# The exact condition number of the truncated singular value solution of a linear ill-posed problem 

El Houcine Bergou, Serge Gratton, Jean Tshimanga Ilunga

## To cite this version:

El Houcine Bergou, Serge Gratton, Jean Tshimanga Ilunga. The exact condition number of the truncated singular value solution of a linear ill-posed problem. SIAM Journal on Mathematical Analysis, 2014, 35 (3), pp.1073-1085. 10.1137/120869286 . hal-02147972

HAL Id: hal-02147972
https://hal.science/hal-02147972
Submitted on 5 Jun 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is a publisher's version published in:
http://oatao.univ-toulouse.fr/22597

## Official URL

DOI : https://doi.org/10.1137/120869286

To cite this version: Bergou, El Houcine and Gratton, Serge and Tshimanga Ilunga, Jean The exact condition number of the truncated singular value solution of a linear ill-posed problem. (2014) SIAM Journal on Mathematical Analysis, 35 (3). 1073-1085. ISSN 0036-1410

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

# THE EXACT CONDITION NUMBER OF THE TRUNCATED SINGULAR VALUE SOLUTION OF A LINEAR ILL-POSED PROBLEM* 

EL HOUCINE BERGOU ${ }^{\dagger}$, SERGE GRATTON ${ }^{\dagger}$, AND JEAN TSHIMANGA ${ }^{\dagger}$


#### Abstract

The main result of this paper is the formulation of an explicit expression for the condition number of the truncated least squares solution of $A x=b$. This expression is given in terms of the singular values of $A$ and the Fourier coefficients of $b$. The result is derived using the notion of the Fréchet derivative together with the product norm on the data $[A, b]$ and the 2 -norm on the solution. Numerical experiments are given to confirm our results by comparing them to those obtained by means of a finite difference approach.


Key words. truncated singular value decomposition, condition number, Fréchet derivative, least squares solution, perturbation theory

DOI. 10.1137/120869286

1. Introduction. Perturbation analysis is the study of the sensitivity of the solution of a given problem to perturbations in the data. The concept of condition number allows us to assess the sensitivity of the solution. Sensitivity and conditioning theory has been applied to many fundamental problems of linear algebra, such as linear systems, linear least squares, or eigenvalue problems [1, 2, 8, 4, 13]. In this paper, we investigate the condition number for the so-called truncated singular value decomposition (TSVD) solution to linear least squares problems. TSVD solutions arise in a wide variety of applications in science, technology, and engineering. In inverse problems, for example, the TSVD can be considered as a regularization technique for ill-conditioned matrices with well-determined numerical rank; see [5, 6, 14]. Applications of TSVD solutions in this area include computational tomography, image deblurring, digital signal processing, and geophysical inversion in seismology. Some numerical solutions of partial differential equations may also require techniques such as TSVD; see [11].

Let $A$ be an $n \times p$ matrix $(n \geq p)$ with $\operatorname{rank}(A)=r^{*} \leq p$ and let

$$
A=U \Sigma V^{T}
$$

be the full singular value decomposition of $A$ with singular values of $A$ arranged in descending order in $\Sigma$. Then, given an $n$-vector $b$, the least squares problem

$$
\min _{x \in \Re^{p}}\|A x-b\|_{2}
$$

has the minimum 2-norm solution $x^{*}=V_{r^{*}} \Sigma_{r^{*}}^{-1} U_{r^{*}}^{T}$, where $\Sigma_{r^{*}}$ is the diagonal matrix consisting of the first $r^{*}$ singular values of $A$ in descending order, and $U_{r^{*}}$ and $V_{r^{*}}$ are

[^0]http://www.siam.org/journals/simax/35-3/86928.html
${ }^{\dagger}$ CERFACS, 42, avenue Gaspard Coriolis, 31057 Toulouse, France (elhoucine.bergou@cerfacs.fr, serge.gratton@cerfacs.fr, jean.tshimanga@cerfacs.fr).
formed from the first $r^{*}$ columns of $U$ and $V$, respectively. In some applications (e.g., problems arising from the discretization of an ill-posed problem), a better solution, in the sense that it is less sensitive than the original one to errors in the data $(A, b)$, is obtained by a truncated least squares solution of the form
$$
x_{r}=V_{r} \Sigma_{r}^{-1} U_{r}^{T} b,
$$
for some $r<r^{*}$, and where $V_{r}, \Sigma_{r}$, and $U_{r}$ are defined as before but with $r$ replacing $r^{*}$. It turns out that if $\hat{U}_{r}$ and $\hat{V}_{r}$ are any orthonormal bases for range $\left(U_{r}\right)$ and range $\left(V_{r}\right)$, then
$$
x_{r}=\hat{V}_{r}\left(\hat{U}_{r}^{T} A \hat{V}_{r}\right)^{-1} \hat{U}_{r}^{T} b
$$

Now let $A$ and $b$ be perturbed to yield $\tilde{A}=A+E$ and $\tilde{b}=b+f$, and let $\tilde{U}_{r}$ and $\tilde{V}_{r}$ form a pair of bases for the left and right singular subspaces associated with the $r$ first singular values of $\tilde{A}$. The corresponding truncated least squares solution of the perturbed problem is then

$$
\begin{equation*}
\tilde{x}_{r}=\tilde{V}_{r}\left(\tilde{U}_{r}^{T} \tilde{A} \tilde{V}_{r}\right)^{-1} \tilde{U}_{r}^{T} b \tag{1.1}
\end{equation*}
$$

Now it turns out that if the Fréchet derivative, $x_{r}^{\prime}$, of the function $x_{r}$ exists, then we have

$$
\tilde{x}_{r}=x_{r}+x_{r}^{\prime} \cdot(E, f)+\mathrm{o}(\|(E, f)\|) .
$$

Here, $x_{r}^{\prime} .(E, f)$ is the application of a linear operator to $(E, f)$. Given a norm on $(E, f)$, call it $\|\cdot\|_{(\alpha, \beta)}$, the condition number of $x_{r}$ is defined to be the operator norm

$$
\left\|\mid x_{r}^{\prime}\right\| \|_{(\alpha, \beta), 2}=\max _{[\alpha E, \beta f] \neq 0} \frac{\left\|x_{r}^{\prime} \cdot(E, f)\right\|_{2}}{\|[E, f]\|_{(\alpha, \beta)}}
$$

The particular norm we use is defined by

$$
\|(E, f)\|_{(\alpha, \beta)}=\sqrt{\alpha^{2}\|E\|_{F}^{2}+\beta^{2}\|f\|_{F}^{2}}
$$

where $\|\cdot\|_{F}$ is the usual Frobenius norm and $\left.\alpha \in\right] 0,+\infty[, \beta \in] 0,+\infty[$. Note that the purpose of the norm $\|\cdot\|_{(\alpha, \beta)}$ is to tag the contributions of perturbations of $A$ and $b$ in the condition number; see [3].

The purpose of this paper is to exhibit the square of the condition number of $x_{r}$ as the 2 -norm of a symmetric nonnegative matrix $\Delta$ that can be formed from the singular values of $A$, and the Fourier coefficients given by the entries of $U^{T} b$. The paper is organized as follows. In section 2, we state preliminary results based on results from [12]. Section 3 is devoted to an expression for the first-order expansion of $x_{r}$ with respect to the data $(A, b)$. The main result of this section is the matrix representation for the corresponding Fréchet derivative leading to the formula for the condition number of $x_{r}$ using the singular values of $A$ and the Fourier coefficients of $b$. We perform some numerical tests to validate our analysis by comparing it with results produced by a finite difference approach in section 4. A brief conclusion is given in section 5 .
2. Preliminary results. It will be worthwhile to define the following matrix partitions:

$$
V=\left[V_{r}, V_{\perp}\right] \in \Re^{p \times p}, \quad U=\left[U_{r}, U_{\perp}\right] \in \Re^{n \times n}, \quad \Sigma=\left[\begin{array}{ll}
\Sigma_{r} & \\
& \Sigma_{\perp}
\end{array}\right] \in \Re^{n \times p}
$$

where

$$
\begin{aligned}
& V_{r} \in \Re^{p \times r}, \quad V_{\perp} \in \Re^{p \times(p-r)}, \quad \begin{array}{c}
U_{r} \in \Re^{n \times r}, \\
\Sigma_{r}=\operatorname{diag}\left(\sigma_{1}, \ldots, \sigma_{r}\right) \in \Re^{r \times r}, \quad \Sigma_{\perp}=\left[\begin{array}{c}
\operatorname{diag}\left(\sigma_{r+1}, \ldots, \sigma_{p}\right) \\
0
\end{array}\right] \in \Re^{n \times(n-r)},
\end{array} . \begin{array}{l}
(n-r) \times(p-r)
\end{array} .
\end{aligned}
$$

Furthermore, we define matrices $E_{r r}=U_{r}^{T} E V_{r}, E_{r \perp}=U_{r}^{T} E V_{\perp}, E_{\perp r}=U_{\perp}^{T} E V_{r}$, and $E_{\perp \perp}=U_{\perp}^{T} E V_{\perp}$ and vectors $b_{r}=U_{r}^{T} b, b_{\perp}=U_{\perp}^{T} b$, and $f_{r}=U_{r}^{T} f$. Finally, we shall denote by $I_{r}, I_{n-r}$ and $I_{p-r}$ the identity matrices of order $r, n-r$, and $p-r$, respectively.

The operator vec $(\cdot)$ and the Kronecker product $\otimes$ will be of particular importance in what follows. The vec $(\cdot)$ operator stacks the columns of the matrix argument into one long vector. For any matrices $B$ and $C$, the matrix $B \otimes C=\left(b_{i j} C\right)$. It is enough for our purpose to recall the following properties concerning these operators. ${ }^{1}$ For any matrices $B, X$, and $C$ having compatible dimensions with respect to the involved products, we have

$$
\begin{align*}
\operatorname{vec}(B X C) & =\left(C^{T} \otimes B\right) \operatorname{vec}(X)  \tag{2.1}\\
\operatorname{vec}\left(X^{T}\right) & =\Psi_{(n, p)} \operatorname{vec}(X) \text { for all } X \in \Re^{n \times p} \tag{2.2}
\end{align*}
$$

where $\Psi_{(n, p)} \in \Re^{n p \times n p}$ is the permutation matrix defined by

$$
\Psi_{(n, p)}=\sum_{i=1}^{n} \sum_{j=1}^{p} L_{i j} \otimes L_{i j}^{T}
$$

Here each $L_{i j} \in \Re^{n \times p}$ has entry 1 in position $(i, j)$ and all other entries are zero.
The following assumption will be of particular importance in what follows.
Assumption 2.1. Let

$$
\gamma=\left\|\left(E_{\perp r}^{T}, E_{r \perp}\right)\right\|_{F}
$$

suppose that

$$
\delta=\left|\sigma_{r}-\sigma_{r+1}\right|-\left\|E_{r r}\right\|_{2}-\left\|E_{\perp \perp}\right\|_{2}>0
$$

and assume that

$$
\gamma / \delta<1 / 2
$$

Roughly speaking, the statement of Assumption 2.1 is that the existence of a gap between $\sigma_{r}$ and $\sigma_{r+1}>0$ is required and that $\|E\|_{2}$ must be small enough compared to this gap.

[^1]Now, we state and adapt results from [12] to our context in the following two theorems.

Theorem 2.2 (see [12, Theorem 6.4]). Let an $n \times p$ perturbation matrix $E$ be given and partition $U^{T} E V$ with respect to $U=\left[U_{r}, U_{\perp}\right]$ and $V=\left[V_{r}, V_{\perp}\right]$ in the form

$$
U^{T} E V=\left(\begin{array}{cc}
E_{r r} & E_{r \perp} \\
E_{\perp r} & E_{\perp \perp}
\end{array}\right)
$$

Then under Assumption 2.1, there are matrices $Q \in \Re^{(n-r) \times r}$ and $P \in \Re^{(p-r) \times r}$ satisfying

$$
\left\|\left(Q^{T}, P^{T}\right)\right\|_{F}<2 \frac{\gamma}{\delta}<1
$$

such that range $\left(V_{r}+V_{\perp} P\right)$ and range $\left(U_{r}-U_{\perp} Q\right)$ form a pair of singular subspaces for $\tilde{A}=A+E$.

Among other things, the theorem above tells us that $Q$ and $P$ approach 0 as $E$ approaches 0 . Other useful results related to the ones above are given in the following theorem (see again [12] and [13, p. 266]).

Theorem 2.3. Suppose Assumption 2.1 holds. Then there exist matrices $Q \in$ $\Re^{(n-r) \times r}$ and $P \in \Re^{(p-r) \times r}$ such that
(2.3) $\quad \tilde{U}_{r}=\left(U_{r}-U_{\perp} Q\right)\left(I+Q^{T} Q\right)^{-1 / 2}, \quad \tilde{U}_{\perp}=\left(U_{r} Q^{T}+U_{\perp}\right)\left(I+Q Q^{T}\right)^{-1 / 2}$,
(2.4) $\tilde{V}_{r}=\left(V_{r}+V_{\perp} P\right)\left(I+P^{T} P\right)^{-1 / 2}, \quad \tilde{V}_{\perp}=\left(-V_{r} P^{T}+V_{\perp}\right)\left(I+P P^{T}\right)^{-1 / 2}$,
with $\tilde{U}_{r}^{T} \tilde{A} \tilde{V}_{\perp}=0$ and $\tilde{U}_{\perp}^{T} \tilde{A} \tilde{V}_{r}=0$. Furthermore, $\tilde{U}=\left[\tilde{U}_{r}, \tilde{U}_{\perp}\right] \in \Re^{n \times n}$ and $\tilde{V}=$ $\left[\tilde{V}_{r}, \tilde{V}_{\perp}\right] \in \Re^{p \times p}$ are orthogonal matrices.

Since the overall aim of this investigation is to derive the condition number as the norm of the Fréchet derivative of $x_{r}$, our intermediate goal will be to write a first-order expansion of (1.1) in terms of quantities in (2.3) and (2.4) and then replace $Q$ and $P$ with their respective first-order expansions with respect to $E$. The next theorem exploits (2.3) and (2.4) together with properties of singular decomposition to establish these expansions.

Theorem 2.4. Suppose that $\sigma_{r}-\sigma_{r+1}>0$. Then the first-order expansions for $Q$ and $P$ are given by

$$
\begin{align*}
\operatorname{vec}\left(\mathrm{Q}^{\mathrm{T}}\right)= & -\left(I_{n-r} \otimes \Sigma_{r}^{2}-\left(\Sigma_{\perp} \Sigma_{\perp}^{T}\right) \otimes I_{r}\right)^{-1}  \tag{2.5}\\
& \times\left[I_{n-r} \otimes \Sigma_{r}, \Sigma_{\perp} \otimes I_{r}\right]\left[\begin{array}{c}
\Psi_{(n-r, r)}\left(V_{r}^{T} \otimes U_{\perp}^{T}\right) \\
V_{\perp}^{T} \otimes U_{r}^{T}
\end{array}\right] \operatorname{vec}(\mathrm{E}) \\
& +\mathrm{o}(\|E\|), \\
\operatorname{vec}(\mathrm{P})= & \left(\Sigma_{r}^{2} \otimes I_{p-r}-I_{r} \otimes\left(\Sigma_{\perp}^{T} \Sigma_{\perp}\right)\right)^{-1} \\
& \times\left[I_{r} \otimes \Sigma_{\perp}^{T}, \Sigma_{r} \otimes I_{p-r}\right]\left[\begin{array}{c}
V_{r}^{T} \otimes U_{\perp}^{T} \\
\Psi_{(r, p-r)}\left(V_{\perp}^{T} \otimes U_{r}^{T}\right)
\end{array}\right] \operatorname{vec}(\mathrm{E}) \\
& +\mathrm{o}(\|E\|) .
\end{align*}
$$

Proof. In agreement with

$$
U^{T} A V=\left[\begin{array}{cc}
U_{r}^{T} A V_{r} & U_{r}^{T} A V_{\perp}  \tag{2.7}\\
U_{\perp}^{T} A V_{r} & U_{\perp}^{T} A V_{\perp}
\end{array}\right]=\left[\begin{array}{cc}
\Sigma_{r} & 0 \\
0 & \Sigma_{\perp}
\end{array}\right] \in \Re^{n \times p},
$$

together with the results of Theorem 2.3, we have

$$
\begin{align*}
U^{T} \tilde{A} V & =\left[\begin{array}{cc}
U_{r}^{T}(A+E) V_{r} & U_{r}^{T}(A+E) V_{\perp} \\
U_{\perp}^{T}(A+E) V_{r} & U_{\perp}^{T}(A+E) V_{\perp}
\end{array}\right] \\
& \stackrel{\text { def }}{=}\left[\begin{array}{cc}
\Sigma_{r}+E_{r r} & E_{r \perp} \\
E_{\perp r} & \Sigma_{\perp}+E_{\perp \perp}
\end{array}\right],  \tag{2.8}\\
\tilde{U}^{T} \tilde{A} \tilde{V} & =\left[\begin{array}{cc}
\tilde{U}_{r}^{T} \tilde{A} \tilde{V}_{r} & \tilde{U}_{r}^{T} \tilde{A} \tilde{V}_{\perp} \\
\tilde{U}_{\perp}^{T} \tilde{A} \tilde{V}_{r} & \tilde{U}_{\perp}^{T} \tilde{A} \tilde{V}_{\perp}
\end{array}\right]=\left[\begin{array}{cc}
\star & 0 \\
0 & \star
\end{array}\right] . \tag{2.9}
\end{align*}
$$

If we substitute (2.3)-(2.4) into the extradiagonal blocks of (2.9) (that are zero), we obtain

$$
\begin{aligned}
& -\left(Q U_{r}^{T} A V_{r}+Q U_{r}^{T} A V_{\perp} P+Q U_{r}^{T} E V_{r}+Q U_{r}^{T} E V_{\perp} P\right. \\
& \left.\quad-U_{\perp}^{T} A V_{r}-U_{\perp}^{T} A V_{\perp} P-U_{\perp}^{T} E V_{r}-U_{\perp}^{T} E V_{\perp} P\right)=0 \\
& -\left(U_{r}^{T} A V_{r} P^{T}-U_{r}^{T} A V_{\perp}+U_{r}^{T} E V_{r} P^{T}-U_{r}^{T} E V_{\perp}\right. \\
& \left.\quad+U_{\perp}^{T} A V_{r} P^{T}-Q^{T} U_{\perp}^{T} A V_{\perp}+Q^{T} U_{\perp}^{T} E V_{r} P^{T}-Q^{T} U_{\perp}^{T} E V_{\perp}\right)=0 .
\end{aligned}
$$

Furthermore, using relations (2.7) and (2.8) and after rearranging terms, we obtain (see also [12, equation (6.2)]) the pair of quadratic matrix equations

$$
\begin{align*}
& Q\left(\Sigma_{r}+E_{r r}\right)+\left(\Sigma_{\perp}+E_{\perp \perp}\right) P=-E_{\perp r}-Q E_{r \perp} P  \tag{2.10}\\
& P\left(\Sigma_{r}+E_{r r}^{T}\right)+\left(\Sigma_{\perp}^{T}+E_{\perp \perp}^{T}\right) Q=E_{r \perp}^{T}+P E_{\perp r}^{T} Q \tag{2.11}
\end{align*}
$$

where unknowns are $Q$ and $P$. We retain only first-order terms ${ }^{2}$ in $\|E\|$ in (2.10) and (2.11) leading to

$$
\begin{align*}
& Q \Sigma_{r}+\Sigma_{\perp} P=-E_{\perp r}+\mathrm{o}(\|E\|)  \tag{2.12}\\
& P \Sigma_{r}+\Sigma_{\perp}^{T} Q=E_{r \perp}^{T}+\mathrm{o}(\|E\|) \tag{2.13}
\end{align*}
$$

from which we obtain the system

$$
\begin{align*}
& Q=-\Sigma_{\perp} P \Sigma_{r}^{-1}-E_{\perp r} \Sigma_{r}^{-1}+\mathrm{o}(\|E\|),  \tag{2.14}\\
& P=-\Sigma_{\perp}^{T} Q \Sigma_{r}^{-1}+E_{r \perp}^{T} \Sigma_{r}^{-1}+\mathrm{o}(\|E\|) \tag{2.15}
\end{align*}
$$

by a postmultiplication of both (2.12) and (2.13) by $\Sigma_{r}$ (which exists because $\sigma_{1} \geq \cdots \geq \sigma_{r}>\sigma_{r+1} \geq 0$ ). Replacing $P$ in (2.14) by the right-hand side of (2.15) and conversely replacing $Q$ in (2.15) by the right-hand side of (2.14) we have

$$
\begin{align*}
& Q=-\Sigma_{\perp}\left(-\Sigma_{\perp}^{T} Q \Sigma_{r}^{-1}+E_{r \perp}^{T} \Sigma_{r}^{-1}\right) \Sigma_{r}^{-1}-E_{\perp r} \Sigma_{r}^{-1}+\mathrm{o}(\|E\|),  \tag{2.16}\\
& P=-\Sigma_{\perp}^{T}\left(-\Sigma_{\perp} P \Sigma_{r}^{-1}-E_{\perp r} \Sigma_{r}^{-1}\right) \Sigma_{r}^{-1}+E_{r \perp}^{T} \Sigma_{r}^{-1}+\mathrm{o}(\|E\|) . \tag{2.17}
\end{align*}
$$

Postmultiplying (2.16) and (2.17) by $\Sigma_{r}^{2}$ and rearranging terms yields

$$
\begin{align*}
\Sigma_{r}^{2} Q^{T}-Q^{T} \Sigma_{\perp} \Sigma_{\perp}^{T} & =-E_{r \perp} \Sigma_{\perp}^{T}-\Sigma_{r} E_{\perp r}^{T}+\mathrm{o}(\|E\|)  \tag{2.18}\\
P \Sigma_{r}^{2}-\Sigma_{\perp}^{T} \Sigma_{\perp} P & =\Sigma_{\perp}^{T} E_{\perp r}+E_{r \perp}^{T} \Sigma_{r}+\mathrm{o}(\|E\|)
\end{align*}
$$

[^2]According to property (2.1), (2.18) and (2.19) may be rewritten as

$$
\begin{aligned}
& \left(I_{n-r} \otimes \Sigma_{r}^{2}-\left(\Sigma_{\perp} \Sigma_{\perp}^{T}\right) \otimes I_{r}\right) \operatorname{vec}\left(Q^{T}\right) \\
& \quad=-\operatorname{vec}\left(E_{r \perp} \Sigma_{\perp}^{T}+\Sigma_{r} E_{\perp r}^{T}\right)+\mathrm{o}(\|E\|) \\
& \quad=-\left[I_{n-r} \otimes \Sigma_{r}, \Sigma_{\perp} \otimes I_{r}\right]\left[\begin{array}{l}
\operatorname{vec}\left(E_{\perp r}^{T}\right) \\
\operatorname{vec}\left(E_{r \perp}\right)
\end{array}\right]+\mathrm{o}(\|E\|), \\
& \left(\Sigma_{r}^{2} \otimes I_{p-r}-I_{r} \otimes\left(\Sigma_{\perp}^{T} \Sigma_{\perp}\right)\right) \operatorname{vec}(P) \\
& \quad=\operatorname{vec}\left(\Sigma_{\perp}^{T} E_{\perp r}+E_{r \perp}^{T} \Sigma_{r}\right)+\mathrm{o}(\|E\|) \\
& \quad=\left[I_{r} \otimes \Sigma_{\perp}^{T}, \Sigma_{r} \otimes I_{p-r}\right]\left[\begin{array}{c}
\operatorname{vec}\left(E_{\perp r}\right) \\
\operatorname{vec}\left(E_{r \perp}^{T}\right)
\end{array}\right]+\mathrm{o}(\|E\|)
\end{aligned}
$$

One can replace $\operatorname{vec}\left(E_{\perp r}^{T}\right)$ and $\operatorname{vec}\left(E_{r \perp}^{T}\right)$ by $\Psi(n-r, r) \operatorname{vec}\left(E_{\perp r}\right)$ and $\Psi(r, p-r) \operatorname{vec}\left(E_{r \perp}\right)$, respectively, based on property (2.2). Note that $\left(I_{n-r} \otimes \Sigma_{r}^{2}-\left(\Sigma_{\perp} \Sigma_{\perp}^{T}\right) \otimes I_{r}\right)$ and $\left(\Sigma_{r}^{2} \otimes I_{p-r}-I_{r} \otimes\left(\Sigma_{\perp}^{T} \Sigma_{\perp}\right)\right)$ are diagonal matrices of order $(n-r) r$ and $(p-r) r$, respectively. In addition, their diagonal entries are strictly positive since $\sigma_{r}>\sigma_{r+1}$. Hence, their inverses exist. To conclude the proof, observe that

$$
\begin{aligned}
\operatorname{vec}\left(E_{\perp r}\right) & =\left(V_{r}^{T} \otimes U_{\perp}^{T}\right) \operatorname{vec}(E), & & \operatorname{vec}\left(E_{r \perp}\right)=\left(V_{\perp}^{T} \otimes U_{r}^{T}\right) \operatorname{vec}(E), \\
\operatorname{vec}\left(E_{\perp \perp}\right) & =\left(V_{\perp}^{T} \otimes U_{\perp}^{T}\right) \operatorname{vec}(E), & & \operatorname{vec}\left(E_{r r}\right)=\left(V_{r}^{T} \otimes U_{r}^{T}\right) \operatorname{vec}(E) .
\end{aligned}
$$

In what follows, we use the results in Theorem 2.3 to introduce the first-order expansion for $x_{r}$ around $(A, b)$ in terms of the partitioned singular value decomposition matrices of $A$, the perturbation matrix $E$, the vector $b$, and the perturbation vector $f$.
3. The Fréchet derivative and the condition number of $x_{r}$. The continuity and the differentiability of $x_{r}$ rely on the fact that one supposes that there is a gap between $\sigma_{r}$ and $\sigma_{r+1}$, that is, $\sigma_{r}-\sigma_{r+1}>0$. Consider the following counterexample. Let

$$
A=\left(\begin{array}{ll}
1 & 0 \\
0 & 1
\end{array}\right), \quad E=\left(\begin{array}{cc}
\epsilon^{2} \sin \left(\frac{1}{\epsilon}\right) & 0 \\
0 & -\epsilon^{2} \sin \left(\frac{1}{\epsilon}\right)
\end{array}\right), \quad b=\binom{1}{1}, \quad f=\binom{0}{0} .
$$

We take $r=1$. Thus

$$
\tilde{x}_{r}= \begin{cases}\frac{1}{1+\epsilon^{2} \sin \left(\frac{1}{\epsilon}\right)} e_{1} & \text { if } \sin \left(\frac{1}{\epsilon}\right)>0, \\ \frac{1}{1-\epsilon^{2} \sin \left(\frac{1}{\epsilon}\right)} e_{2} & \text { if } \sin \left(\frac{1}{\epsilon}\right)<0, \\ x_{r} & \text { if } \sin \left(\frac{1}{\epsilon}\right)=0,\end{cases}
$$

where $e_{1}=(1,0)^{T}$ and $e_{2}=(0,1)^{T}$ are the canonical vectors of $\Re^{2}$. The above counterexample shows that the unit-vector of $\tilde{x}_{r}$ fluctuates between $e_{1}$ and $e_{2}$ as $\epsilon$ tends to 0 . In this case $x_{r}$ is not continuous, and a fortiori not differentiable, around $A$. We know from Theorem 2.3 that the singular values of $\tilde{A}$ are the disjoint union of the singular values of $\tilde{U}_{r}^{T} \tilde{A} \tilde{V}_{r}$ and those of $\tilde{U}_{\perp}^{T} \tilde{A} \tilde{V}_{\perp}$. To define $\tilde{x}_{r}$ by (1.1) it is required that the $r$ leading singular values of $\tilde{A}$ be those of $\tilde{U}_{r}^{T} \tilde{A} \tilde{V}_{r}$. This is achieved if $\sigma_{r}-\sigma_{r+1}>0$ and $E$, sufficiently small. ${ }^{3}$

[^3]Now, let us state the following lemma.
Lemma 3.1. Suppose $\sigma_{r}-\sigma_{r+1}>0$. Then the first-order expansion of $x_{r}$ can be written in the form

$$
\begin{align*}
\tilde{x}_{r}= & x_{r}+V\left[\begin{array}{c}
I_{r} \\
0
\end{array}\right] \Sigma_{r}^{-1} f_{r}-V\left[\begin{array}{c}
I_{r} \\
0
\end{array}\right] \Sigma_{r}^{-1} Q^{T} b_{\perp}  \tag{3.1}\\
& +V\left[\begin{array}{c}
0 \\
I_{p-r}
\end{array}\right] P \Sigma_{r}^{-1} b_{r}-V\left[\begin{array}{c}
I_{r} \\
0
\end{array}\right] \Sigma_{r}^{-1} E_{r r} \Sigma_{r}^{-1} b_{r}+\mathrm{o}(\|[E, f]\|) .
\end{align*}
$$

Proof. We insert (2.3) and (2.4) in expression (1.1) to yield

$$
\begin{aligned}
\tilde{x}_{r} & =\left(V_{r}+V_{\perp} P\right)\left(\left(U_{r}-U_{\perp} Q\right)^{T}(A+E)\left(V_{r}+V_{\perp} P\right)\right)^{-1}\left(U_{r}-U_{\perp} Q\right)^{T} \tilde{b} \\
& =\left(V_{r}+V_{\perp} P\right)\left(\Sigma_{r}^{-1}-\Sigma_{r}^{-1} U_{r}^{T} E V_{r} \Sigma_{r}^{-1}\right)\left(U_{r}-U_{\perp} Q\right)^{T} \tilde{b}+\mathrm{o}(\|[E, f]\|),
\end{aligned}
$$

where we used the following result concerning a perturbation of the inverse of a matrix $(F+G)^{-1}=F^{-1}-F^{-1} G F^{-1}+\mathrm{o}(\|G\|)$; see [13, p. 131]. Developing this equation and recalling that $E_{r r} \stackrel{\text { def }}{=} U_{r}^{T} E V_{r}$ gives, after rearranging terms,

$$
\begin{aligned}
\tilde{x}_{r}= & x_{r}+V_{r} \Sigma_{r}^{-1} U_{r}^{T} f-V_{r} \Sigma_{r}^{-1} Q^{T} U_{\perp}^{T} b+V_{\perp} P \Sigma_{r}^{-1} U_{r}^{T} b-V_{r} \Sigma_{r}^{-1} E_{r r} \Sigma_{r}^{-1} U_{r}^{T} b \\
& +\mathrm{o}(\|[E, f]\|) \\
= & x_{r}+V_{r} \Sigma_{r}^{-1} f_{r}-V_{r} \Sigma_{r}^{-1} Q^{T} b_{\perp}+V_{\perp} P \Sigma_{r}^{-1} b_{r}-V_{r} \Sigma_{r}^{-1} E_{r r} \Sigma_{r}^{-1} b_{r}+\mathrm{o}(\|[E, f]\|) .
\end{aligned}
$$

From the properties

$$
V V^{T}=I, \quad V^{T} V_{r}=\left[\begin{array}{c}
I_{r} \\
0
\end{array}\right] \quad \text { and } \quad V^{T} V_{\perp}=\left[\begin{array}{c}
0 \\
I_{p-r}
\end{array}\right],
$$

we have

$$
\begin{aligned}
\tilde{x}_{r}= & x_{r}+V V^{T} V_{r} \Sigma_{r}^{-1} f_{r}-V V^{T} V_{r} \Sigma_{r}^{-1} Q^{T} b_{\perp} \\
& +V V^{T} V_{\perp} P \Sigma_{r}^{-1} b_{r}-V V^{T} V_{r} \Sigma_{r}^{-1} E_{r r} \Sigma_{r}^{-1} b_{r}+\mathrm{o}(\|[E, f]\|),
\end{aligned}
$$

which implies (3.1).
Now, we are ready to give the expression of the matrix $x_{r}^{\prime}$ that represents the Fréchet derivative of $x_{r}$, with respect to the data $(A, b)$. The expression is given in terms of the singular value decomposition information of $A$ and the vector $b$. For that, we simply use results in Theorem 2.4 to eliminate $Q$ and $P$ from (3.1).

Proposition 3.2. Suppose that $\sigma_{r}-\sigma_{r+1}>0$. Then the application

$$
x_{r}:\left(\Re^{n \times p}, \Re^{n}\right) \longrightarrow \Re^{p}:(A, b) \longrightarrow x_{r}
$$

is a differentiable function of $(A, b)$. In addition, we have

$$
\tilde{x}_{r}=x_{r}+x_{r}^{\prime}\left[\begin{array}{c}
\alpha \operatorname{vec}(\mathrm{E}) \\
\beta f
\end{array}\right]+\mathrm{o}(\|[E, f]\|)
$$

with

$$
x_{r}^{\prime}=V\left[\frac{1}{\alpha} M, \frac{1}{\beta}\left(\begin{array}{ll}
\Sigma_{r}^{-1} &  \tag{3.2}\\
& 0
\end{array}\right)\right] W \in \Re^{n \times(n p+n)} .
$$

Here, $W$ is an orthogonal matrix defined by

$$
W=\left[\begin{array}{cc}
V_{r}^{T} \otimes U_{\perp}^{T} & 0 \\
V_{\perp}^{T} \otimes U_{r}^{T} & 0 \\
V_{r}^{T} \otimes U_{r}^{T} & 0 \\
V_{\perp}^{T} \otimes U_{\perp}^{T} & 0 \\
0 & U^{T}
\end{array}\right] \in \Re^{(n p+n) \times(n p+n)}
$$

and $M$ is the partitioned matrix given by

$$
M=\left[\begin{array}{cccc}
R_{r} & S_{r} & -T_{r} & 0 \\
R_{\perp} & S_{\perp} & 0 & 0
\end{array}\right] \in \Re^{p \times(n p)}
$$

with

$$
\begin{array}{ll}
\text { (3.3) } & R_{r}=\left(b_{\perp}^{T} \otimes \Sigma_{r}^{-1}\right)\left(I_{n-r} \otimes \Sigma_{r}^{2}-\left(\Sigma_{\perp} \Sigma_{\perp}^{T}\right) \otimes I_{r}\right)^{-1}\left(I_{n-r} \otimes \Sigma_{r}\right) \Psi_{(n-r, r)},  \tag{3.3}\\
\text { (3.4) } & S_{r}=\left(b_{\perp}^{T} \otimes \Sigma_{r}^{-1}\right)\left(I_{n-r} \otimes \Sigma_{r}^{2}-\left(\Sigma_{\perp} \Sigma_{\perp}^{T}\right) \otimes I_{r}\right)^{-1}\left(\Sigma_{\perp} \otimes I_{r}\right) \\
\text { (3.5) } & R_{\perp}=\left(\left(b_{r}^{T} \Sigma_{r}^{-1}\right) \otimes I_{p-r}\right)\left(\Sigma_{r}^{2} \otimes I_{p-r}-I_{r} \otimes\left(\Sigma_{\perp}^{T} \Sigma_{\perp}\right)\right)^{-1}\left(I_{r} \otimes \Sigma_{\perp}^{T}\right) \\
(3.6) & S_{\perp}=\left(\left(b_{r}^{T} \Sigma_{r}^{-1}\right) \otimes I_{p-r}\right)\left(\Sigma_{r}^{2} \otimes I_{p-r}-I_{r} \otimes\left(\Sigma_{\perp}^{T} \Sigma_{\perp}\right)\right)^{-1}\left(\Sigma_{r} \otimes I_{p-r}\right) \Psi_{(r, p-r)}, \\
(3.7) & T_{r}=\left(b_{r}^{T} \Sigma_{r}^{-1}\right) \otimes \Sigma_{r}^{-1} .
\end{array}
$$

The dimensions of these matrices are given in the following:

$$
R_{r}, S_{r} \in \Re^{r \times(n-r) r}, \quad R_{\perp}, S_{\perp} \in \Re^{(p-r) \times(n-r) r}, \quad \text { and } T_{r} \in \Re^{r \times r^{2}}
$$

Proof. Consider the quantities in (3.1). Using the properties of the vec operator applied to a vector, we obtain

$$
\left[\begin{array}{c}
I_{r} \\
0
\end{array}\right] \Sigma_{r}^{-1} E_{r r} \Sigma_{r}^{-1} b_{r}=\left[\begin{array}{c}
\left(b_{r}^{T} \Sigma_{r}^{-1}\right) \otimes \Sigma_{r}^{-1} \\
0
\end{array}\right] \operatorname{vec}\left(E_{r r}\right)=\left[\begin{array}{c}
T_{r} \\
0
\end{array}\right]\left(V_{r}^{T} \otimes U_{r}^{T}\right) \operatorname{vec}(E)
$$

Taking the expressions for $\operatorname{vec}\left(Q^{T}\right)$ and $\operatorname{vec}(P)$ given in (2.5) and (2.6), we have

$$
\begin{aligned}
{\left[\begin{array}{c}
I_{r} \\
0
\end{array}\right] \Sigma_{r}^{-1} Q^{T} b_{\perp} } & =\left[\begin{array}{c}
b_{\perp}^{T} \otimes \Sigma_{r}^{-1} \\
0
\end{array}\right] \operatorname{vec}\left(Q^{T}\right) \\
& =-\left[\begin{array}{cc}
R_{r} & S_{r} \\
0 & 0
\end{array}\right]\left[\begin{array}{c}
V_{r}^{T} \otimes U_{\perp}^{T} \\
V_{\perp}^{T} \otimes U_{r}^{T}
\end{array}\right] \operatorname{vec}(E)+\mathrm{o}(\|[E, f]\|), \\
{\left[\begin{array}{c}
0 \\
I_{p-r}
\end{array}\right] P \Sigma_{r}^{-1} b_{r} } & =\left[\begin{array}{c}
0 \\
\left(b_{r}^{T} \Sigma_{r}^{-1}\right) \otimes I_{p-r}
\end{array}\right] \operatorname{vec}(P) \\
& =\left[\begin{array}{cc}
0 & 0 \\
R_{\perp} & S_{\perp}
\end{array}\right]\left[\begin{array}{c}
V_{r}^{T} \otimes U_{1}^{T} \\
V_{\perp}^{T} \otimes U_{r}^{T}
\end{array}\right] \operatorname{vec}(E)+\mathrm{o}(\|[E, f]\|) .
\end{aligned}
$$

Injecting these quantities in (3.1) results in

$$
\begin{aligned}
\tilde{x}_{r}= & x_{r}+V\left[\begin{array}{c}
\Sigma_{r}^{-1} \\
0
\end{array}\right] U^{T} f+V\left[\begin{array}{cccc}
R_{r} & S_{r} & -T_{r} & 0 \\
R_{\perp} & S_{\perp} & 0 & 0
\end{array}\right]\left[\begin{array}{c}
V_{r}^{T} \otimes U_{\perp}^{T} \\
V_{\perp}^{T} \otimes U_{r}^{T} \\
V_{r}^{T} \otimes U_{r}^{T} \\
V_{\perp}^{T} \otimes U_{\perp}^{T}
\end{array}\right] \operatorname{vec}(E) \\
& +\mathrm{o}(\|[E, f]\|)
\end{aligned}
$$

from which the results are derived.
We can now establish the expression of the $x_{r}$ condition number. We know by definition that

$$
\left\|\left|x_{r}^{\prime}\right|\right\|_{(\alpha, \beta), 2}=\max _{[\alpha E, \beta f] \neq 0} \frac{\left\|x_{r}^{\prime} \cdot(E, f)\right\|_{2}}{\|\operatorname{vec}[E, f]\|_{(\alpha, \beta)}}
$$

Thus, from (3.2) we conclude that the exact condition number of $x_{r}$ is

$$
\left\|\left|x_{r}^{\prime}\right|\right\|_{(\alpha, \beta), 2}=\lambda_{\max }^{1 / 2}(\Delta)
$$

where

$$
\Delta \stackrel{\text { def }}{=} V^{T} x_{r}^{\prime}\left(x_{r}^{\prime}\right)^{T} V=\frac{1}{\alpha^{2}} M M^{T}+\frac{1}{\beta^{2}}\left(\begin{array}{cc}
\Sigma_{r}^{-2} & 0 \\
0 & 0
\end{array}\right) \in \Re^{p \times p} .
$$

It remains to show how $\Delta$ can be expressed with the singular values of $A$ and the Fourier coefficients given by the elements of $U^{T} b$.

Proposition 3.3. Assume that the singular values of the matrix $A$ are such that

$$
\sigma_{1} \geq \cdots \geq \sigma_{r}>\sigma_{r+1} \geq \cdots \geq \sigma_{p} \geq 0
$$

Then

$$
\Delta=\left(\begin{array}{cc}
\frac{1}{\alpha^{2}} \Delta_{r r}+\frac{1}{\beta^{2}} \Sigma_{r}^{-2} & \frac{1}{\alpha^{2}} \Gamma_{\perp r}^{T} \\
\frac{1}{\alpha^{2}} \Gamma_{\perp r} & \frac{1}{\alpha^{2}} \Delta_{\perp \perp}
\end{array}\right)
$$

where

$$
\begin{aligned}
& \Delta_{r r}=\operatorname{diag}\left(\sum_{\mathrm{k}=1}^{\mathrm{r}} \frac{\theta_{\mathrm{k}}^{2}}{\sigma_{\mathrm{k}}^{2} \sigma_{\mathrm{t}}^{2}}+\sum_{\mathrm{k}=\mathrm{r}+1}^{\mathrm{p}}\left(\pi_{\mathrm{k}}^{(\mathrm{t})}\right)^{2} \frac{\sigma_{\mathrm{k}}^{2}+\sigma_{\mathrm{t}}^{2}}{\sigma_{\mathrm{t}}^{2}} \theta_{\mathrm{k}}^{2}+\sum_{\mathrm{k}=\mathrm{p}+1}^{\mathrm{n}} \frac{\theta_{\mathrm{k}}^{2}}{\sigma_{\mathrm{t}}^{4}}\right), \quad 1 \leq \mathrm{t} \leq \mathrm{r}, \\
& \Delta_{\perp \perp}=\operatorname{diag}\left(\sum_{\mathrm{k}=1}^{\mathrm{r}}\left(\pi_{\mathrm{k}}^{(\mathrm{t})}\right)^{2} \frac{\sigma_{\mathrm{k}}^{2}+\sigma_{\mathrm{t}}^{2}}{\sigma_{\mathrm{k}}^{2}} \theta_{\mathrm{k}}^{2}\right), \\
& \Gamma_{\perp r}=R_{\perp} R_{r}^{T}+S_{\perp} S_{r}^{T} \\
& =2\left(\begin{array}{cccc}
\left(\pi_{r+1}^{(1)}\right)^{2} \frac{\sigma_{r+1}}{\sigma_{1}} \theta_{1} \theta_{r+1} & \left(\pi_{r+1}^{(2)}\right)^{2} \frac{\sigma_{r+1}}{\sigma_{2}} \theta_{2} \theta_{r+1} & \cdots & \left(\pi_{r+1}^{(r)}\right)^{2} \frac{\sigma_{r+1}}{\sigma_{r}} \theta_{r} \theta_{r+1} \\
\left(\pi_{r+2}^{(1)}\right)^{2} \frac{\sigma_{r+2}}{\sigma_{1}} \theta_{1} \theta_{r+2} & \left(\pi_{r+2}^{(2)}\right)^{2} \frac{\sigma_{r+2}}{\sigma_{2}} \theta_{2} \theta_{r+2} & \cdots & \left(\pi_{r+2}^{(r)}\right)^{2} \frac{\sigma_{r+2}}{\sigma_{r}} \theta_{r} \theta_{r+2} \\
\vdots & \vdots & \ddots & \vdots \\
\left(\pi_{p}^{(1)}\right)^{2} \frac{\sigma_{p}}{\sigma_{1}} \theta_{1} \theta_{p} & \left(\pi_{p}^{(2)}\right)^{2} \frac{\sigma_{p}}{\sigma_{2}} \theta_{2} \theta_{p} & \cdots & \left(\pi_{p}^{(r)}\right)^{2} \frac{\sigma_{p}}{\sigma_{r}} \theta_{r} \theta_{p}
\end{array}\right),
\end{aligned}
$$

with $\left(\theta_{1}, \ldots, \theta_{n}\right)=b^{T} U$, and $\pi_{k}^{(t)}=\frac{1}{\sigma_{t}^{2}-\sigma_{k}^{2}}$, with either $t=1, \ldots, r$ and $k=r+1, \ldots, p$ or $k=1, \ldots, r$ and $t=r+1, \ldots, p$.

Moreover, the quantity $\pi_{k}^{(t)}$ is well defined, since whenever it appears, $\sigma_{t}^{2}-\sigma_{k}^{2} \neq 0$ holds.

Proof. First we consider the $p \times p$ symmetric matrix

$$
M M^{T}=\left[\begin{array}{cc}
R_{r} R_{r}^{T}+S_{r} S_{r}^{T}+T_{r} T_{r}^{T} & -R_{r} R_{\perp}^{T}-S_{r} S_{\perp}^{T} \\
-R_{\perp} R_{r}^{T}-S_{\perp} S_{r}^{T} & R_{\perp} R_{\perp}^{T}+S_{\perp} S_{\perp}^{T}
\end{array}\right] \stackrel{\text { def }}{=}\left[\begin{array}{cc}
\Delta_{r r} & \Gamma_{\perp r} \\
\Gamma_{r \perp} & \Delta_{\perp \perp}
\end{array}\right] .
$$

Exploiting their structure, we can write the matrices (3.3)-(3.7) as

$$
\begin{align*}
R_{r}= & {\left[\theta_{r+1}\left(\Sigma_{r}^{2}-\sigma_{r+1}^{2} I_{r}\right)^{-1}, \ldots, \theta_{p}\left(\Sigma_{r}^{2}\right.\right.}  \tag{3.8}\\
& \left.\left.-\sigma_{p}^{2} I_{r}\right)^{-1}, \theta_{p+1} \Sigma_{r}^{-2}, \ldots, \theta_{n} \Sigma_{r}^{-2}\right] \Psi_{(n-r, r)}  \tag{3.9}\\
S_{r}= & {\left[\theta_{r+1} \sigma_{r+1} \Sigma_{r}^{-1}\left(\Sigma_{r}^{2}-\sigma_{r+1}^{2} I_{r}\right)^{-1}, \ldots, \theta_{p} \sigma_{p} \Sigma_{r}^{-1}\left(\Sigma_{r}^{2}-\sigma_{p}^{2} I_{r}\right)^{-1}, 0, \ldots, 0\right] }  \tag{3.10}\\
R_{\perp}= & {\left[\theta_{1} \sigma_{1}^{-1}\left(\sigma_{1}^{2} I_{p-r}-\Sigma_{\perp}^{T} \Sigma_{\perp}\right)^{-1} \Sigma_{\perp}^{T}, \ldots, \theta_{r} \sigma_{r}^{-1}\left(\sigma_{r}^{2} I_{p-r}-\Sigma_{\perp}^{T} \Sigma_{\perp}\right)^{-1} \Sigma_{\perp}^{T}\right] } \\
S_{\perp}= & {\left[\theta_{1}\left(\sigma_{1}^{2} I_{p-r}-\Sigma_{\perp}^{T} \Sigma_{\perp}\right)^{-1}, \ldots, \theta_{r}\left(\sigma_{r}^{2} I_{p-r}-\Sigma_{\perp}^{T} \Sigma_{\perp}\right)^{-1}\right] \Psi_{(r, p-r)}, } \\
T_{r}= & {\left[\theta_{1} \sigma_{1}^{-1} \Sigma_{r}^{-1}, \ldots, \theta_{r} \sigma_{r}^{-1} \Sigma_{r}^{-1}\right] . }
\end{align*}
$$

In (3.8), the first of the two factors,

$$
\begin{equation*}
\left[\theta_{r+1}\left(\Sigma_{r}^{2}-\sigma_{r+1}^{2} I_{r}\right)^{-1}, \ldots, \theta_{p}\left(\Sigma_{r}^{2}-\sigma_{p}^{2} I_{r}\right)^{-1}, \theta_{p+1} \Sigma_{r}^{-2}, \ldots, \theta_{n} \Sigma_{r}^{-2}\right] \tag{3.14}
\end{equation*}
$$

is a $1 \times(n-r)$ partitioned matrix. Its blocks consist of $r$-order diagonal matrices. Recall that the second factor in (3.8) is

$$
\begin{equation*}
\Psi_{(n-r, r)}=\sum_{i=1}^{n-r} \sum_{j=1}^{r} L_{i j} \otimes L_{i j}^{T}, \tag{3.15}
\end{equation*}
$$

where $L_{i j}=e_{i} e_{j}^{T} \in \Re^{(n-r) r}$ with $e_{i} \in \Re^{(n-r)}$ and $e_{j} \in \Re^{r}$. Observe that $L_{i j} \otimes L_{i j}^{T}$ is an $(n-r) \times r$ partitioned matrix where each block has $r$ rows and $n-r$ columns. Furthermore, it has the block $L_{i j}^{T}$ in position $i, j$ and 0 in the remaining blocks. The multiplication of the partitioned matrices (3.14) and (3.15) results in the $1 \times r$ partitioned matrix

$$
\begin{aligned}
R_{r}= & \sum_{i=1}^{n-r} \sum_{j=1}^{r}\left[\theta_{r+1}\left(\Sigma_{r}^{2}-\sigma_{r+1}^{2} I_{r}\right)^{-1}, \ldots, \theta_{p}\left(\Sigma_{r}^{2}\right.\right. \\
& \left.\left.-\sigma_{p}^{2} I_{r}\right)^{-1}, \theta_{p+1} \Sigma_{r}^{-2}, \ldots, \theta_{n} \Sigma_{r}^{-2}\right] L_{i j} \otimes L_{i j}^{T},
\end{aligned}
$$

whose block $j$ can be written as

$$
\sum_{i=1}^{p-r} \theta_{r+i}\left(\Sigma_{r}^{2}-\sigma_{r+i}^{2} I_{r}\right)^{-1} L_{i j}^{T}+\sum_{i=p-r+1}^{n-r} \theta_{r+i} \Sigma_{r}^{-2} L_{i j}^{T}
$$

Consequently, multiplying $R_{\perp}$ and $R_{r}$ block by block yields

$$
\begin{align*}
R_{\perp} R_{r}^{T}= & \sum_{j=1}^{r} \theta_{j} \sigma_{j}^{-1}\left(\sigma_{j}^{2} I_{p-r}-\Sigma_{\perp}^{T} \Sigma_{\perp}\right)^{-1} \Sigma_{\perp}^{T} \sum_{i=1}^{p-r} L_{i j} \theta_{r+i}\left(\Sigma_{r}^{2}-\sigma_{r+i} I_{r}\right)^{-1}  \tag{3.16}\\
& +\sum_{j=1}^{r} \theta_{j} \sigma_{j}^{-1}\left(\sigma_{j}^{2} I_{p-r}-\Sigma_{\perp}^{T} \Sigma_{\perp}\right)^{-1} \Sigma_{\perp}^{T} \sum_{i=p-r+1}^{n-r} L_{i j} \theta_{r+i} \Sigma_{r}^{-2}
\end{align*}
$$

Since $\Sigma_{\perp}^{T} \sum_{i=p-r+1}^{n-r} e_{i}=0$, one has $\Sigma_{\perp}^{T} \sum_{i=p-r+1}^{n-r} L_{i j}=\Sigma_{\perp}^{T} \sum_{i=p-r+1}^{n-r} e_{i} e j^{T}=0$ and hence the last term in (3.16) vanishes. Thus

$$
R_{\perp} R_{r}^{T}=\sum_{i=1}^{p-r} \sum_{j=1}^{r} \theta_{r+i} \theta_{j} \sigma_{j}^{-1}\left(\sigma_{j}^{2} I_{p-r}-\Sigma_{\perp}^{T} \Sigma_{\perp}\right)^{-1} \Sigma_{\perp}^{T} e_{i} e_{j}^{T}\left(\Sigma_{r}^{2}-\sigma_{r+i}^{2} I_{r}\right)^{-1}
$$

A direct computation gives

$$
\theta_{r+i} \theta_{j} \sigma_{j}^{-1}\left(\sigma_{j}^{2} I_{p-r}-\Sigma_{\perp}^{T} \Sigma_{\perp}\right)^{-1} \Sigma_{\perp}^{T} e_{i}= \begin{cases}\frac{\theta_{r+i} \theta_{j} \sigma_{r+i}}{\sigma_{j}\left(\sigma_{j}^{2}-\sigma_{r+i}^{2}\right)} e_{i}, & i=1, \ldots, p-r \\ 0, & i=p-r+1, \ldots, n-r\end{cases}
$$

where $e_{i} \in \Re^{(n-r)}$ in the left-hand side and $e_{i} \in \Re^{(p-r)}$ on the right-hand side. Then from

$$
e_{j}^{T}\left(\Sigma_{r}^{2}-\sigma_{r+i}^{2} I_{r}\right)^{-1}=\frac{1}{\left(\sigma_{j}^{2}-\sigma_{r+i}^{2}\right)} e_{j}^{T},
$$

where $e_{j} \in \Re^{r}$ on both sides, we deduce that

$$
\begin{aligned}
R_{\perp} R_{r}^{T} & =\sum_{i=1}^{p-r} \sum_{j=1}^{r} \frac{1}{\left(\sigma_{j}^{2}-\sigma_{r+i}^{2}\right)^{2}} \frac{\sigma_{r+i}}{\sigma_{j}} \theta_{r+i} \theta_{j} e_{i} e_{j}^{T} \\
& =\sum_{i=1}^{p-r} \sum_{j=1}^{r}\left(\pi_{r+i}^{(j)}\right)^{2} \frac{\sigma_{r+i}}{\sigma_{j}} \theta_{r+i} \theta_{j} e_{i} e_{j}^{T} \in \Re^{(p-r) \times r}
\end{aligned}
$$

with $\pi_{r+i}^{(j)}=\frac{1}{\left(\sigma_{j}^{2}-\sigma_{r+i}^{2}\right)}$. In the same manner we can compute and show that $S_{\perp} S_{r}^{T}$ is equivalent to $R_{\perp} R_{r}^{T}$.

The remaining blocks in $M M^{T}$ are computed by performing the block matrixmatrix multiplications. So,

$$
\begin{aligned}
R_{r} R_{r}^{T}=\sum_{k=r+1}^{p} \theta_{k}^{2}\left(\Sigma_{r}^{2}-\sigma_{k}^{2} I_{r}\right)^{-2}+\sum_{k=p+1}^{n} \theta_{k}^{2} \Sigma_{r}^{-4} & =\operatorname{diag}\left(\sum_{k=r+1}^{p}\left(\pi_{k}^{(t)}\right)^{2} \theta_{k}^{2}+\sum_{k=r+1}^{n} \frac{\theta_{k}^{2}}{\sigma_{t}^{4}}\right) \\
S_{r} S_{r}^{T}=\sum_{k=r+1}^{p} \theta_{k}^{2} \sigma_{k}^{2} \Sigma_{r}^{-2}\left(\Sigma_{r}^{2}-\sigma_{k}^{2} I_{r}\right)^{-2} & =\operatorname{diag}\left(\sum_{k=r+1}^{p}\left(\pi_{k}^{(t)}\right)^{2} \frac{\sigma_{k}^{2}}{\sigma_{t}^{2}} \theta_{k}^{2}\right) \\
T_{r} T_{r}^{T}=\sum_{k=1}^{r} \frac{\theta_{k}^{2}}{\sigma_{k}^{2}} \Sigma_{r}^{-2} & =\operatorname{diag}\left(\sum_{k=1}^{r} \frac{\theta_{k}^{2}}{\sigma_{k}^{2} \sigma_{t}^{2}}\right)
\end{aligned}
$$

$$
\text { for } t=1, \ldots, r
$$

$$
\begin{array}{ll}
R_{\perp} R_{\perp}^{T}=\sum_{k=1}^{r} \frac{\theta_{k}^{2}}{\sigma_{k}^{2}} \Sigma_{\perp}^{T} \Sigma_{\perp}\left(\sigma_{k}^{2} I_{p-r}-\Sigma_{\perp}^{T} \Sigma_{\perp}\right)^{-2} & =\operatorname{diag}\left(\sum_{k=1}^{r}\left(\pi_{k}^{(t)}\right)^{2} \frac{\sigma_{t}^{2}}{\sigma_{k}^{2}} \theta_{k}^{2}\right) \\
S_{\perp} S_{\perp}^{T}=\sum_{k=1}^{r} \theta_{k}^{2}\left(\sigma_{k}^{2} I_{p-r}-\Sigma_{\perp}^{T} \Sigma_{\perp}\right)^{-2} & =\operatorname{diag}\left(\sum_{k=1}^{r}\left(\pi_{k}^{(t)}\right)^{2} \theta_{k}^{2}\right)
\end{array}
$$

for $t=1, \ldots, p-r$.
Putting the above results together yields the result.

Table 1
The exact value of cond $\left(x_{r}\right)$ using the expression in Proposition 3.3 versus the finite difference estimate value using jacobianest for 12 problems.

| Problem | cond $\left(x_{r}\right)$ <br> from 3.3 | Finite difference estimate <br> value of cond $\left(x_{r}\right)$ | $n$ | $p$ | $r$ |
| :--- | :--- | :--- | :--- | :--- | :--- |
| baart | $7.156 \mathrm{e}+3$ | $7.087 \mathrm{e}+3$ | 20 | 20 | 5 |
| blur | $2.516 \mathrm{e}+1$ | $2.516 \mathrm{e}+1$ | 16 | 16 | 6 |
| derive | $1.698 \mathrm{e}+3$ | $1.698 \mathrm{e}+3$ | 12 | 12 | 10 |
| foxgood | $2.896 \mathrm{e}+1$ | $2.896 \mathrm{e}+1$ | 20 | 20 | 2 |
| heat | $4.486 \mathrm{e}+1$ | $4.478 \mathrm{e}+1$ | 12 | 12 | 10 |
| i_laplace | $1.448 \mathrm{e}+4$ | $1.367 \mathrm{e}+4$ | 20 | 20 | 7 |
| parallax | $1.412 \mathrm{e}+5$ | $1.411 \mathrm{e}+5$ | 26 | 12 | 10 |
| phillips | $5.731 \mathrm{e}+1$ | $5.731 \mathrm{e}+1$ | 12 | 12 | 10 |
| shaw | $1.044 \mathrm{e}+3$ | $1.044 \mathrm{e}+3$ | 12 | 12 | 8 |
| spikes | $8.178 \mathrm{e}+2$ | $8.178 \mathrm{e}+2$ | 12 | 12 | 4 |
| ursell | $3.716 \mathrm{e}+5$ | $3.716 \mathrm{e}+5$ | 20 | 20 | 3 |
| wing | $3.429 \mathrm{e}+6$ | $3.010 \mathrm{e}+6$ | 20 | 20 | 5 |

Let us point out the fact that an early result in [3], when $r=p$, that is, when we do not perform truncation (i.e., we assume that $A$ is a full rank matrix), is a particular case of the results above. In fact, in this case, $\Delta$ becomes diagonal and simplifies to

$$
\Delta_{r r}=\operatorname{diag}\left(\sum_{k=1}^{p} \frac{\theta_{k}^{2}}{\sigma_{k}^{2} \sigma_{t}^{2}}+\sum_{k=p+1}^{n} \frac{\theta_{k}^{2}}{\sigma_{t}^{4}}\right)=\operatorname{diag}\left(\frac{1}{\sigma_{t}^{2}}\left(\sum_{k=1}^{p} \frac{\theta_{k}^{2}}{\sigma_{k}^{2}}+\sum_{k=p+1}^{n} \frac{\theta_{k}^{2}}{\sigma_{t}^{2}}\right)\right)
$$

for $t=1, \ldots, p$. This implies the result given in [3], that is,

$$
\begin{aligned}
\left\|x_{r}^{\prime}\right\|_{(\alpha, \beta), 2} & =\sqrt{\frac{1}{\alpha^{2}} \frac{1}{\sigma_{\min }^{2}}\left[\sum_{k=1}^{p}\left(\frac{\theta_{k}}{\sigma_{k}}\right)^{2}+\frac{1}{\sigma_{\min }^{2}} \sum_{k=p+1}^{n} \theta_{k}^{2}\right]+\frac{1}{\beta^{2}} \frac{1}{\sigma_{\min }^{2}}} \\
& =\left\|A^{\dagger}\right\| \sqrt{\frac{1}{\alpha^{2}}\left(\|x\|^{2}+\left\|A^{\dagger}\right\|^{2}\|r\|^{2}\right)+\frac{1}{\beta^{2}}}
\end{aligned}
$$

where $A^{\dagger}$ denotes the Moore-Penrose inverse (see [9, p. 421]) of $A$, and $x$ denotes the solution of the linear least squares problem associated with $A$ and $b$.

Looking at the general result of Proposition 3.3, we see that the quantities (scalars) involved in the computation of the $x_{r}$ condition number are nothing but the singular values $\sigma_{k}$ of $A$ and the components $\theta_{k}$ of $b$ along singular vectors $u_{k}$. Finally, observe that the critical gap is $\sigma_{r}-\sigma_{r+1}$.
4. Numerical experiments. We now describe some numerical tests carried out in MATLAB. Our test cases come from the package Regularization Tools ${ }^{4}$ by Hansen [7]. We arbitrarily choose values of $n, p$, and $r$. By means of a specific routine in this package, we generate pairs $(A, b)$ associated with some test problems indicated by their name. To validate the expression of the exact condition number, we use the numerical derivative code ${ }^{5}$ by D'Errico, called jacpbianest.m, to estimate uses a centered finite differences approach with Romberg extrapolation to improve

[^4]the estimates to sixth order. For our purpose, we recast $x_{r}$ as $f: z=\operatorname{vec}([A, b]) \rightarrow x_{r}$ prior to the use of jacobianest.m, and then we approximate the condition number of $x_{r}$ as the 2 -norm of the estimated Jacobian. Note that in all tests we set $\alpha=\beta=1$.

Table 1 displays the exact condition number versus an estimate of the condition number produced with jacobianest.m. The results show how the derived expression of the exact condition compares well with the finite difference estimate.
5. Conclusion. We solved the problem of the determination of a closed formula for the condition number of the truncated singular value solution of an ill-posed problem which relies on a singular value decomposition of the problem. We anticipate that the presented formula will therefore stimulate research in several directions. Finding good estimates of the condition number using iterative techniques would, for instance, be of crucial relevance for large-scale problems. From a theoretical point of view, we also believe that the condition number may bring new insight into the problem of the detection of the truncation index of the singular value decomposition. One of the topics of future research will be to explore this issue for practical problems.

Acknowledgment. The authors thank the anonymous referees for the valuable comments and suggestions.

## REFERENCES

[1] Å. Björck, Numerical Methods for Least Squares Problems, SIAM, Philadelphia, 1996.
[2] G. H. Golub and C. F. Van Loan, Matrix Computations, 3rd ed., Johns Hopkins University Press, Baltimore, MD, 1996.
[3] S. Gratton, On the condition number of linear least squares problems in Frobenius norm, BIT, 36 (1995), pp. 523-530.
[4] S. Gratton, D. Titley-Peloquin, and J. Tshimanga, Sensitivity and conditioning of the truncated total least squares solution, SIAM J. Matrix. Anal. Appl., 34 (2013), pp. 1257-1276.
[5] P. C. HANSEN, The truncated SVD as a method for regularization, BIT, 27 (1987), pp. 534-553.
[6] P. C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion, SIAM, Philadelphia, 1998.
[7] P. C. HANSEN, Regularization tools: A MATLAB package for analysis and solution of discrete ill-posed problems, Numer. Algorithms, 46 (2007), pp. 187-194.
[8] N. J. Higham, Accuracy and Stability of Numerical Algorithms, 2nd ed., SIAM, Philadelphia, 2002.
[9] R. A. Horn and C. R. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, UK, 1990.
[10] R. A. Horn and C. R. Johnson, Topics in Matrix Analysis, Cambridge University Press, Cambridge, UK, 1991.
[11] Z.-C. Li, H.-T. Huang, and Y. Wei, Ill-conditioning of the truncated singular value decomposition, Tikhonov regularization and their applications to numerical partial differential equations, Numer. Linear Algebra Appl., 18 (2011), pp. 205-221.
[12] G. W. Stewart, Error and perturbation bounds for subspaces associated with certain eigenvalue problems, SIAM Rev., 15 (1973), pp. 727-764.
[13] G. W. Stewart and J.-G. Sun, Matrix Perturbation Theory, Academic Press, New York, 1990.
[14] C. R. Vogel, Computational Methods for Inverse Problems, Frontiers in Appl. Math., SIAM, Philadelphia, 2002.


[^0]:    *This research was supported by the "Assimilation de Données pour la Terre, l'Atmosphère et l'Oc éa n (ADTAO)" project, funded by the Fondation "Sciences et Technologies pour l'Aéoro-nautique et l 'Espace (STAE)", Toulouse, France, within the "Réseau Thématique de Recherche Avancée (RTRA) ".

[^1]:    ${ }^{1}$ We refer to [10, Chapter 4] for further properties of these operators.

[^2]:    ${ }^{2}$ This is why the terms $P E_{r r}^{T}, E_{\perp \perp}^{T} Q, P E_{\perp r}^{T} Q, Q E_{r r}, E_{\perp \perp} P$, and $Q E_{r \perp} P$ no longer appear.

[^3]:    ${ }^{3}$ Observe that in the presence of a gap $\sigma_{r}-\sigma_{r+1}>0$, the bases of the involved singular subspaces of $\tilde{A}$ tend continuously to those of $A$ as $E$ tends 0 .

[^4]:    ${ }^{4}$ See http://www2.imm.dtu.dk/ ~pch/Regutools.
    ${ }^{5}$ See http://www.mathworks.com/matlabcentral/fileexchang/13490-automatic-numericaldifferentiation.

