values, Appl. Numer. Math. 109 (2016) 138–156.

[28] D. F. Sun, J. Sun, Strong semismoothness of eigenvalues of symmetric matrices and its application to inverse eigenvalue problems, SIAM J. Numer. Anal. 40 (2003) 2352–2367.

[29] J. A. Tropp, I. S. Dhillon, R.W. Heath, Finite-step algorithms for constructing optimal CDMA signature sequences, IEEE Trans. Inform. Theory 50 (2004) 2916–2921.

[30] G. Uhlmann, S. Levy (Eds.), Inside Out: Inverse Problems and Applications, Cambridge University Press, 2003.

[31] S. W. Vong, Z. J. Bai, X. Q. Jin, A Ulm-like method for inverse singular value problems, SIAM J. Matrix Anal. Appl. 32 (2011) 412–429.

[32] S. F. Xu, An Introduction to Inverse Eigenvalue Problems, Peking University Press, 1998.

[33] S. F. Yuan, A. P. Liao, G. Z. Yao, Parameterized inverse singular value problems for anti-bisymmetric matrices, Numer. Algorithms 60 (2012) 501–522.