

Small singular values can increase in lower precision

Christos Boutsikas
Purdue University
cboutsik@purdue.edu

Petros Drineas
Purdue University
pdrineas@purdue.edu

Ilse C.F. Ipsen
North Carolina State University
ipsen@ncsu.edu

Abstract

We perturb a real matrix A of full column rank, and derive lower bounds for the smallest singular values of the perturbed matrix, for two classes of perturbations: deterministic normwise absolute, and probabilistic componentwise relative. Both classes of bounds, which extend existing lower-order expressions, demonstrate a potential increase in the smallest singular values. Our perturbation results represent a qualitative model for the increase in the small singular values after a matrix has been demoted to a lower arithmetic precision. Numerical experiments confirm the qualitative validity of the model and its ability to predict singular values changes in the presence of decreased arithmetic precision.

1 Introduction

Given a real, full column-rank matrix \mathbf{A} , we present two types of lower bounds for the smallest singular values of a perturbed matrix $\mathbf{A} + \mathbf{E}$: deterministic normwise absolute bounds, and probabilistic componentwise relative bounds.

1.1 Motivation

We investigate the change in the computed singular values of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $\text{rank}(\mathbf{A}) = n$ when \mathbf{A} is demoted to a lower arithmetic precision.

We have observed that the demotion to lower precision can improve the conditioning of \mathbf{A} by significantly increasing the computed small singular values, while leaving large singular values mostly unharmed. For instance, if the smallest singular value of \mathbf{A} is on the order of double precision roundoff,

$$\sigma_{\min}(\text{double}(\mathbf{A})) \approx 10^{-16},$$

then demotion of \mathbf{A} to single precision can increase the smallest singular value to single precision roundoff,

$$\sigma_{\min}(\text{double}(\text{single}(\mathbf{A}))) \approx 10^{-8}.$$

This phenomenon has been observed before, as the following quotes illustrate:

*... small singular values tend to increase [SS90, page 266]
... even an approximate inverse of an arbitrarily ill-conditioned
matrix does, in general, contain useful information [Rum09, page 260]
This is due to a kind of regularization by rounding to working precision [Rum09, page 261]*

1.2 Modelling demotion to lower precision in terms of perturbations

We model the demotion of a matrix to lower precision with the help of two classes of perturbations: deterministic normwise absolute, and probabilistic componentwise relative.

Deterministic normwise absolute perturbations The accumulated error from typical singular value algorithms in Matlab and Julia is a normwise absolute error [GV13, section 8.6].

Thus, if we demote a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m \geq n$ to single precision and compute the singular values of the demoted matrix, the resulting error can be represented as an absolute perturbation \mathbf{E} . According to Weyl's inequality [GV13, Corollary 8.6.2], corresponding singular values¹ of \mathbf{A} change by at most $\|\mathbf{E}\|_2$,

$$|\sigma_j(\mathbf{A} + \mathbf{E}) - \sigma_j(\mathbf{A})| \leq \|\mathbf{E}\|_2 \approx 10^{-8}, \quad 1 \leq j \leq n.$$

The bound implies that singular values larger than single precision roundoff, i.e. $\sigma_j(\mathbf{A}) \gg \|\mathbf{E}\|_2$, remain essentially the same,

$$\sigma_j(\mathbf{A}) \approx \sigma_j(\mathbf{A}) - \underbrace{\|\mathbf{E}\|_2}_{10^{-8}} \leq \sigma_j(\mathbf{A} + \mathbf{E}) \leq \sigma_j(\mathbf{A}) + \underbrace{\|\mathbf{E}\|_2}_{10^{-8}} \approx \sigma_j(\mathbf{A}),$$

while it is inconclusive about small singular values on the order of double precision roundoff,

$$\underbrace{\sigma_\ell(\mathbf{A})}_{10^{-16}} - \underbrace{\|\mathbf{E}\|_2}_{10^{-8}} \leq \sigma_\ell(\mathbf{A} + \mathbf{E}) \leq \underbrace{\sigma_\ell(\mathbf{A})}_{10^{-16}} + \underbrace{\|\mathbf{E}\|_2}_{10^{-8}}.$$

Probabilistic componentwise relative perturbations For high relative accuracy implementations, like Jacobi's method [DV07a; DV07b] and QR-preconditioned QR SVD methods [Drm17], we model demotion to lower precision in terms of componentwise relative perturbations with independent uniform random variables in $(-1, 1)$.

1.3 Our contributions

Our two main results are lower bounds on the smallest singular value cluster of a perturbed matrix: a deterministic normwise absolute bound, summarized in Theorem 1; and a probabilistic componentwise relative expression, summarized in Theorem 2. Both results show a definitive increase in the perturbed small singular values. The numerical experiments in section 5 confirm the qualitative validity of these results. Our assumptions are not restrictive and merely require the smallest singular value cluster to be separated by a small gap from the remaining singular values.

The deterministic bound in Theorem 1 improves the second-order perturbation expansions in [Ste84; Ste06], [SS90, Section V.4.2], because it is a true lower bound and it needs no assumptions on singular vectors.

Theorem 1. *Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m \geq n$ have $\text{rank}(\mathbf{A}) \geq n - r$ for some $r \geq 1$. Let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ be a full singular value decomposition, where $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ is diagonal, and $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V} \in \mathbb{R}^{n \times n}$ are orthogonal matrices. Partition commensurately,*

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{U}^T \mathbf{E} \mathbf{V} = \begin{bmatrix} \mathbf{E}_{11} & \mathbf{E}_{12} \\ \mathbf{E}_{21} & \mathbf{E}_{22} \\ \mathbf{E}_{31} & \mathbf{E}_{32} \end{bmatrix},$$

where $\mathbf{\Sigma}_1, \mathbf{E}_{11} \in \mathbb{R}^{(n-r) \times (n-r)}$ with $\mathbf{\Sigma}_1$ nonsingular diagonal; and $\mathbf{\Sigma}_2, \mathbf{E}_{22} \in \mathbb{R}^{r \times r}$ with $\mathbf{\Sigma}_2$ diagonal.

If $1/\|\mathbf{\Sigma}_1^{-1}\|_2 > 4\|\mathbf{E}\|_2$ and $\|\mathbf{\Sigma}_2\|_2 < \|\mathbf{E}\|_2$, then

$$\sigma_{n-r+j}(\mathbf{A} + \mathbf{E})^2 \geq \sigma_j(\mathbf{E}_{32}^T \mathbf{E}_{32} + (\mathbf{\Sigma}_2 + \mathbf{E}_{22})^T (\mathbf{\Sigma}_2 + \mathbf{E}_{22}) - \mathbf{R}_3)^2 - r_4, \quad 1 \leq j \leq r,$$

where \mathbf{R}_3 contains terms of order 3

$$\mathbf{R}_3 \equiv \mathbf{E}_{12}^T \mathbf{W} + \mathbf{W}^T \mathbf{E}_{12} \quad \mathbf{W} \equiv (\mathbf{\Sigma}_1 + \mathbf{E}_{11})^{-T} \begin{bmatrix} \mathbf{E}_{21}^T & \mathbf{E}_{31}^T \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_2 + \mathbf{E}_{22} \\ \mathbf{E}_{32} \end{bmatrix}$$

and r_4 contains terms of order 4 and higher,

$$r_4 \equiv \|\mathbf{W}\|_2^2 + 4 \frac{\|\mathbf{E}\|_2^2 \|(\mathbf{\Sigma}_1 + \mathbf{E}_{11})^{-1} (\mathbf{E}_{12} + \mathbf{W})\|_2^2}{1 - 4\|\mathbf{E}\|_2^2 \|(\mathbf{\Sigma}_1 + \mathbf{E}_{11})^{-1}\|_2^2}.$$

¹The singular values of a matrix \mathbf{A} are labelled in non-increasing order, $\sigma_1(\mathbf{A}) \geq \sigma_2(\mathbf{A}) \geq \dots$

Proof. This is a restatement of Assumptions 3.1 and Theorem 8. For the special case $r = 1$, this reduces to Assumptions 2.1 and Theorem 5. \square

For componentwise relative perturbations, we consider a matrix with two singular value clusters: a large one and a small one, under the assumption that the perturbation is sufficiently small to preserve the average large singular value. The resulting expression below for the average perturbed small singular values suggests a definitive increase over the original small singular values.

Theorem 2. Let $\mathbf{A}, \mathbf{F} \in \mathbb{R}^{m \times n}$ where $m \geq n$, and the elements of \mathbf{F} are independent random variables with

$$f_{ij} = \epsilon_{lower} a_{ij} \omega_{ij}, \quad \omega_{ij} \in \mathcal{U}(-1, 1), \quad 1 \leq i \leq m, \quad 1 \leq j \leq n.$$

Assume that \mathbf{A} has a cluster of large and a cluster of small singular values,

$$\sigma_r(\mathbf{A}) > \sigma_{r+1}(\mathbf{A}), \quad \mathbb{E} \left[\sum_{j=1}^r \sigma_j(\mathbf{A} + \mathbf{F})^2 \right] = \sum_{j=1}^r \sigma_j(\mathbf{A})^2,$$

and define the cluster averages of the small singular values as

$$\begin{aligned} \sigma_{small}(\mathbf{A})^2 &\equiv \frac{1}{n-r} \sum_{j=r+1}^n \sigma_j(\mathbf{A})^2, \\ \sigma_{small}(\mathbf{A} + \mathbf{F})^2 &\equiv \frac{1}{n-r} \sum_{j=r+1}^n \sigma_j(\mathbf{A} + \mathbf{F})^2. \end{aligned}$$

Then

$$\mathbb{E}[\sigma_{small}(\mathbf{A} + \mathbf{F})^2] = \sigma_{small}(\mathbf{A})^2 + \frac{\epsilon_{lower}^2}{3(n-r)} \|\mathbf{A}\|_F^2.$$

Proof. This is a restatement of Theorem 9 and Corollary 10. \square

Future work, sketched in section 6, will refine the above results towards a quantitative analysis that predicts the order of magnitude of the increase, and the influential matrix properties, in particular, the role of the singular value gap.

1.4 Existing work

There are many deterministic bounds for the smallest singular value of general, unstructured matrices. The bounds for nonsingular matrices $\mathbf{A} \in \mathbb{C}^{n \times n}$ in [LX21; Shu22; Zou12] involve the factor $|\det(\mathbf{A})|^2 \left(\frac{n-1}{\|\mathbf{A}\|_F^2} \right)^{n-1}$, while the ones in [HP92; YG97] contain factors like $|\det(\mathbf{A})|^2 \left(\frac{n-1}{n} \right)^{(n-1)/2}$ and row and column norms. The Schur complement-based bounds for strictly diagonally dominant matrices in [Hua08; Li20; San21; Var75] depend on the degree of diagonal dominance, as do the Gerschgorin type bounds for rectangular matrices in [Joh89; JS98].

In contrast, we are bounding the smallest singular values of *perturbed* matrices. The expressions for small singular values in [Ste84, Theorem], [Ste06, Theorem 8], [SS90, Section V.4.2] are second-order perturbation expansions rather than bounds, and require assumptions on the singular vectors.

There is an abundance of probabilistic approaches to estimating small singular values, including: probabilistic bounds [Dem88], and probabilistic analysis [Cuc16] of condition numbers; expansions and expected values of norms for stochastic perturbation theory [Arm10; LN20; Ste90]; statistical condition estimation [KLR98]; smallest singular values of random matrices [Coo18]; smoothed analysis [SST06; TV07; TV09; TV10]; and mixed precision computations [CD22].

In contrast, our straightforward expression for the expectation of small singular values applies to all matrices and requires a minimum of assumptions.

1.5 Overview

Our deterministic and probabilistic lower bounds for small singular values of $\mathbf{A} + \mathbf{E}$ are based on eigenvalue bounds for $(\mathbf{A} + \mathbf{E})^T(\mathbf{A} + \mathbf{E})$. We present deterministic normwise absolute bounds for a single smallest singular value (section 2) and for a cluster of small singular values (section 3); and a probabilistic expression for a cluster of small singular value (section 4). The numerical experiments (section 5) confirm the qualitative increase in small singular values resulting from the demotion of the matrix to lower precision. A brief discussion of future work (section 6) concludes the paper.

2 A single smallest singular value

We perturb a matrix that has a single smallest singular value, and derive a lower bound for the smallest singular value of the perturbed matrix in terms of normwise absolute perturbations (Section 2.2), based on eigenvalue bounds (Section 2.1).

2.1 Auxiliary eigenvalue results

We square the singular values of $\mathbf{A} \in \mathbb{R}^{m \times n}$ and consider instead the eigenvalues of the symmetric positive semi-definite matrix $\mathbf{B} \equiv \mathbf{A}^T \mathbf{A} \in \mathbb{R}^{n \times n}$.

For a symmetric positive semi-definite matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$ with a single smallest eigenvalue $\lambda_{\min}(\mathbf{B})$, we present two expressions for $\lambda_{\min}(\mathbf{B})$ with different assumptions (Lemmas 1 and 2), and two lower bounds in terms of normwise absolute perturbations (Theorems 3 and 4).

We assume that $\lambda_{\min}(\mathbf{B})$ is separated from the remaining eigenvalues, in the sense that it is strictly smaller than the smallest eigenvalue of the leading principal submatrix \mathbf{B}_{11} of order $n - 1$. The equality below expresses $\lambda_{\min}(\mathbf{B})$ in terms of itself.

Lemma 1 (Exact expression). *Let $\mathbf{B} \in \mathbb{R}^{n \times n}$ be symmetric positive semi-definite with $\text{rank}(\mathbf{B}) \geq n - 1$, and partition*

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{b} \\ \mathbf{b}^T & \beta \end{bmatrix} \quad \text{where } \mathbf{B}_{11} \in \mathbb{R}^{(n-1) \times (n-1)}.$$

Then

$$(1) \quad 0 \leq \lambda_{\min}(\mathbf{B}) \leq \beta.$$

If also $\lambda_{\min}(\mathbf{B}) < \lambda_{\min}(\mathbf{B}_{11})$ then

$$\lambda_{\min}(\mathbf{B}) = \beta - \mathbf{b}^T (\mathbf{B}_{11} - \lambda_{\min}(\mathbf{B}) \mathbf{I})^{-1} \mathbf{b},$$

Proof. Abbreviate $\tilde{\lambda}_{\min} \equiv \lambda_{\min}(\mathbf{B})$. The positive semi-definiteness of \mathbf{B} implies the lower bound in (1), while the variational inequalities imply the upper bound,

$$0 \leq \tilde{\lambda}_{\min} = \min_{\|\mathbf{x}\|_2=1} \mathbf{x}^T \mathbf{B} \mathbf{x} \leq \mathbf{e}_n^T \mathbf{B} \mathbf{e}_n = \beta.$$

To show the expression for $\tilde{\lambda}_{\min}$, observe that the shifted matrix

$$\mathbf{B} - \tilde{\lambda}_{\min} \mathbf{I} = \begin{bmatrix} \mathbf{B}_{11} - \tilde{\lambda}_{\min} \mathbf{I} & \mathbf{b} \\ \mathbf{b}^T & \beta - \tilde{\lambda}_{\min} \end{bmatrix}$$

is singular. From the assumption $\tilde{\lambda}_{\min} < \lambda_{\min}(\mathbf{B}_{11})$ follows that $\mathbf{B}_{11} - \tilde{\lambda}_{\min} \mathbf{I}$ is nonsingular. So we can determine the block LU decomposition $\mathbf{B} - \tilde{\lambda}_{\min} \mathbf{I} = \mathbf{L} \hat{\mathbf{U}}$ with

$$\mathbf{L} \equiv \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{b}^T (\mathbf{B}_{11} - \tilde{\lambda}_{\min} \mathbf{I})^{-1} & 1 \end{bmatrix},$$

$$\hat{\mathbf{U}} \equiv \begin{bmatrix} \mathbf{B}_{11} - \tilde{\lambda}_{\min} \mathbf{I} & \mathbf{b} \\ \mathbf{0} & \beta - \tilde{\lambda}_{\min} - \mathbf{b}^T (\mathbf{B}_{11} - \tilde{\lambda}_{\min} \mathbf{I})^{-1} \mathbf{b} \end{bmatrix}.$$

Since $\mathbf{B} - \tilde{\lambda}_{\min} \mathbf{I}$ is singular and the unit triangular matrix \mathbf{L} is nonsingular, the block upper triangular matrix $\hat{\mathbf{U}}$ has no choice but to be singular. Its leading principal submatrix $\mathbf{B}_{11} - \tilde{\lambda}_{\min} \mathbf{I}$ is nonsingular by assumption, which leaves the (2,2) element to be singular, but it being a scalar implies

$$\beta - \tilde{\lambda}_{\min} - \mathbf{b}^T (\mathbf{B}_{11} - \tilde{\lambda}_{\min} \mathbf{I})^{-1} \mathbf{b} = 0.$$

This gives the expression for $\tilde{\lambda}_{\min}$. □

If $\mathbf{b} = \mathbf{0}$ then Lemma 1 correctly asserts that $\lambda_{\min}(\mathbf{B}) = \beta$.

Lemma 2 below presents the same expression for $\lambda_{\min}(\mathbf{B})$ as in Lemma 1, but under a stronger albeit more useful assumption.

Lemma 2 (Exact expression with stronger assumption). *Let $\mathbf{B} \in \mathbb{R}^{n \times n}$ be symmetric positive semi-definite with $\text{rank}(\mathbf{B}) \geq n - 1$, and partition*

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{b} \\ \mathbf{b}^T & \beta \end{bmatrix} \quad \text{where } \mathbf{B}_{11} \in \mathbb{R}^{(n-1) \times (n-1)}.$$

If $\beta < \lambda_{\min}(\mathbf{B}_{11})$ then

$$\lambda_{\min}(\mathbf{B}) = \beta - \mathbf{b}^T (\mathbf{B}_{11} - \lambda_{\min}(\mathbf{B}) \mathbf{I})^{-1} \mathbf{b} \geq 0.$$

Proof. The upper bound (1) combined with the assumption $\beta < \lambda_{\min}(\mathbf{B}_{11})$ implies the assumption in Lemma 1,

$$(2) \quad 0 \leq \lambda_{\min}(\mathbf{B}) \leq \beta < \lambda_{\min}(\mathbf{B}_{11}).$$

□

The subsequent lower bounds for $\lambda_{\min}(\mathbf{B})$ are informative if the offdiagonal part has small norm.

Theorem 3 (First lower bound). *Let $\mathbf{B} \in \mathbb{R}^{n \times n}$ be symmetric positive semi-definite with $\text{rank}(\mathbf{B}) \geq n - 1$, and partition*

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{b} \\ \mathbf{b}^T & \beta \end{bmatrix} \quad \text{where } \mathbf{B}_{11} \in \mathbb{R}^{(n-1) \times (n-1)}.$$

If $\beta < \lambda_{\min}(\mathbf{B}_{11})$ then

$$\lambda_{\min}(\mathbf{B}) \geq \beta - \mathbf{b}^T \mathbf{B}_{11}^{-1} \mathbf{b} - \frac{\beta \|\mathbf{B}_{11}^{-1} \mathbf{b}\|_2^2}{1 - \beta \|\mathbf{B}_{11}^{-1}\|_2}.$$

Proof. Abbreviate $\tilde{\lambda}_{\min} \equiv \lambda_{\min}(\mathbf{B})$. From (2) follows that \mathbf{B}_{11} and $\mathbf{B}_{11} - \tilde{\lambda}_{\min} \mathbf{I}$ are nonsingular. Combined with the symmetric positive semi-definiteness of \mathbf{B}_{11} this gives

$$\tilde{\lambda}_{\min} < \lambda_{\min}(\mathbf{B}_{11}) = 1/\|\mathbf{B}_{11}^{-1}\|_2,$$

hence

$$(3) \quad \|\tilde{\lambda}_{\min} \mathbf{B}_{11}^{-1}\|_2 < 1.$$

Thus we can apply the Sherman-Morrison formula [GV13, Section 2.1.4],

$$(\mathbf{B}_{11} - \tilde{\lambda}_{\min} \mathbf{I})^{-1} = \mathbf{B}_{11}^{-1} + \tilde{\lambda}_{\min} \mathbf{B}_{11}^{-1} (\mathbf{I} - \tilde{\lambda}_{\min} \mathbf{B}_{11}^{-1})^{-1} \mathbf{B}_{11}^{-1},$$

and substitute the above into the expression for $\tilde{\lambda}_{\min}$ from Lemma 1,

$$(4) \quad \tilde{\lambda}_{\min} = \beta - \mathbf{b}^T \mathbf{B}_{11}^{-1} \mathbf{b} - \tilde{\lambda}_{\min} \mathbf{b}^T \mathbf{B}_{11}^{-1} (\mathbf{I} - \tilde{\lambda}_{\min} \mathbf{B}_{11}^{-1})^{-1} \mathbf{B}_{11}^{-1} \mathbf{b}.$$

The symmetric positive semi-definiteness of \mathbf{B} implies that $\beta \geq 0$ and $\mathbf{b}^T \mathbf{B}_{11}^{-1} \mathbf{b} \geq 0$, hence it remains to bound the norm of the remaining summand. From the symmetry of \mathbf{B}_{11} and the invariance of the two-norm under transposition follows

$$(5) \quad \begin{aligned} \|\mathbf{b}^T \mathbf{B}_{11}^{-1} (\mathbf{I} - \tilde{\lambda}_{\min} \mathbf{B}_{11}^{-1})^{-1} \mathbf{B}_{11}^{-1} \mathbf{b}\|_2 &\leq \|\mathbf{b}^T \mathbf{B}_{11}^{-1}\|_2 \|(\mathbf{I} - \tilde{\lambda}_{\min} \mathbf{B}_{11}^{-1})^{-1}\|_2 \|\mathbf{B}_{11}^{-1} \mathbf{b}\|_2 \\ &\leq \|\mathbf{B}_{11}^{-1} \mathbf{b}\|_2^2 \|(\mathbf{I} - \tilde{\lambda}_{\min} \mathbf{B}_{11}^{-1})^{-1}\|_2. \end{aligned}$$

The inequality (3) allows us to apply the Banach lemma [GV13, Lemma 2.3.3] to bound the norm of the inverse by

$$\|(\mathbf{I} - \tilde{\lambda}_{\min} \mathbf{B}_{11}^{-1})^{-1}\|_2 \leq \frac{1}{1 - \|\tilde{\lambda}_{\min} \mathbf{B}_{11}^{-1}\|_2} = \frac{1}{1 - \tilde{\lambda}_{\min} \|\mathbf{B}_{11}^{-1}\|_2}.$$

Substitute this into (5) and the resulting bound into the expression for $\tilde{\lambda}_{\min}$ in (4),

$$\tilde{\lambda}_{\min} \geq \beta - \mathbf{b}^T \mathbf{B}_{11}^{-1} \mathbf{b} - \frac{\tilde{\lambda}_{\min} \|\mathbf{B}_{11}^{-1} \mathbf{b}\|_2^2}{1 - \tilde{\lambda}_{\min} \|\mathbf{B}_{11}^{-1}\|_2},$$

and at last apply the upper bound (1). □

The lower bound in Theorem 3 is positive if $\|\mathbf{b}\|_2$ is sufficiently small, in which case $\lambda_{\min}(\mathbf{B}) \geq \beta - \mathcal{O}(\|\mathbf{b}\|_2^2)$. If $\mathbf{b} = \mathbf{0}$ then (1) and Theorem 3 imply $\lambda_{\min}(\mathbf{B}) = \beta$.

The slightly weaker bound below focusses on a ‘dominant part’ of \mathbf{B}_{11} .

Theorem 4 (Second lower bound). *Let $\mathbf{B} \in \mathbb{R}^{n \times n}$ be symmetric positive semi-definite with $\text{rank}(\mathbf{B}) \geq n-1$, and partition*

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{b} \\ \mathbf{b}^T & \beta \end{bmatrix} \quad \text{where } \mathbf{B}_{11} \in \mathbb{R}^{(n-1) \times (n-1)}.$$

If $\mathbf{B}_{11} = \mathbf{C}_{11} + \mathbf{C}_{12}$ where \mathbf{C}_{11} is symmetric positive definite with $\lambda_{\min}(\mathbf{C}_{11}) > \beta$, and \mathbf{C}_{12} is symmetric positive semi-definite then

$$\lambda_{\min}(\mathbf{B}) \geq \beta - \mathbf{b}^T \mathbf{C}_{11}^{-1} \mathbf{b} - \frac{\beta \|\mathbf{C}_{11}^{-1} \mathbf{b}\|_2^2}{1 - \beta \|\mathbf{C}_{11}^{-1}\|_2}.$$

Proof. Abbreviate $\tilde{\lambda}_{\min} \equiv \lambda_{\min}(\mathbf{B})$. From (1) and the assumption follows $\tilde{\lambda}_{\min} \leq \beta < \lambda_{\min}(\mathbf{C}_{11})$, hence \mathbf{C}_{11} and $\mathbf{C}_{11} - \tilde{\lambda}_{\min} \mathbf{I}$ are nonsingular. Write

$$\mathbf{B}_{11} - \tilde{\lambda}_{\min} \mathbf{I} = \underbrace{\mathbf{C}_{11} - \tilde{\lambda}_{\min} \mathbf{I}}_{\mathbf{G}} + \mathbf{C}_{12} = \mathbf{G}^{1/2} (\mathbf{I} + \underbrace{\mathbf{G}^{-1/2} \mathbf{C}_{12} \mathbf{G}^{-1/2}}_{\mathbf{H}}) \mathbf{G}^{1/2},$$

where \mathbf{G} is symmetric positive definite and \mathbf{H} is symmetric positive semi-definite. The Loewner ordering implies $\mathbf{I} \preceq \mathbf{I} + \mathbf{H}$. From [HJ13, Corollary 7.7.4] follows $(\mathbf{I} + \mathbf{H})^{-1} \preceq \mathbf{I}^{-1} = \mathbf{I}$. Thus

$$\begin{aligned} (\mathbf{B}_{11} - \tilde{\lambda}_{\min} \mathbf{I})^{-1} &= \mathbf{G}^{-1/2} (\mathbf{I} + \mathbf{H})^{-1} \mathbf{G}^{-1/2} \\ &\preceq \mathbf{G}^{-1/2} \mathbf{I} \mathbf{G}^{-1/2} = \mathbf{G}^{-1} = (\mathbf{C}_{11} - \tilde{\lambda}_{\min} \mathbf{I})^{-1}. \end{aligned}$$

Substituting $(\mathbf{B}_{11} - \tilde{\lambda}_{\min} \mathbf{I})^{-1} \preceq (\mathbf{C}_{11} - \tilde{\lambda}_{\min} \mathbf{I})^{-1}$ into the expression for $\tilde{\lambda}_{\min}$ in Lemma 1 gives

$$\tilde{\lambda}_{\min} \geq \beta - \mathbf{b}^T (\mathbf{C}_{11} - \tilde{\lambda}_{\min} \mathbf{I})^{-1} \mathbf{b}.$$

We continue as in the proof of Theorem 3 with the Sherman-Morrison formula [GV13, Section 2.1.4],

$$\begin{aligned} \tilde{\lambda}_{\min} &\geq \beta - \mathbf{b}^T \mathbf{C}_{11}^{-1} \mathbf{b} - \tilde{\lambda}_{\min} \mathbf{b}^T \mathbf{C}_{11}^{-1} (\mathbf{I} - \tilde{\lambda}_{\min} \mathbf{C}_{11}^{-1})^{-1} \mathbf{C}_{11}^{-1} \mathbf{b} \\ &\geq \beta - \mathbf{b}^T \mathbf{C}_{11}^{-1} \mathbf{b} - \frac{\tilde{\lambda}_{\min} \|\mathbf{C}_{11}^{-1} \mathbf{b}\|_2^2}{1 - \tilde{\lambda}_{\min} \|\mathbf{C}_{11}^{-1}\|_2}, \end{aligned}$$

and at last apply (1). □

If $\mathbf{C}_{12} = \mathbf{0}$, then Theorem 4 reduces to Theorem 3.

2.2 A lower bound for the smallest singular value

We consider a matrix with a distinct smallest singular value. Based on the eigenvalue bounds in section 2.1, we derive a lower bound for the smallest singular value of a perturbed matrix (Theorem 5) in terms of normwise absolute perturbations. We start with a summary of all assumptions (Assumptions 2.1), and end with a discussion of their generality (Remark 2.1).

Assumptions 2.1. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m \geq n$ have $\text{rank}(\mathbf{A}) \geq n - 1$. Let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ be a full singular value decomposition, where $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ is diagonal, and $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V} \in \mathbb{R}^{n \times n}$ are orthogonal matrices. Partition commensurately,

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \sigma_{\min} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{E} = \mathbf{U} \begin{bmatrix} \mathbf{E}_{11} & \mathbf{e}_{12} \\ \mathbf{e}_{21}^T & e_{22} \\ \mathbf{E}_{31} & \mathbf{e}_{32} \end{bmatrix} \mathbf{V}^T,$$

where $\mathbf{\Sigma}_1 \in \mathbb{R}^{(n-1) \times (n-1)}$ is nonsingular diagonal, and $\sigma_{\min} \geq 0$.

For a matrix with a single smallest singular value, we corroborate the observation that ‘small singular values tend to increase’ [SS90, page 266]. Motivated by the second-order expressions in terms of absolute perturbations [SS90, Section V.4.2] and [Ste06, Theorem 8], we derive a true lower bound.

Theorem 5. Let $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{m \times n}$ satisfy Assumptions 2.1. If $1/\|\mathbf{\Sigma}_1^{-1}\|_2 > 4\|\mathbf{E}\|_2$ and $\sigma_{\min} < \|\mathbf{E}\|_2$, then

$$\sigma_{\min}(\mathbf{A} + \mathbf{E})^2 \geq \|\mathbf{e}_{32}\|_2^2 + (\sigma_{\min} + e_{22})^2 - r_3 - r_4,$$

where r_3 contains terms of order 3,

$$r_3 \equiv 2\mathbf{e}_{12}^T \mathbf{w} \quad \mathbf{w} \equiv (\mathbf{\Sigma}_1 + \mathbf{E}_{11})^{-T} \begin{bmatrix} \mathbf{e}_{21} & \mathbf{E}_{31}^T \end{bmatrix} \begin{bmatrix} e_{22} + \sigma_{\min} \\ \mathbf{e}_{32} \end{bmatrix},$$

and r_4 contains terms of order 4 and higher,

$$r_4 \equiv \|\mathbf{w}\|_2^2 + 4 \frac{\|\mathbf{E}\|_2^2 \|(\mathbf{\Sigma}_1 + \mathbf{E}_{11})^{-1}(\mathbf{e}_{12} + \mathbf{w})\|_2^2}{1 - 4\|\mathbf{E}\|_2^2 \|(\mathbf{\Sigma}_1 + \mathbf{E}_{11})^{-1}\|_2^2}.$$

Proof. We square the singular values of $\mathbf{A} + \mathbf{E}$, and consider the eigenvalues of

$$\mathbf{B} \equiv (\mathbf{A} + \mathbf{E})^T (\mathbf{A} + \mathbf{E}) = \mathbf{V} \begin{bmatrix} \mathbf{B}_{11} & \mathbf{b} \\ \mathbf{b}^T & \beta \end{bmatrix} \mathbf{V}^T$$

where

$$(6) \quad \mathbf{B}_{11} = \underbrace{(\mathbf{\Sigma}_1 + \mathbf{E}_{11})^T (\mathbf{\Sigma}_1 + \mathbf{E}_{11})}_{\mathbf{C}_{11}} + \underbrace{\mathbf{e}_{21} \mathbf{e}_{21}^T + \mathbf{E}_{31}^T \mathbf{E}_{31}}_{\mathbf{C}_{12}}$$

$$(7) \quad \begin{aligned} \beta &= \|\mathbf{e}_{12}\|_2^2 + (\sigma_{\min} + e_{22})^2 + \|\mathbf{e}_{32}\|_2^2 \\ \mathbf{b} &= (\mathbf{\Sigma}_1 + \mathbf{E}_{11})^T \mathbf{e}_{12} + \mathbf{e}_{21}(\sigma_{\min} + e_{22}) + \mathbf{E}_{31}^T \mathbf{e}_{32}. \end{aligned}$$

From $\sigma_{\min}(\mathbf{\Sigma}_1) > 4\|\mathbf{E}\|_2$ follows that \mathbf{C}_{11} is symmetric positive definite, while \mathbf{C}_{12} is symmetric positive semi-definite and contains only second order terms. Abbreviate $\tilde{\lambda}_{\min} \equiv \lambda_{\min}(\mathbf{B}) = \sigma_{\min}(\mathbf{A} + \mathbf{E})^2$.

The proof proceeds in two steps:

1. Confirming that \mathbf{C}_{11} satisfies the assumptions of Theorem 4.
2. Deriving the lower bound for $\tilde{\lambda}_{\min}$ from Theorem 4.

1. Confirm that \mathbf{C}_{11} satisfies the assumptions of Theorem 4 We show that $\lambda_{\min}(\mathbf{C}_{11}) > \beta$, by bounding β from above and $\lambda_{\min}(\mathbf{C}_{11})$ from below.

Regarding the upper bound for β , the expression (7) and the assumption $\sigma_{\min} < \|\mathbf{E}\|_2$ imply

$$(8) \quad \beta = \left\| \begin{bmatrix} \mathbf{e}_{12}^T & e_{22} + \sigma_{\min} & \mathbf{e}_{32}^T \end{bmatrix}^T \right\|_2^2 \leq (\sigma_{\min} + \|\mathbf{E}\mathbf{e}_n\|_2)^2 \leq 4\|\mathbf{E}\|_2^2.$$

Regarding the lower bound for $\lambda_{\min}(\mathbf{C}_{11})$, view $\mathbf{C}_{11} = (\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^T(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})$ as a singular value problem, so that $\lambda_{\min}(\mathbf{C}_{11}) = \sigma_{\min}(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^2$. The well-conditioning of singular values [GV13, Corollary 8.6.2] implies

$$|\sigma_{\min}(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11}) - \sigma_{\min}(\boldsymbol{\Sigma}_1)| \leq \|\mathbf{E}_{11}\|_2 \leq \|\mathbf{E}\|_2.$$

Adding the assumption $\sigma_{\min}(\boldsymbol{\Sigma}_1) = 1/\|\boldsymbol{\Sigma}_1^{-1}\|_2 > 4\|\mathbf{E}\|_2$ gives

$$\sigma_{\min}(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11}) \geq \sigma_{\min}(\boldsymbol{\Sigma}_1) - \|\mathbf{E}\|_2 > 4\|\mathbf{E}\|_2 - \|\mathbf{E}\|_2 = 3\|\mathbf{E}\|_2.$$

Now combine this lower bound for $\lambda_{\min}(\mathbf{C}_{11})$ with (8),

$$\lambda_{\min}(\mathbf{C}_{11}) = \sigma_{\min}(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^2 > 9\|\mathbf{E}\|_2^2 > 4\|\mathbf{E}\|_2^2 \geq \beta.$$

Hence $\lambda_{\min}(\mathbf{C}_{11}) > \beta$, and \mathbf{C}_{11} satisfies the assumptions of Theorem 4.

2. Derive the lower bound for $\tilde{\lambda}_{\min}$ from Theorem 4 In this bound,

$$(9) \quad \tilde{\lambda}_{\min} \geq \beta - \mathbf{b}^T \mathbf{C}_{11}^{-1} \mathbf{b} - \frac{\beta \|\mathbf{C}_{11}^{-1} \mathbf{b}\|_2^2}{1 - \beta \|\mathbf{C}_{11}^{-1}\|_2},$$

where the key term is $\mathbf{C}_{11}^{-1} \mathbf{b}$. Insert the expression for \mathbf{b} from (7),

$$(10) \quad \begin{aligned} (\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-T} \mathbf{b} &= \mathbf{e}_{12} + (\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-T} \left(\mathbf{e}_{21}(\sigma_{\min} + e_{22}) + \mathbf{E}_{31}^T \mathbf{e}_{32} \right) \\ &= \mathbf{e}_{12} + \mathbf{w}. \end{aligned}$$

Combine the expression for \mathbf{C}_{11} from (6) with (10),

$$(11) \quad \mathbf{C}_{11}^{-1} \mathbf{b} = (\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-1} \underbrace{(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-T} \mathbf{b}}_{\mathbf{e}_{12} + \mathbf{w}} = (\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-1} (\mathbf{e}_{12} + \mathbf{w})$$

Multiply the above by \mathbf{b}^T on the left, and use (10)

$$\begin{aligned} \mathbf{b}^T \mathbf{C}_{11}^{-1} \mathbf{b} &= \mathbf{b}^T (\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-1} (\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-T} \mathbf{b} = (\mathbf{e}_{12} + \mathbf{w})^T (\mathbf{e}_{12} + \mathbf{w}) \\ &= \|\mathbf{e}_{12} + \mathbf{w}\|_2^2 = \|\mathbf{e}_{12}\|_2^2 + 2\mathbf{e}_{12}^T \mathbf{w} + \|\mathbf{w}\|_2^2. \end{aligned}$$

Substitute the above, and β from (7) into the first two summands of (9),

$$(12) \quad \begin{aligned} \beta - \mathbf{b}^T \mathbf{C}_{11}^{-1} \mathbf{b} &= (\|\mathbf{e}_{12}\|_2^2 + (\sigma_{\min} + e_{22})^2 + \|\mathbf{e}_{32}\|_2^2) - (\|\mathbf{e}_{12}\|_2^2 + 2\mathbf{e}_{12}^T \mathbf{w} + \|\mathbf{w}\|_2^2) \\ &= \|\mathbf{e}_{32}\|_2^2 + (\sigma_{\min} + e_{22})^2 - \underbrace{2\mathbf{e}_{12}^T \mathbf{w}}_{r_3} - \|\mathbf{w}\|_2^2. \end{aligned}$$

Substitute the bound for β in (8), and (11) into the third summand of (9),

$$(13) \quad \frac{\beta \|\mathbf{C}_{11}^{-1} \mathbf{b}\|_2^2}{1 - \beta \|\mathbf{C}_{11}^{-1}\|_2} \leq 4 \frac{\|\mathbf{E}\|_2^2 \|(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-1} (\mathbf{e}_{12} + \mathbf{w})\|_2^2}{1 - 4\|\mathbf{E}\|_2^2 \|(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-1}\|_2^2}.$$

Inserting (12) and (13) into (9) gives

$$\begin{aligned} \tilde{\lambda}_{\min} &\geq \|\mathbf{e}_{32}\|_2^2 + (\sigma_{\min} + e_{22})^2 \\ &\quad - \underbrace{\left(\|\mathbf{w}\|_2^2 + 4 \frac{\|\mathbf{E}\|_2^2 \|(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-1} (\mathbf{e}_{12} + \mathbf{w})\|_2^2}{1 - 4\|\mathbf{E}\|_2^2 \|(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-1}\|_2^2} \right)}_{r_4}. \end{aligned}$$

□

Remark 2.1. *The assumptions in Theorem 5 are not restrictive. Only a small gap of $3\|\mathbf{E}\|_2$ is required to separate the smallest singular value of \mathbf{A} from the remaining singular values,*

$$\sigma_{\min}(\mathbf{A}) < \|\mathbf{E}\|_2 < 4\|\mathbf{E}\|_2 \leq 1/\|\Sigma_1^{-1}\|_2.$$

3 A cluster of small singular values

We extend the results in Section 2 from a single smallest singular value to a cluster of small singular values. To this end, we derive lower bounds for the small singular values of the perturbed matrix in terms of normwise absolute perturbations (Section 3.2), based on eigenvalue bounds (Section 3.1).

3.1 Auxiliary eigenvalue results

We square the singular values of $\mathbf{A} \in \mathbb{R}^{m \times n}$ and consider instead the eigenvalues of the symmetric positive semi-definite matrix $\mathbf{B} \equiv \mathbf{A}^T \mathbf{A} \in \mathbb{R}^{n \times n}$.

For a symmetric positive semi-definite matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$ with a cluster of r small eigenvalues, we present an expression for these eigenvalues (Lemma 3), and two lower bounds in terms of normwise absolute perturbations (Theorems 6 and 7).

We assume that the r small eigenvalues are separated from the remaining ones, in the sense that they are strictly smaller than the smallest eigenvalue of the leading principal submatrix \mathbf{B}_{11} of order $n - r$. The eigenvalues are labelled so that

$$\lambda_n(\mathbf{B}) \leq \dots \leq \lambda_{n-r+1}(\mathbf{B}) < \lambda_{n-r}(\mathbf{B}) \leq \dots \leq \lambda_1(\mathbf{B}).$$

The equality below expresses the smallest eigenvalues in terms of themselves, and represents an extension of Lemma 2 to clusters.

Lemma 3 (Exact expression). *Let $\mathbf{B} \in \mathbb{R}^{n \times n}$ be symmetric positive semi-definite with $\text{rank}(\mathbf{B}) \geq n - r$ for some $r \geq 1$, and partition*

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{12}^T & \mathbf{B}_{22} \end{bmatrix} \quad \text{where } \mathbf{B}_{11} \in \mathbb{R}^{(n-r) \times (n-r)}, \quad \mathbf{B}_{22} \in \mathbb{R}^{r \times r}.$$

If $\|\mathbf{B}_{22}\|_2 < \lambda_{\min}(\mathbf{B}_{11})$ then

$$\lambda_{n-r+j}(\mathbf{B}) = \lambda_j \left(\mathbf{B}_{22} - \mathbf{B}_{12}^T (\mathbf{B}_{11} - \lambda_{n-r+j}(\mathbf{B}) \mathbf{I})^{-1} \mathbf{B}_{12} \right), \quad 1 \leq j \leq r,$$

where

$$(14) \quad 0 \leq \lambda_{n-r+j}(\mathbf{B}) \leq \|\mathbf{B}_{22}\|_2, \quad 1 \leq j \leq r.$$

Proof. Abbreviate $\tilde{\lambda}_{n-r+j} \equiv \lambda_{n-r+j}(\mathbf{B})$, $1 \leq j \leq r$. The lower bound in (14) follows from the positive semi-definiteness of \mathbf{B} , and the upper bound from the Cauchy interlace theorem [Par80, Section 10.1]

$$\tilde{\lambda}_{n-r+j} \leq \lambda_j(\mathbf{B}_{22}) \leq \lambda_{\max}(\mathbf{B}_{22}) = \|\mathbf{B}_{22}\|_2, \quad 1 \leq j \leq r.$$

Combining this with the assumption $\|\mathbf{B}_{22}\|_2 < \lambda_{\min}(\mathbf{B}_{11})$ shows

$$(15) \quad \tilde{\lambda}_{n-r+j} \leq \|\mathbf{B}_{22}\|_2 < \lambda_{\min}(\mathbf{B}_{11}), \quad 1 \leq j \leq r.$$

Hence is $\mathbf{B}_{11} - \tilde{\lambda}_{n-r+j} \mathbf{I}$ nonsingular.

To derive the expression for $\tilde{\lambda}_{n-r+j}$, we start as in the proof of [Par80, Theorem (10-1-2)]. The shifted matrix $\mathbf{B} - \tilde{\lambda}_{n-r+j} \mathbf{I}$ has at most $n - r + j - 1$ positive eigenvalues, at least one zero eigenvalue, and at most $r - j$ negative eigenvalues, $1 \leq j \leq r$. Perform the congruence transformation

$$\mathbf{B} - \tilde{\lambda}_{n-r+j} \mathbf{I} = \mathbf{L} \begin{bmatrix} \mathbf{B}_{11} - \tilde{\lambda}_{n-r+j} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{bmatrix} \mathbf{L}^T, \quad \mathbf{L} \equiv \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{B}_{12}^T (\mathbf{B}_{11} - \tilde{\lambda}_{n-r+j} \mathbf{I})^{-1} & \mathbf{I} \end{bmatrix}$$

where

$$(16) \quad \mathbf{S} \equiv \mathbf{B}_{22} - \tilde{\lambda}_{n-r+j} \mathbf{I} - \mathbf{B}_{12}^T (\mathbf{B}_{11} - \tilde{\lambda}_{n-r+j} \mathbf{I})^{-1} \mathbf{B}_{12}, \quad 1 \leq j \leq r.$$

From (15) follows that $\mathbf{B}_{11} - \tilde{\lambda}_{n-r+j} \mathbf{I}$ has $n - r$ positive eigenvalues. Combining this with the inertia preservation of congruence transformations implies that \mathbf{S} has at most $r - j$ positive eigenvalues, at least one zero eigenvalue $\lambda_j(\mathbf{S}) = 0$, and at least $j - 1$ negative eigenvalues, $1 \leq j \leq r$. Insert (16) into $\lambda_j(\mathbf{S}) = 0$, and exploit the fact that the shift $\tilde{\lambda}_{n-r+j} \mathbf{I}$ does not change the algebraic eigenvalue ordering, to obtain the expression for $\tilde{\lambda}_{n-r+j}$, $1 \leq j \leq r$. \square

By restricting ourselves to a ‘dominant part’ of \mathbf{B}_{11} , we weaken the expression in Lemma 3 to a lower bound, which allows the eigenvalues to be negative.

Lemma 4 (Lower bound). *Let $\mathbf{B} \in \mathbb{R}^{n \times n}$ be symmetric positive semi-definite with $\text{rank}(\mathbf{B}) \geq n - r$ for some $r \geq 1$, and partition*

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{12}^T & \mathbf{B}_{22} \end{bmatrix} \quad \text{where} \quad \mathbf{B}_{11} \in \mathbb{R}^{(n-r) \times (n-r)}, \quad \mathbf{B}_{22} \in \mathbb{R}^{r \times r}.$$

Let $\mathbf{B}_{11} = \mathbf{C}_{11} + \mathbf{C}_{12}$ where $\mathbf{C}_{11} \in \mathbb{R}^{(n-r) \times (n-r)}$ is symmetric positive definite and $\mathbf{C}_{12} \in \mathbb{R}^{(n-r) \times (n-r)}$ is symmetric positive semi-definite. If

$$\widehat{\mathbf{B}} \equiv \begin{bmatrix} \mathbf{C}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{12}^T & \mathbf{B}_{22} \end{bmatrix}$$

with $\lambda_{\min}(\mathbf{C}_{11}) > \|\mathbf{B}_{22}\|_2$, then

$$(17) \quad \lambda_{n-r+j}(\mathbf{B}) \geq \lambda_{n-r+j}(\widehat{\mathbf{B}})$$

$$(18) \quad = \lambda_j \left(\mathbf{B}_{22} - \mathbf{B}_{12}^T (\mathbf{C}_{11} - \lambda_{n-r+j}(\widehat{\mathbf{B}}) \mathbf{I})^{-1} \mathbf{B}_{12} \right), \quad 1 \leq j \leq r,$$

where

$$(19) \quad \lambda_{n-r+j}(\widehat{\mathbf{B}}) \leq \|\mathbf{B}_{22}\|_2,$$

$$(20) \quad \left\| \left(\mathbf{C}_{11} - \lambda_{n-r+j}(\widehat{\mathbf{B}}) \mathbf{I} \right)^{-1} \right\|_2 \leq \frac{\|\mathbf{C}_{11}^{-1}\|_2}{1 - \|\mathbf{B}_{22}\|_2 \|\mathbf{C}_{11}^{-1}\|_2}, \quad 1 \leq j \leq r.$$

Proof. The proof proceeds in four steps.

Proof of (17) The symmetric positive semi-definiteness of \mathbf{C}_{12} and Weyl’s monotonicity theorem [HJ13, Corollary 4.3.3] imply

$$\lambda_j(\mathbf{B}) \geq \lambda_j(\widehat{\mathbf{B}}), \quad 1 \leq j \leq n.$$

Now we concentrate on the eigenvalues of $\widehat{\mathbf{B}}$, and abbreviate $\widehat{\lambda}_{n-r+j} \equiv \lambda_{n-r+j}(\widehat{\mathbf{B}})$, $1 \leq j \leq r$.

Proof of (19) Apply the Cauchy interlace theorem [Par80, Section 10.1] to $\widehat{\mathbf{B}}$,

$$\widehat{\lambda}_{n-r+j} \leq \lambda_j(\mathbf{B}_{22}) \leq \lambda_{\max}(\mathbf{B}_{22}) = \|\mathbf{B}_{22}\|_2, \quad 1 \leq j \leq r.$$

Combining this with the assumption $\|\mathbf{B}_{22}\|_2 < \lambda_{\min}(\mathbf{C}_{11})$ shows

$$\widehat{\lambda}_{n-r+j} \leq \|\mathbf{B}_{22}\|_2 < \lambda_{\min}(\mathbf{C}_{11}), \quad 1 \leq j \leq r.$$

Hence $\mathbf{C}_{11} - \widehat{\lambda}_{n-r+j} \mathbf{I}$ is nonsingular, which holds in particular if $\widehat{\lambda}_{n-r+j} < 0$.

Proof of (18) To derive the expression for $\widehat{\lambda}_{n-r+j}$, apply the proof of Lemma 3 to the eigenvalues of $\widehat{\mathbf{B}}$. This proof relies only on the signs of eigenvalues of shifted matrices, and does not require positive semi-definiteness of the host matrix $\widehat{\mathbf{B}}$.

Proof of (20) Fix some $1 \leq j \leq r$ for the inverse in (18). Then factor out \mathbf{C}_{11}^{-1} ,

$$(\mathbf{C}_{11} - \widehat{\lambda}_{n-r+j} \mathbf{I})^{-1} = \mathbf{C}_{11}^{-1} \mathbf{D} \quad \text{where} \quad \mathbf{D} \equiv (\mathbf{I} - \widehat{\lambda}_{n-r+j} \mathbf{C}_{11}^{-1})^{-1},$$

and take norms,

$$\|(\mathbf{C}_{11} - \widehat{\lambda}_{n-r+j} \mathbf{I})^{-1}\|_2 \leq \|\mathbf{C}_{11}^{-1}\|_2 \|\mathbf{D}\|_2.$$

To bound $\|\mathbf{D}\|_2$, consider the eigenvalue decomposition $\mathbf{C}_{11} = \mathbf{W} \mathbf{\Lambda} \mathbf{W}^T$, where \mathbf{W} is an orthogonal matrix, and the diagonal matrix

$$\mathbf{\Lambda} = \text{diag}(\gamma_1 \quad \cdots \quad \gamma_{n-r}) \in \mathbb{R}^{(n-r) \times (n-r)}$$

has positive diagonal elements $\gamma_\ell > 0$. Thus \mathbf{D} has an eigenvalue decomposition $\mathbf{D} = \mathbf{W}(\mathbf{I} - \widehat{\lambda}_{n-r+j} \mathbf{\Lambda}^{-1})^{-1} \mathbf{W}^T$ with eigenvalues

$$\lambda_\ell(\mathbf{D}) = 1 / \left(1 - \frac{\widehat{\lambda}_{n-r+j}}{\gamma_\ell} \right), \quad 1 \leq \ell \leq n-r.$$

Case 1: If $\widehat{\lambda}_{n-r+j} \geq 0$, then (19) implies

$$0 \leq \frac{\widehat{\lambda}_{n-r+j}}{\gamma_\ell} \leq \frac{\|\mathbf{B}_{22}\|_2}{\lambda_{\min}(\mathbf{C}_{11})} = \|\mathbf{B}_{22}\|_2 \|\mathbf{C}_{11}^{-1}\|_2 < 1, \quad 1 \leq \ell \leq n-r.$$

Hence

$$(21) \quad \|\mathbf{D}\|_{22} = \max_{1 \leq \ell \leq n-r} |\lambda_j(\mathbf{D})| \leq \frac{1}{1 - \|\mathbf{B}_{22}\|_2 \|\mathbf{C}_{11}^{-1}\|_2}.$$

Case 2: If $\widehat{\lambda}_{n-r+j} < 0$, then $\gamma_\ell > 0$ and (19) imply

$$1 - \frac{\widehat{\lambda}_{n-r+j}}{\gamma_\ell} = 1 + \frac{|\widehat{\lambda}_{n-r+j}|}{\gamma_\ell} > 1 > 1 - \|\mathbf{B}_{22}\|_2 \|\mathbf{C}_{11}^{-1}\|_2, \quad 1 \leq \ell \leq n-r.$$

Again, as in (21) we conclude

$$\|\mathbf{D}\|_{22} = \max_{1 \leq \ell \leq n-r} |\lambda_j(\mathbf{D})| \leq \frac{1}{1 - \|\mathbf{B}_{22}\|_2 \|\mathbf{C}_{11}^{-1}\|_2}.$$

Since we fixed an arbitrary j to show (20), it holds for all $1 \leq j \leq r$. \square

The subsequent lower bounds are informative if the offdiagonal part has small norm. We start with an extension of Theorem 3.

Theorem 6 (First lower bound). *Let $\mathbf{B} \in \mathbb{R}^{n \times n}$ be symmetric positive semi-definite with $\text{rank}(\mathbf{B}) \geq n-r$ for some $r \geq 1$, and partition*

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{12}^T & \mathbf{B}_{22} \end{bmatrix} \quad \text{where} \quad \mathbf{B}_{11} \in \mathbb{R}^{(n-r) \times (n-r)}, \quad \mathbf{B}_{22} \in \mathbb{R}^{r \times r}.$$

If $\|\mathbf{B}_{22}\|_2 < \lambda_{\min}(\mathbf{B}_{11})$ then $\lambda_{n-r+j}(\mathbf{B}) \geq \lambda_j(\mathbf{Z}_j)$, $1 \leq j \leq r$, where

$$\mathbf{Z}_j \equiv \mathbf{B}_{22} - \mathbf{B}_{12}^T \mathbf{B}_{11}^{-1} \mathbf{B}_{12} - \|\mathbf{B}_{22}\|_2 \mathbf{B}_{12}^T \mathbf{B}_{11}^{-1} (\mathbf{I} - \lambda_{n-r+j}(\mathbf{B}) \mathbf{B}_{11}^{-1})^{-1} \mathbf{B}_{11}^{-1} \mathbf{B}_{12}.$$

Proof. Abbreviate $\tilde{\lambda}_{n-r+j} \equiv \lambda_{n-r+j}(\mathbf{B})$, $1 \leq j \leq r$. As in the proof of Theorem 3, apply the Sherman-Morrison formula [GV13, Section 2.1.4]

$$(\mathbf{B}_{11} - \tilde{\lambda}_{n-r+j} \mathbf{I})^{-1} = \mathbf{B}_{11}^{-1} + \tilde{\lambda}_{n-r+j} \mathbf{B}_{11}^{-1} (\mathbf{I} - \tilde{\lambda}_{n-r+j} \mathbf{B}_{11}^{-1})^{-1} \mathbf{B}_{11}^{-1},$$

and substitute the above into the expressions for

$$(22) \quad \tilde{\lambda}_{n-r+j} = \lambda_j(\mathbf{M}_j), \quad 1 \leq j \leq r$$

from Lemma 3 where

$$\begin{aligned} \mathbf{M}_j &\equiv \mathbf{B}_{22} - \mathbf{B}_{12}^T (\mathbf{B}_{11} - \tilde{\lambda}_{n-r+j} \mathbf{I})^{-1} \mathbf{B}_{12} \\ &= \mathbf{B}_{22} - \mathbf{B}_{12}^T \mathbf{B}_{11}^{-1} \mathbf{B}_{12} - \tilde{\lambda}_{n-r+j} \mathbf{B}_{12}^T \mathbf{B}_{11}^{-1} (\mathbf{I} - \tilde{\lambda}_{n-r+j} \mathbf{B}_{11}^{-1})^{-1} \mathbf{B}_{11}^{-1} \mathbf{B}_{12}. \end{aligned}$$

From (14) follows the Loewner bound

$$\mathbf{M}_j \succeq \mathbf{Z}_j \equiv \mathbf{B}_{22} - \mathbf{B}_{12}^T \mathbf{B}_{11}^{-1} \mathbf{B}_{12} - \|\mathbf{B}_{22}\|_2 \mathbf{B}_{12}^T \mathbf{B}_{11}^{-1} (\mathbf{I} - \tilde{\lambda}_{r+j} \mathbf{B}_{11}^{-1})^{-1} \mathbf{B}_{11}^{-1} \mathbf{B}_{12}.$$

This and (22) imply $\tilde{\lambda}_{n-r+j} = \lambda_j(\mathbf{M}_j) \geq \lambda_j(\mathbf{Z}_j)$, $1 \leq j \leq r$ [HJ13, Corollary 7.7.4]. \square

The slightly weaker bound below extends Theorem 4 and focusses on a 'dominant' part of \mathbf{B}_{11} . This establishes the connection to Theorem 8, where \mathbf{B} represents the perturbed matrix and the low order terms in \mathbf{B}_{11} are captured by \mathbf{C}_{11} .

Theorem 7 (Second lower bound). *Let $\mathbf{B} \in \mathbb{R}^{n \times n}$ be symmetric positive semi-definite with $\text{rank}(\mathbf{B}) \geq n - r$ for some $r \geq 1$, and partition*

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{12}^T & \mathbf{B}_{22} \end{bmatrix} \quad \text{where } \mathbf{B}_{11} \in \mathbb{R}^{(n-r) \times (n-r)}, \quad \mathbf{B}_{22} \in \mathbb{R}^{r \times r}.$$

If $\mathbf{B}_{11} = \mathbf{C}_{11} + \mathbf{C}_{12}$ where \mathbf{C}_{11} is symmetric positive definite with $\lambda_{\min}(\mathbf{C}_{11}) > \|\mathbf{B}_{22}\|_2$, and \mathbf{C}_{12} is symmetric positive semi-definite, then

$$\lambda_{n-r+j}(\mathbf{B}) \geq \lambda_j \left(\mathbf{B}_{22} - \mathbf{B}_{12}^T \mathbf{C}_{11}^{-1} \mathbf{B}_{12} \right) - \frac{\|\mathbf{B}_{22}\|_2 \|\mathbf{C}_{11}^{-1} \mathbf{B}_{12}\|_2^2}{1 - \|\mathbf{B}_{22}\|_2 \|\mathbf{C}_{11}^{-1}\|_2}, \quad 1 \leq j \leq r.$$

Proof. Define

$$\hat{\mathbf{B}} \equiv \begin{bmatrix} \mathbf{C}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{12}^T & \mathbf{B}_{22} \end{bmatrix},$$

and abbreviate $\hat{\lambda}_{n-r+j} \equiv \lambda_{n-r+j}(\hat{\mathbf{B}})$, $1 \leq j \leq r$. From (18) in Lemma 4 follows

$$\lambda_{n-r+j}(\mathbf{B}) \geq \hat{\lambda}_{n-r+j} = \lambda_j \left(\mathbf{B}_{22} - \mathbf{B}_{12}^T (\mathbf{C}_{11} - \hat{\lambda}_{n-r+j} \mathbf{I})^{-1} \mathbf{B}_{12} \right), \quad 1 \leq j \leq r,$$

We proceed as in the proof of Theorem 6, and apply the Sherman-Morrison formula [GV13, Section 2.1.4],

$$(\mathbf{C}_{11} - \hat{\lambda}_{n-r+j} \mathbf{I})^{-1} = \mathbf{C}_{11}^{-1} + \hat{\lambda}_{n-r+j} \mathbf{C}_{11}^{-1} (\mathbf{I} - \hat{\lambda}_{n-r+j} \mathbf{C}_{11}^{-1})^{-1} \mathbf{C}_{11}^{-1},$$

and (19) to the expression for

$$(23) \quad \hat{\lambda}_{n-r+j} = \lambda_j(\mathbf{M}_j), \quad 1 \leq j \leq r,$$

from Lemma 4, where

$$\begin{aligned} \mathbf{M}_j &\equiv \mathbf{B}_{22} - \mathbf{B}_{12}^T (\mathbf{C}_{11} - \hat{\lambda}_{n-r+j} \mathbf{I})^{-1} \mathbf{B}_{12}, \quad 1 \leq j \leq r \\ &= \mathbf{B}_{22} - \mathbf{B}_{12}^T \mathbf{C}_{11}^{-1} \mathbf{B}_{12} - \hat{\lambda}_{n-r+j} \mathbf{B}_{12}^T \mathbf{C}_{11}^{-1} (\mathbf{I} - \hat{\lambda}_{n-r+j} \mathbf{C}_{11}^{-1})^{-1} \mathbf{C}_{11}^{-1} \mathbf{B}_{12}. \end{aligned}$$

From (19) and (20) follows the lower bound

$$\begin{aligned} \mathbf{M}_j &\succeq \mathbf{B}_{22} - \mathbf{B}_{12}^T \mathbf{C}_{11}^{-1} \mathbf{B}_{12} - \|\mathbf{B}_{22}\|_2 \mathbf{B}_{12}^T \mathbf{C}_{11}^{-1} (\mathbf{I} - \tilde{\lambda}_{r+j} \mathbf{C}_{11}^{-1})^{-1} \mathbf{C}_{11}^{-1} \mathbf{B}_{12} \\ &\succeq \mathbf{Z} \equiv \mathbf{B}_{22} - \mathbf{B}_{12}^T \mathbf{C}_{11}^{-1} \mathbf{B}_{12} - \underbrace{\frac{\|\mathbf{B}_{22}\|_2 \|\mathbf{C}_{11}^{-1} \mathbf{B}_{12}\|_2^2}{1 - \|\mathbf{B}_{22}\|_2 \|\mathbf{C}_{11}^{-1}\|_2}}_{\gamma} \mathbf{I}, \quad 1 \leq j \leq r. \end{aligned}$$

Thus, $\mathbf{M}_j \succeq \mathbf{Z}$, $1 \leq j \leq r$. The Loewner properties [HJ13, Corollary 7.7.4] imply the same for the eigenvalues, $\lambda_j(\mathbf{M}_j) \geq \lambda_j(\mathbf{Z})$, $1 \leq j \leq r$. Combine this with the well conditioning of eigenvalues [GV13, Theorem 8.1.5],

$$\tilde{\lambda}_{n-r+j} = \lambda_j(\mathbf{M}_j) \geq \lambda_j(\mathbf{Z}) \geq \lambda_j(\mathbf{B}_{22} - \mathbf{B}_{12}^T \mathbf{C}_{11}^{-1} \mathbf{B}_{12}) - \gamma, \quad 1 \leq j \leq r.$$

□

Theorem 7 reduces to Theorem 4 for $r = 1$, and to Theorem 6 for $\mathbf{C}_{12} = \mathbf{0}$.

3.2 A lower bound for a cluster of smallest singular values

We extend the bound for a single smallest singular value in section 2.2 to a cluster of smallest singular values. The resulting lower bound for the cluster of perturbed smallest singular values (Theorem 8) is based on the eigenvalue bounds in section 3.1, and expressed in terms of normwise absolute perturbations. We start with a summary of all assumptions (Assumptions 3.1), and end with a discussion of their generality (Remark 3.1).

Assumptions 3.1. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m \geq n$ have $\text{rank}(\mathbf{A}) \geq n - r$ for some $r \geq 1$. Let $\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$ be a full singular value decomposition, where $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ is diagonal, and $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V} \in \mathbb{R}^{n \times n}$ are orthogonal matrices. Partition commensurately,

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{E} = \mathbf{U} \begin{bmatrix} \mathbf{E}_{11} & \mathbf{E}_{12} \\ \mathbf{E}_{21} & \mathbf{E}_{22} \\ \mathbf{E}_{31} & \mathbf{E}_{32} \end{bmatrix} \mathbf{V}^T,$$

where $\mathbf{\Sigma}_1 \in \mathbb{R}^{(n-r) \times (n-r)}$ is nonsingular diagonal, and $\mathbf{\Sigma}_2 \in \mathbb{R}^{r \times r}$ is diagonal.

This bound below extends Theorem 5, and reduces to it for $r = 1$.

Theorem 8. Let $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{m \times n}$ satisfy Assumptions 3.1. If $1/\|\mathbf{\Sigma}_1^{-1}\|_2 > 4\|\mathbf{E}\|_2$ and $\|\mathbf{\Sigma}_2\|_2 < \|\mathbf{E}\|_2$, then

$$\sigma_{n-r+j}(\mathbf{A} + \mathbf{E})^2 \geq \lambda_j(\mathbf{E}_{32}^T \mathbf{E}_{32} + (\mathbf{\Sigma}_2 + \mathbf{E}_{22})^T (\mathbf{\Sigma}_2 + \mathbf{E}_{22}) - \mathbf{R}_3) - r_4, \quad 1 \leq j \leq r,$$

where \mathbf{R}_3 contains terms of order 3

$$\mathbf{R}_3 \equiv \mathbf{E}_{12}^T \mathbf{W} + \mathbf{W}^T \mathbf{E}_{12}, \quad \mathbf{W} \equiv (\mathbf{\Sigma}_1 + \mathbf{E}_{11})^{-T} \begin{bmatrix} \mathbf{E}_{21} & \mathbf{E}_{31}^T \\ \mathbf{E}_{22} & \mathbf{E}_{32} \end{bmatrix}$$

and r_4 contains terms of order 4 and higher,

$$r_4 \equiv \|\mathbf{W}\|_2^2 + 4 \frac{\|\mathbf{E}\|_2^2 \|(\mathbf{\Sigma}_1 + \mathbf{E}_{11})^{-1} (\mathbf{E}_{12} + \mathbf{W})\|_2^2}{1 - 4\|\mathbf{E}\|_2^2 \|(\mathbf{\Sigma}_1 + \mathbf{E}_{11})^{-1}\|_2^2}.$$

Proof. We square the singular values of $\mathbf{A} + \mathbf{E}$ and consider the eigenvalues of the perturbed matrix

$$\mathbf{B} \equiv (\mathbf{A} + \mathbf{E})^T (\mathbf{A} + \mathbf{E}) = \mathbf{V} \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{12}^T & \mathbf{B}_{22} \end{bmatrix} \mathbf{V}^T$$

where

$$(24) \quad \mathbf{B}_{11} = \underbrace{(\mathbf{\Sigma}_1 + \mathbf{E}_{11})^T (\mathbf{\Sigma}_1 + \mathbf{E}_{11})}_{\mathbf{C}_{11}} + \underbrace{\mathbf{E}_{21}^T \mathbf{E}_{21} + \mathbf{E}_{31}^T \mathbf{E}_{31}}_{\mathbf{C}_{12}}$$

$$(25) \quad \begin{aligned} \mathbf{B}_{22} &= \mathbf{E}_{12}^T \mathbf{E}_{12} + (\mathbf{\Sigma}_2 + \mathbf{E}_{22})^T (\mathbf{\Sigma}_2 + \mathbf{E}_{22}) + \mathbf{E}_{32}^T \mathbf{E}_{32} \\ \mathbf{B}_{12} &= (\mathbf{\Sigma}_1 + \mathbf{E}_{11})^T \mathbf{E}_{12} + \mathbf{E}_{21}^T (\mathbf{\Sigma}_2 + \mathbf{E}_{22}) + \mathbf{E}_{31}^T \mathbf{E}_{32}. \end{aligned}$$

From $\sigma_{\min}(\boldsymbol{\Sigma}_1) > 4\|\mathbf{E}\|_2$ follows that \mathbf{C}_{11} is symmetric positive definite, while \mathbf{C}_{12} is symmetric positive semi-definite and contains only second order terms. Abbreviate $\tilde{\lambda}_{n-r+j} \equiv \lambda_{n-r+j}(\mathbf{B}) = \sigma_{n-r+j}(\mathbf{A} + \mathbf{E})^2$, $1 \leq j \leq r$.

The proof proceeds in two steps:

1. Confirming that \mathbf{C}_{11} satisfies the assumptions of Theorem 7.
2. Deriving the lower bounds for $\tilde{\lambda}_{n-r+j}$ from Theorem 7.

1. Confirm that \mathbf{C}_{11} satisfies the assumptions of Theorem 7 We show that $\lambda_{\min}(\mathbf{C}_{11}) > \|\mathbf{B}_{22}\|_2$, by bounding $\|\mathbf{B}_{22}\|_2$ from above and $\lambda_{\min}(\mathbf{C}_{11})$ from below.

Regarding the upper bound for $\|\mathbf{B}_{22}\|_2$, the expression for \mathbf{B}_{22} in (25) and the assumption $\|\boldsymbol{\Sigma}_2\|_2 < \|\mathbf{E}\|_2$ imply

$$(26) \quad \|\mathbf{B}_{22}\|_2 = \left\| \begin{bmatrix} \mathbf{E}_{12}^T & (\mathbf{E}_{22} + \boldsymbol{\Sigma}_2)^T & \mathbf{E}_{32}^T \end{bmatrix}^T \right\|_2^2 \leq (\|\boldsymbol{\Sigma}_2\|_2 + \|\mathbf{E}\|_2)^2 \leq 4\|\mathbf{E}\|_2^2.$$

Regarding the lower bound for $\lambda_{\min}(\mathbf{C}_{11})$, view $\mathbf{C}_{11} = (\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^T(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})$ as a singular value problem, so that $\lambda_{\min}(\mathbf{C}_{11}) = \sigma_{\min}(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^2$. The well-conditioning of singular values [GV13, Corollary 8.6.2] implies

$$|\sigma_{\min}(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11}) - \sigma_{\min}(\boldsymbol{\Sigma}_1)| \leq \|\mathbf{E}_{11}\|_2 \leq \|\mathbf{E}\|_2.$$

Adding the assumption $\sigma_{\min}(\boldsymbol{\Sigma}_1) = 1/\|\boldsymbol{\Sigma}_1^{-1}\|_2 > 4\|\mathbf{E}\|_2$ gives

$$\sigma_{\min}(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11}) \geq \sigma_{\min}(\boldsymbol{\Sigma}_1) - \|\mathbf{E}\|_2 > 4\|\mathbf{E}\|_2 - \|\mathbf{E}\|_2 = 3\|\mathbf{E}\|_2.$$

Now combine this lower bound for $\lambda_{\min}(\mathbf{C}_{11})$ with (26),

$$\lambda_{\min}(\mathbf{C}_{11}) = \sigma_{\min}(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^2 > 9\|\mathbf{E}\|_2^2 > 4\|\mathbf{E}\|_2^2 \geq \|\mathbf{B}_{22}\|_2.$$

Hence $\lambda_{\min}(\mathbf{C}_{11}) > \|\mathbf{B}_{22}\|_2$, and \mathbf{C}_{11} satisfies the assumptions of Theorem 4.

2. Derive the lower bounds for $\tilde{\lambda}_{n-r+j}$ from Theorem 7 In these bounds,

$$(27) \quad \tilde{\lambda}_{n-r+j} \geq \lambda_j(\mathbf{S}) - \frac{\|\mathbf{B}_{22}\|_2 \|\mathbf{C}_{11}^{-1} \mathbf{B}_{12}\|_2^2}{1 - \|\mathbf{B}_{22}\|_2 \|\mathbf{C}_{11}^{-1}\|_2}, \quad \mathbf{S} \equiv \mathbf{B}_{22} - \mathbf{B}_{12}^T \mathbf{C}_{11}^{-1} \mathbf{B}_{12},$$

the key term is $\mathbf{C}_{11}^{-1} \mathbf{B}_{12}$. Insert the expression for \mathbf{B}_{12} from (25),

$$(28) \quad \begin{aligned} (\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-T} \mathbf{B}_{12} &= \mathbf{E}_{12} + (\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-T} \left(\mathbf{E}_{21}^T (\boldsymbol{\Sigma}_2 + \mathbf{E}_{22}) + \mathbf{E}_{31}^T \mathbf{E}_{32} \right) \\ &= \mathbf{E}_{12} + \mathbf{W}. \end{aligned}$$

Combine the expression for \mathbf{C}_{11} from (24) with the above,

$$(29) \quad \mathbf{C}_{11}^{-1} \mathbf{B}_{12} = (\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-1} \underbrace{(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-T} \mathbf{B}_{12}}_{\mathbf{E}_{12} + \mathbf{W}} = (\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-1} (\mathbf{E}_{12} + \mathbf{W})$$

Multiply the above by \mathbf{B}_{12}^T on the left, and use (28),

$$\begin{aligned} \mathbf{B}_{12}^T \mathbf{C}_{11}^{-1} \mathbf{B}_{12} &= \mathbf{B}_{12}^T (\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-1} (\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-T} \mathbf{B}_{12} \\ &= (\mathbf{E}_{12} + \mathbf{W})^T (\mathbf{E}_{12} + \mathbf{W}) = \mathbf{E}_{12}^T \mathbf{E}_{12} + \mathbf{E}_{12}^T \mathbf{W} + \mathbf{W}^T \mathbf{E}_{12} + \mathbf{W}^T \mathbf{W}. \end{aligned}$$

Substitute the above, and \mathbf{B}_{22} from (25) into \mathbf{S} from (27),

$$\begin{aligned} \mathbf{S} &= \mathbf{E}_{12}^T \mathbf{E}_{12} + (\boldsymbol{\Sigma}_2 + \mathbf{E}_{22})^T (\boldsymbol{\Sigma}_2 + \mathbf{E}_{22}) + \mathbf{E}_{32}^T \mathbf{E}_{32} \\ &\quad - (\mathbf{E}_{12}^T \mathbf{E}_{12} + \mathbf{E}_{12}^T \mathbf{W} + \mathbf{W}^T \mathbf{E}_{12} + \mathbf{W}^T \mathbf{W}) \\ &= \mathbf{E}_{32}^T \mathbf{E}_{32} + (\boldsymbol{\Sigma}_2 + \mathbf{E}_{22})^T (\boldsymbol{\Sigma}_2 + \mathbf{E}_{22}) - \mathbf{R}_3 - \mathbf{W}^T \mathbf{W}. \end{aligned}$$

The well conditioning of eigenvalues [GV13, Theorem 8.1.5] implies

$$(30) \quad \lambda_j(\mathbf{S}) \geq \lambda_j(\mathbf{E}_{32}^T \mathbf{E}_{32} + (\boldsymbol{\Sigma}_2 + \mathbf{E}_{22})^T (\boldsymbol{\Sigma}_2 + \mathbf{E}_{22}) - \mathbf{R}_3) - \|\mathbf{W}\|_2^2.$$

Substitute the bound for $\|\mathbf{B}_{22}\|_2$ from (26), and (28) into the second summand of (27),

$$(31) \quad \frac{\|\mathbf{B}_{22}\|_2 \|\mathbf{C}_{11}^{-1} \mathbf{B}_{12}\|_2^2}{1 - \|\mathbf{B}_{22}\|_2 \|\mathbf{C}_{11}^{-1}\|_2} \leq 4 \frac{\|\mathbf{E}\|_2^2 \|(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-1} (\mathbf{E}_{12} + \mathbf{W})\|_2^2}{1 - 4\|\mathbf{E}\|_2^2 \|(\boldsymbol{\Sigma}_1 + \mathbf{E}_{11})^{-1}\|_2^2}.$$

At last insert (30) and (31) into (27). \square

Remark 3.1. *The assumptions in Theorem 8 are not restrictive. Only a small gap of $3\|\mathbf{E}\|_2$ is required to separate the small singular value cluster of \mathbf{A} from the remaining singular values,*

$$\|\boldsymbol{\Sigma}_2\|_2 < \|\mathbf{E}\|_2 < 4\|\mathbf{E}\|_2 \leq 1/\|\boldsymbol{\Sigma}_1^{-1}\|_2.$$

4 Probabilistic componentwise relative perturbations

We consider componentwise relative perturbations based on independent uniform random variables, and derive an expression for the expectation of a sum of squared perturbed singular values (Theorem 9), which can also be interpreted as a relative perturbation of the Frobenius norm of a perturbed matrix (Remark 4.1) and leads to a probabilistic expression for a perturbed cluster of small singular values (Corollary 10).

We denote by ϵ_{lower} the unit roundoff of the lower precision to which the matrix is demoted; and by a_{ij} element (i, j) of the matrix \mathbf{A} .

Theorem 9. *Let $\mathbf{A}, \mathbf{F} \in \mathbb{R}^{m \times n}$ where $m \geq n$, and the elements of \mathbf{F} are independent random variables with*

$$(32) \quad f_{ij} \equiv \epsilon_{\text{lower}} a_{ij} \omega_{ij}, \quad \omega_{ij} \in \mathcal{U}(-1, 1), \quad 1 \leq i \leq m, \quad 1 \leq j \leq n.$$

Then

$$\mathbb{E} \left[\sum_{j=1}^n \sigma_j(\mathbf{A} + \mathbf{F})^2 \right] = \sum_{j=1}^n \sigma_j(\mathbf{A})^2 + \frac{\epsilon_{\text{lower}}^2}{3} \|\mathbf{A}\|_F^2.$$

Proof. The linearity of the trace implies

$$\begin{aligned} \sum_{j=1}^n \sigma_j(\mathbf{A} + \mathbf{F})^2 &= \text{trace}((\mathbf{A} + \mathbf{F})^T (\mathbf{A} + \mathbf{F})) \\ &= \text{trace}(\mathbf{A}^T \mathbf{A}) + 2 \text{trace}(\mathbf{A}^T \mathbf{F}) + \text{trace}(\mathbf{F}^T \mathbf{F}) \\ &= \sum_{j=1}^n \sigma_j(\mathbf{A})^2 + 2 \text{trace}(\mathbf{A}^T \mathbf{F}) + \text{trace}(\mathbf{F}^T \mathbf{F}). \end{aligned}$$

From the expressions for the elements of \mathbf{F} in (32) follows

$$\begin{aligned} \text{trace}(\mathbf{A}^T \mathbf{F}) &= \sum_{j=1}^n \sum_{i=1}^m a_{ij} f_{ij} = \epsilon_{\text{lower}} \sum_{j=1}^n \sum_{i=1}^m a_{ij}^2 \omega_{ij} \\ \text{trace}(\mathbf{F}^T \mathbf{F}) &= \sum_{j=1}^n \sum_{i=1}^m f_{ij}^2 = \epsilon_{\text{lower}}^2 \sum_{j=1}^n \sum_{i=1}^m a_{ij}^2 \omega_{ij}^2. \end{aligned}$$

Hence

$$\sum_{j=1}^n \sigma_j(\mathbf{A} + \mathbf{F})^2 = \sum_{j=1}^n \sigma_j(\mathbf{A})^2 + 2\epsilon_{\text{lower}} \sum_{j=1}^n \sum_{i=1}^m a_{ij}^2 \omega_{ij} + \epsilon_{\text{lower}}^2 \sum_{j=1}^n \sum_{i=1}^m a_{ij}^2 \omega_{ij}^2$$

Take the expectation of the above expression, and apply $\mathbb{E}[\omega_{ij}] = 0$ and $\mathbb{E}[\omega_{ij}^2] = \frac{1}{3}$ to obtain

$$\mathbb{E} \left[\sum_{j=1}^n \sum_{i=1}^m a_{ij}^2 \omega_{ij} \right] = 0, \quad \mathbb{E} \left[\sum_{j=1}^n \sum_{i=1}^m a_{ij}^2 \omega_{ij}^2 \right] = \frac{1}{3} \sum_{j=1}^n \sum_{i=1}^m a_{ij}^2 = \frac{1}{3} \|\mathbf{A}\|_F^2.$$

□

Remark 4.1. *Theorem 9 implies that the componentwise perturbation of a matrix induces a relative perturbation of its Frobenius norm,*

$$\mathbb{E} [\|\mathbf{A} + \mathbf{F}\|_F^2] = \left(1 + \frac{\epsilon_{\text{lower}}^2}{3} \right) \|\mathbf{A}\|_F^2.$$

We model the situation where \mathbf{A} has two singular value clusters: a cluster of r large singular values, and a cluster of $n - r$ very small singular values, under the assumption that the perturbation is sufficiently small to preserve the average large singular value.

Corollary 10. *Under the conditions of Theorem 9 assume that*

$$\sigma_r(\mathbf{A}) > \sigma_{r+1}(\mathbf{A}), \quad \mathbb{E} \left[\sum_{j=1}^r \sigma_j(\mathbf{A} + \mathbf{F})^2 \right] = \sum_{j=1}^r \sigma_j(\mathbf{A})^2,$$

and define the cluster averages of the small singular values as

$$\begin{aligned} \sigma_{\text{small}}(\mathbf{A})^2 &\equiv \frac{1}{n-r} \sum_{j=r+1}^n \sigma_j(\mathbf{A})^2, \\ \sigma_{\text{small}}(\mathbf{A} + \mathbf{F})^2 &\equiv \frac{1}{n-r} \sum_{j=r+1}^n \sigma_j(\mathbf{A} + \mathbf{F})^2. \end{aligned}$$

Then

$$\mathbb{E}[\sigma_{\text{small}}(\mathbf{A} + \mathbf{F})^2] = \sigma_{\text{small}}(\mathbf{A})^2 + \frac{\epsilon_{\text{lower}}^2}{3(n-r)} \|\mathbf{A}\|_F^2.$$

Proof. Inserting the assumption $\mathbb{E} \left[\sum_{j=1}^r \sigma_j(\mathbf{A} + \mathbf{F})^2 \right] = \sum_{j=1}^r \sigma_j(\mathbf{A})^2$ into the expectation from Theorem 9 gives

$$\mathbb{E} \left[\sum_{j=r+1}^n \sigma_j(\mathbf{A} + \mathbf{F})^2 \right] = \sum_{j=r+1}^n \sigma_j(\mathbf{A})^2 + \frac{\epsilon_{\text{lower}}^2}{3} \|\mathbf{A}\|_F^2.$$

At last divide both sides by $n - r$.

□

If the small singular values of \mathbf{A} are sufficiently small with $\sigma_{\text{small}}(\mathbf{A}) = \mathcal{O}(\epsilon_{\text{lower}}^2)$; and there are only a few of them, $r \lesssim n$; and \mathbf{A} is well conditioned with $\|\mathbf{A}\|_F \ll 1/\epsilon_{\text{lower}}$, then Corollary 10 suggests that the perturbation significantly increases the small singular values,

$$\mathbb{E}[\sigma_{\text{small}}(\mathbf{A} + \mathbf{F})^2] = \mathcal{O}(\epsilon_{\text{lower}}^2).$$

5 Numerical experiments

We present numerical experiments to illustrate that demotion to lower precision can increase small singular values, thus confirming that our bounds in sections 2 and 3 are informative models for the effects of reduced arithmetic precision.

After describing the algorithms for computing the singular values (Section 5.1), we present the numerical experiments (Section 5.2).

5.1 Generation and computation of singular values

The code for the numerical experiments consists of two algorithms: Algorithm 2 in Appendix A generates the exact singular values Σ , while Algorithm 1 generates the matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ from Σ in double precision, and then computes the singular values of \mathbf{A} , and those of the lower precision versions $\text{single}(\mathbf{A})$, and $\text{half}(\mathbf{A})$.

The n singular values Σ generated by Algorithm 2 consist of two clusters: a cluster Σ_1 of large singular values, and a cluster Σ_2 of small singular values. Each cluster is defined by the following input parameters: the number of singular values; the smallest and largest singular value; and the gap between the two clusters.

Specifically, cluster Σ_1 consists of $k_1 > 0$ singular values, the largest one being 10^{s_1} and the smallest one being $10^{s_1-d_1}$. Here $d_1 \geq 0$ controls the distance between smallest and largest singular value. If $d_1 > 0$ and $k_1 > 2$, then the interior singular values of Σ_1 are sampled uniformly at random in the interval $[10^{s_1-d_1}, 10^{s_1}]$.

The parameter g controls the gap 10^g between the clusters. Cluster Σ_2 consists of $k_2 \equiv n - k_1 \geq 0$ singular values, the largest one being $10^{s_1-d_1-g}$, and the smallest one being $10^{s_1-d_1-g-d_2}$, where $d_2 \geq 0$ controls the distance between smallest and largest singular value in cluster Σ_2 . If $d_2 > 0$ and $k_2 > 2$, then the interior singular values of Σ_2 are sampled uniformly at random in the interval $[10^{s_1-d_1-g-d_2}, 10^{s_1-d_1-g}]$.

5.2 Numerical results and discussion

In accordance with our bounds in sections 2 and 3, we present experiments for a single smallest singular value (section 5.2.1) and for a cluster of small singular values (section 5.2.2).

The matrices $\mathbf{A} \in \mathbb{R}^{m \times n}$ have $\text{rank}(\mathbf{A}) = n$ with $m = 4,096$ and $n = 256$. Changing the matrix dimensions while keeping the aspect ratio m/n constant does not change our conclusions. Figures 1–5 show the exact singular values, and the singular values computed in double, single, and half precision, with the exact increase in the smallest singular values listed in Table 1.

	$\min(\Sigma)$	$\min(\Sigma^d)$	$\min(\Sigma^s)$	$\min(\Sigma^h)$
Figure 1	10^{-3}	10^{-3}	10^{-3}	7×10^{-3}
Figure 2	10^{-5}	10^{-5}	5×10^{-4}	3×10^1
Figure 3	10^{-4}	10^{-4}	7×10^{-4}	5×10^0
Figure 4	10^{-7}	10^{-7}	5×10^{-5}	3×10^{-1}
Figure 5	10^{-7}	10^{-7}	4×10^{-4}	3×10^0

Table 1: Smallest singular values in Figures 1–5: exact (Σ); and double (Σ^d), single (Σ^s) and half (Σ^h) precisions.

5.2.1 A single smallest singular value

We present experiments to illustrate that demotion to lower precision can increase the smallest singular value, thus confirming that Theorem 5 represents a proper qualitative model for the effects of reduced arithmetic precision. In Figures 1–2, the small singular value cluster Σ_2 consists of a single singular value, while the large singular value cluster Σ_1 contains 255 singular values.

Figure 1 The cluster Σ_1 contains 255 distinct singular values in the interval $[10^{-2}, 10^2]$, while Σ_2 contains the single singular value 10^{-3} . The singular values in double and single precision are identical to the exact singular values. However, in half precision the smallest singular value has increased by almost an order of magnitude to 7×10^{-3} .

Figure 2 The cluster Σ_1 contains 255 distinct singular values in the interval $[10^{-3}, 10^6]$, while Σ_2 contains the single singular value 10^{-5} . The singular values in double precision are identical to the exact singular values. In single precision, the smallest singular value has increased a bit, to 5×10^{-4} , while in half precision it has increased by five orders of magnitude, to 33.6.

5.2.2 A cluster of small singular values

We present experiments to illustrate that demotion to lower precision can increase the smallest singular value cluster, thus confirming that Theorem 8 represents a proper qualitative model for the effects of reduced arithmetic precision. In Figures 3–4, the small singular value cluster Σ_2 contains 28 singular values, while the large singular value cluster Σ_1 contains 228 singular values.

Figure 3 The cluster Σ_1 contains 228 distinct singular values in the interval $[10, 10^5]$, while Σ_2 contains 28 singular values in the interval $[10^{-4}, 10^{-1}]$. The singular values in double precision are identical to the exact singular values. In single precision, the smallest singular value of Σ_2 has increased by two orders of magnitude, from 10^{-4} to 7×10^{-7} , while in half precision it has increased by more than four orders of magnitude, to about 5.

Figure 4 The cluster Σ_1 contains 228 distinct singular values in the interval $[10^{-3}, 10^4]$, while Σ_2 contains 28 singular values in the interval $[10^{-7}, 10^{-4}]$. The singular values in double precision are identical to the exact singular values. In single precision, the smallest singular value of Σ_2 has increased by an order of magnitude, from 10^{-7} to 5×10^{-5} , while in half precision it has increased by four orders of magnitude to about 3×10^{-1} .

Figure 5 The cluster Σ_1 contains 228 distinct singular values in the interval $[10^{-4}, 10^5]$, while Σ_2 contains 28 singular values in the interval $[10^{-6}, 10^{-7}]$. The singular values in double precision are identical to the exact singular values. In single precision the smallest singular value of Σ_2 has increased by more than two orders of magnitude, from 10^{-7} to 4×10^{-4} , and in half precision by seven orders of magnitude to about 3.

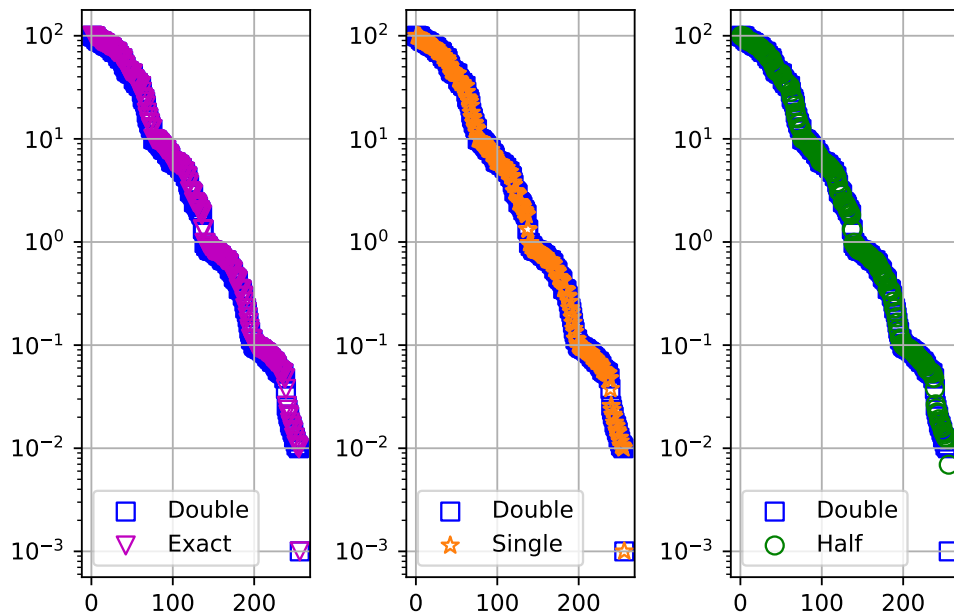


Figure 1: The matrix $\mathbf{A} \in \mathbb{R}^{4096 \times 256}$ has 255 distinct singular values in $[10^{-2}, 10^2]$, and a single small singular value 10^{-3} . All panels: Double precision singular values (squares). Left: Exact singular values (triangles). Middle: Single precision singular values (stars). Right: Half precision singular values (circle).

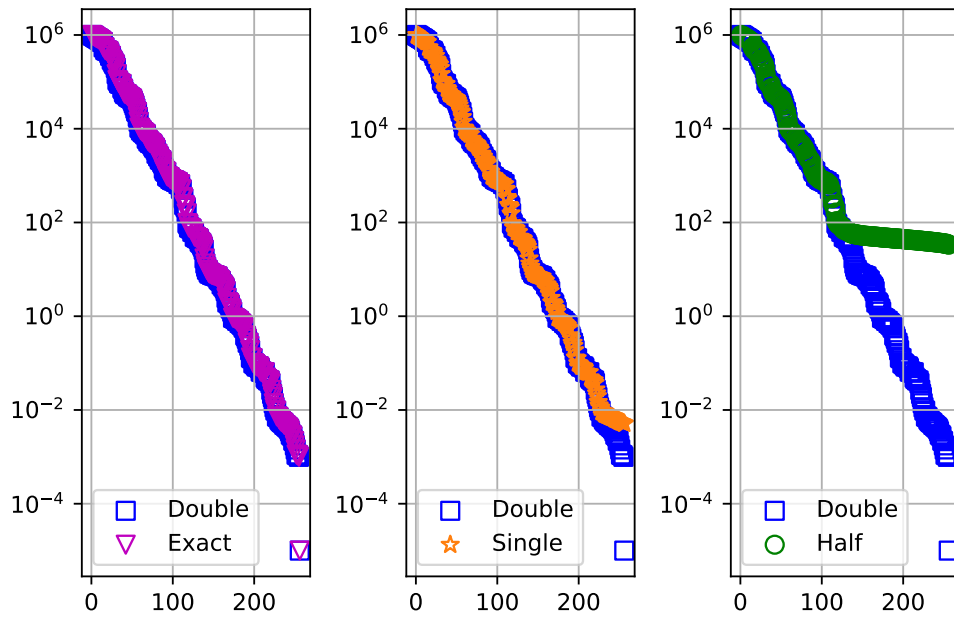


Figure 2: The matrix $\mathbf{A} \in \mathbb{R}^{4096 \times 256}$ has 255 distinct singular values in $[10^{-3}, 10^6]$, and a single small singular value 10^{-5} . All panels: Double precision singular values (squares). Left: Exact singular values (triangles). Middle: Single precision singular values (stars). Right: Half precision singular values (circle).

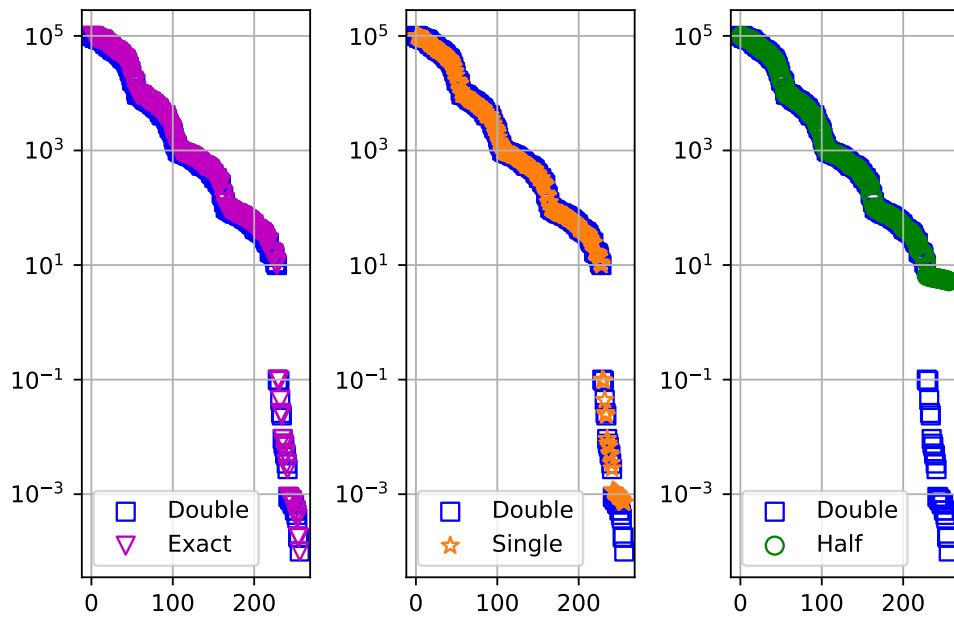


Figure 3: The matrix $\mathbf{A} \in \mathbb{R}^{4096 \times 256}$ has 228 distinct singular values in $[10, 10^5]$, and 28 distinct singular values in $[10^{-4}, 10^{-1}]$. All panels: Double precision singular values (squares). Left: Exact singular values (triangles). Middle: Single precision singular values (stars). Right: Half precision singular values (circle).

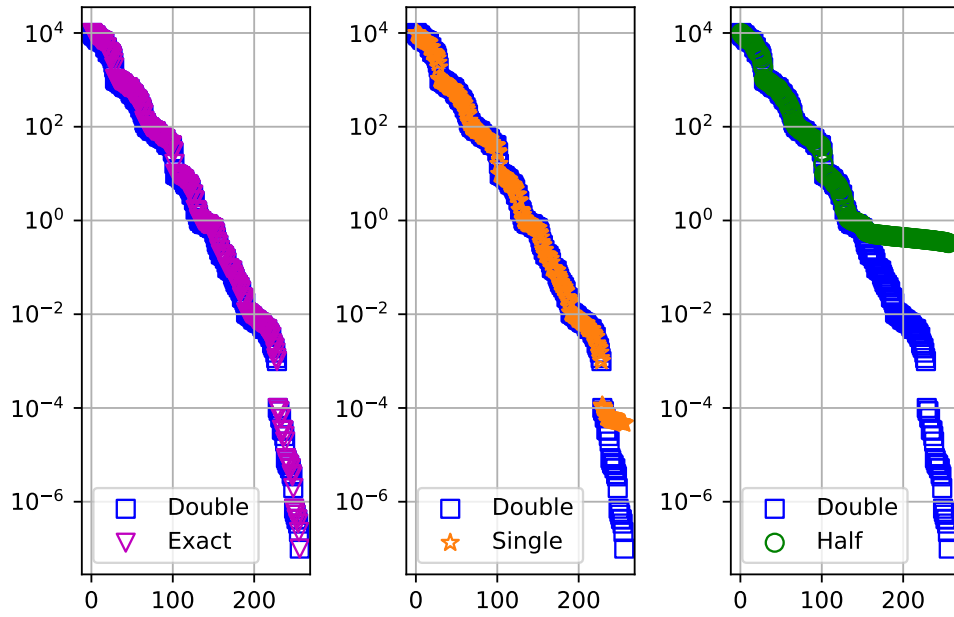


Figure 4: The matrix $\mathbf{A} \in \mathbb{R}^{4096 \times 256}$ has 228 distinct singular values in $[10^{-3}, 10^4]$, and 28 distinct singular values in $[10^{-7}, 10^{-4}]$. All panels: Double precision singular values (squares). Left: Exact singular values (triangles). Middle: Single precision singular values (stars). Right: Half precision singular values (circle).

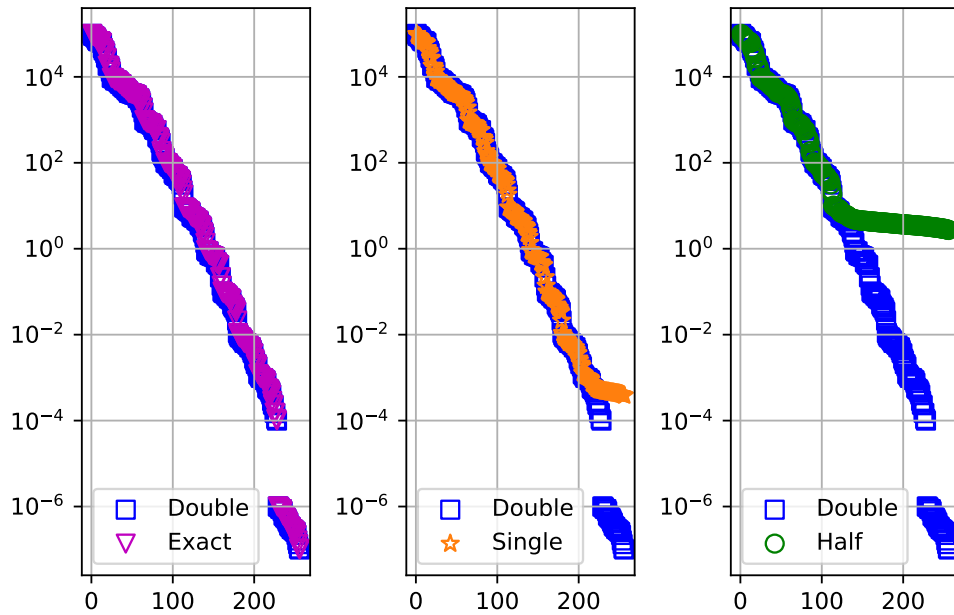


Figure 5: The matrix $\mathbf{A} \in \mathbb{R}^{4096 \times 256}$ has 228 distinct singular values in $[10^{-4}, 10^5]$, and 28 distinct singular values in $[10^{-6}, 10^{-7}]$. All panels: Double precision singular values (squares). Left: Exact singular values (triangles). Middle: Single precision singular values (stars). Right: Half precision singular values (circle).

6 Future Work

The previous sections investigate the change in the computed singular values of a full column-rank matrix \mathbf{A} after it has been demoted to a lower arithmetic precision. Our deterministic and probabilistic lower bounds in Theorems 1 and 2 represent a *qualitative* model for the increase in the smallest singular values of the perturbed matrix $\mathbf{A} + \mathbf{E}$, which is confirmed by the experiments in section 5.]

Future work will consist of a *quantitative* analysis to determine the exact order of magnitude of the increase in the small singular values and the structural properties of \mathbf{A} that can contribute to it, including specifically the size of the gap that separates the small singular values from the larger singular values; and the condition number of \mathbf{A} with respect to left inversion. In addition, the influence of the third order perturbation terms needs to be investigated, as they might possibly become dominant for ill-conditioned matrices \mathbf{A} .

References

- [Arm10] D. Armentano. “Stochastic perturbations and smooth condition numbers”. In: *J. Complexity* 26.2 (2010), pp. 161–171. ISSN: 0885-064X.
- [CD22] E. Carson and I. Daužickaitė. *Single-pass Nystrom approximation in mixed precision*. arXiv:2202.13355. 2022. DOI: 10.48550/ARXIV.2205.13355. URL: <https://arxiv.org/abs/2205.13355>.
- [Coo18] N. Cook. “Lower bounds for the smallest singular value of structured random matrices”. In: *Ann. Probab.* 46.6 (2018), pp. 3442–3500. DOI: 10.1214/17-AOP1251. URL: <https://doi.org/10.1214/17-AOP1251>.
- [Cuc16] F. Cucker. “Probabilistic analyses of condition numbers”. In: *Acta Numer.* 25 (2016), pp. 321–382. ISSN: 0962-4929.
- [Dem88] J. W. Demmel. “The probability that a numerical analysis problem is difficult”. In: *Math. Comp.* 50.182 (1988), pp. 449–480. DOI: 10.2307/2008617. URL: <https://doi.org/10.2307/2008617>.
- [Drm17] Zlatko Drmač. “Algorithm 977: a QR-preconditioned QR SVD method for computing the SVD with high accuracy”. In: *ACM Trans. Math. Software* 44.1 (2017), Art. 11, 30. DOI: 10.1145/3061709. URL: <https://doi.org/10.1145/3061709>.
- [DV07a] Z. Drmač and K. Veselić. “New fast and accurate Jacobi SVD algorithm. I”. In: *SIAM J. Matrix Anal. Appl.* 29.4 (2007), pp. 1322–1342. DOI: 10.1137/050639193. URL: <https://doi.org/10.1137/050639193>.
- [DV07b] Z. Drmač and K. Veselić. “New fast and accurate Jacobi SVD algorithm. II”. In: *SIAM J. Matrix Anal. Appl.* 29.4 (2007), pp. 1343–1362. DOI: 10.1137/05063920X. URL: <https://doi.org/10.1137/05063920X>.
- [GV13] G. H. Golub and C. F. Van Loan. *Matrix computations*. Fourth. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 2013.
- [HJ13] R. A. Horn and C. R. Johnson. *Matrix analysis*. Second. Cambridge University Press, Cambridge, 2013.
- [HP92] Y. P. Hong and C.-T. Pan. “A lower bound for the smallest singular value”. In: *Linear Algebra Appl.* 172 (1992), pp. 27–32. ISSN: 0024-3795. DOI: 10.1016/0024-3795(92)90016-4. URL: [https://doi.org/10.1016/0024-3795\(92\)90016-4](https://doi.org/10.1016/0024-3795(92)90016-4).
- [Hua08] T.-Z. Huang. “Estimation of $\|A^{-1}\|_\infty$ and the smallest singular value”. In: *Comput. Math. Appl.* 55.6 (2008), pp. 1075–1080. ISSN: 0898-1221. DOI: 10.1016/j.camwa.2007.04.036. URL: <https://doi.org/10.1016/j.camwa.2007.04.036>.
- [Joh89] C. R. Johnson. “A Gersgorin-type lower bound for the smallest singular value”. In: *Linear Algebra Appl.* 112 (1989), pp. 1–7. ISSN: 0024-3795. DOI: 10.1016/0024-3795(89)90583-1. URL: [https://doi.org/10.1016/0024-3795\(89\)90583-1](https://doi.org/10.1016/0024-3795(89)90583-1).

- [JS98] C. R. Johnson and T. Szulc. “Further lower bounds for the smallest singular value”. In: *Linear Algebra Appl.* 272 (1998), pp. 169–179. ISSN: 0024-3795. DOI: 10.1016/S0024-3795(97)00330-3. URL: [https://doi.org/10.1016/S0024-3795\(97\)00330-3](https://doi.org/10.1016/S0024-3795(97)00330-3).
- [KLR98] C. S. Kenney, A. J. Laub, and M. S. Reese. “Statistical condition estimation for linear systems”. In: *SIAM J. Sci. Comput.* 19.2 (1998), pp. 566–583. ISSN: 1064-8275.
- [Li20] C. Li. “Schur complement-based infinity norm bounds for the inverse of SDD matrices”. In: *Bull. Malays. Math. Sci. Soc.* 43.5 (2020), pp. 3829–3845. ISSN: 0126-6705. DOI: 10.1007/s40840-020-00895-x. URL: <https://doi.org/10.1007/s40840-020-00895-x>.
- [LN20] M. Lotz and V. Noferini. “Wilkinson’s bus: weak condition numbers, with an application to singular polynomial eigenproblems”. In: *Found. Comput. Math.* 20.6 (2020), pp. 1439–1473.
- [LX21] M. Lin and M. Xie. “On some lower bounds for smallest singular value of matrices”. In: *Appl. Math. Lett.* 121 (2021), Paper No. 107411, 7. ISSN: 0893-9659. DOI: 10.1016/j.aml.2021.107411. URL: <https://doi.org/10.1016/j.aml.2021.107411>.
- [Par80] B. N. Parlett. *The Symmetric Eigenvalue Problem*. Englewood Cliffs: Prentice Hall, 1980.
- [Rum09] S. M. Rump. “Inversion of extremely ill-conditioned matrices in floating-point”. In: *Japan J. Indust. Appl. Math.* 26.2-3 (2009), pp. 249–277. ISSN: 0916-7005. URL: <http://projecteuclid.org/euclid.jjiam/1265033781>.
- [San21] C. Sang. “Schur complement-based infinity norm bounds for the inverse of $DSDD$ matrices”. In: *Bull. Iranian Math. Soc.* 47.5 (2021), pp. 1379–1398. ISSN: 1017-060X. DOI: 10.1007/s41980-020-00447-w. URL: <https://doi.org/10.1007/s41980-020-00447-w>.
- [Shu22] X. Shun. “Two new lower bounds for the smallest singular value”. In: *J. Math. Inequal.* 16.1 (2022), pp. 63–68. DOI: 10.7153/jmi-2022-16-05. URL: <https://doi.org/10.7153/jmi-2022-16-05>.
- [SS90] G. W. Stewart and J. G. Sun. *Matrix perturbation theory*. Computer Science and Scientific Computing. Academic Press, Inc., Boston, MA, 1990.
- [SST06] A. Sankar, D. A. Spielman, and S.-H. Teng. “Smoothed analysis of the condition numbers and growth factors of matrices”. In: *SIAM J. Matrix Anal. Appl.* 28.2 (2006), pp. 446–476. ISSN: 0895-4798. DOI: 10.1137/S0895479803436202. URL: <https://doi.org/10.1137/S0895479803436202>.
- [Ste06] Michael Stewart. “Perturbation of the SVD in the presence of small singular values”. In: *Linear Algebra Appl.* 419.1 (2006), pp. 53–77. ISSN: 0024-3795. DOI: 10.1016/j.laa.2006.04.013.
- [Ste84] G. W. Stewart. “A second order perturbation expansion for small singular values”. In: *Linear Algebra Appl.* 56 (1984), pp. 231–235.
- [Ste90] G. W. Stewart. “Stochastic perturbation theory”. In: *SIAM Rev.* 32.4 (1990), pp. 579–610. ISSN: 0036-1445.
- [TV07] T. Tao and V. Vu. “The condition number of a randomly perturbed matrix”. In: *STOC’07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*. ACM, New York, 2007, pp. 248–255. DOI: 10.1145/1250790.1250828. URL: <https://doi.org/10.1145/1250790.1250828>.
- [TV09] T. Tao and V. Vu. “Smooth analysis of the condition number and the least singular value”. In: *Approximation, randomization, and combinatorial optimization*. Vol. 5687. Lecture Notes in Comput. Sci. Springer, Berlin, 2009, pp. 714–737.
- [TV10] T. Tao and V. Vu. “Smooth analysis of the condition number and the least singular value”. In: *Math. Comp.* 79.272 (2010). DOI: 10.1090/S0025-5718-2010-02396-8. URL: <https://doi.org/10.1090/S0025-5718-2010-02396-8>.
- [Var75] J. M. Varah. “A lower bound for the smallest singular value of a matrix”. In: *Linear Algebra Appl.* 11 (1975), pp. 3–5. ISSN: 0024-3795. DOI: 10.1016/0024-3795(75)90112-3. URL: [https://doi.org/10.1016/0024-3795\(75\)90112-3](https://doi.org/10.1016/0024-3795(75)90112-3).
- [YG97] Y. Yu and D. Gu. “A note on a lower bound for the smallest singular value”. In: *Linear Algebra Appl.* 253 (1997), pp. 25–38. ISSN: 0024-3795. DOI: 10.1016/0024-3795(95)00784-9. URL: [https://doi.org/10.1016/0024-3795\(95\)00784-9](https://doi.org/10.1016/0024-3795(95)00784-9).

[Zou12] L. Zou. “A lower bound for the smallest singular value”. In: *J. Math. Inequal.* 6.4 (2012), pp. 625–629. ISSN: 1846-579X. DOI: 10.7153/jmi-06-60. URL: <https://doi.org/10.7153/jmi-06-60>.

A Algorithms

We present pseudo codes for two algorithms: The function `create_sigmas` in Algorithm 2 computes the singular values Σ according to the specifications in the input parameters `params`. Algorithm 1 constructs \mathbf{A} from Σ in double precision, and then computes the singular values Σ^d of \mathbf{A} , Σ^s of `single(A)`, and Σ^h of `half(A)`. If $d_1 = 0$ or $d_2 = 0$ in Algorithm 2, then the cluster Σ_1 or Σ_2 consists of a single singular value of multiplicity k_1 or k_2 , respectively.

Algorithm 1 Singular values of \mathbf{A} , `single(A)` and `half(A)`

Input: Large matrix dimension m , `params`

Output: Singular values of \mathbf{A} in double, single, half precision

$\Sigma \leftarrow \text{create_sigmas}(\text{params})$ {Exact singular values}

$n \leftarrow \text{length}(\Sigma)$ {Small dimension of \mathbf{A} }

$[U, S, V] \leftarrow \text{SVD}(\text{randn}(m, n))$ {Left and right singular vectors for \mathbf{A} }

$\mathbf{A} \leftarrow U\Sigma V^T$ {Compute \mathbf{A} in double precision}

$\Sigma^d \leftarrow \text{SVD}(\mathbf{A})$ {Singular values of double precision \mathbf{A} }

$\Sigma^s \leftarrow \text{SVD}(\text{double}(\text{single}(\mathbf{A})))$ {Singular values of single precision \mathbf{A} }

$\Sigma^h \leftarrow \text{SVD}(\text{double}(\text{half}(\mathbf{A})))$ {Singular values of half precision \mathbf{A} }

return $\Sigma, \Sigma^d, \Sigma^s, \Sigma^h$

Algorithm 2 Exact singular values: function `create_sigmas`

Input: `params = {s1, g, k1, k2, d1, d2}`

Output: Exact singular values Σ

$\Sigma \leftarrow \mathbf{zeros}(k_1 + k_2, 1)$ {Initialize vector of all singular values}

$\Sigma_1 \leftarrow \mathbf{zeros}(k_1, 1)$ {Initialize cluster of large singular values}

$\Sigma_2 \leftarrow \mathbf{zeros}(k_2, 1)$ {Initialize cluster of small singular values}

$\Sigma_1(1) \leftarrow 10^{s_1}$ {Largest singular value}

if $k_1 > 1$ **then**

$\Sigma_1(k_1) \leftarrow 10^{s_1 - d_1}$ {Smallest singular value in Σ_1 }

end if

{Uniform sampling of interior singular values in cluster Σ_1 }

for $j = 2 : k_1 - 1$ **do**

$\Sigma_1(j) \leftarrow \mathbf{Uniform}([\Sigma_1(k_1), \Sigma_1(1)])$

end for

$\Sigma_2(1) \leftarrow 10^{s_1 - d_1 - g}$ {Largest singular value in Σ_2 }

if $k_2 > 1$ **then**

$\Sigma_2(k_2) \leftarrow 10^{s_1 - d_1 - g - d_2}$ {Smallest singular value in Σ_2 }

end if

{Uniform sampling of interior singular values in cluster Σ_2 }

for $j = 2 : k_2 - 1$ **do**

$\Sigma_2(j) \leftarrow \mathbf{Uniform}([\Sigma_2(k_2), \Sigma_2(1)])$

end for

$\Sigma \leftarrow [\Sigma_1, \Sigma_2]$ {Concatenate the two singular value clusters}

return `sort(Σ)` {Return sorted singular values in non-ascending order}
