Small Singular Values *can Increase* in Lower Precision

Petros Drineas (Purdue CS)

Joint work with I. Ipsen (NCSU) & C. Boutsikas, G. Dexter, L. Ma (Purdue)

PURDUE UNIVERSITY®

Small Singular Values *Increase* in Lower Precision

Petros Drineas (Purdue CS)

Joint work with I. Ipsen (NCSU) & C. Boutsikas, G. Dexter, L. Ma (Purdue)

Small Singular Values *Increase* in Lower Precision

[for sufficiently tall-and-thin matrices; using stochastic rounding; with high probability; etc.]

Petros Drineas (Purdue CS)

Joint work with I. Ipsen (NCSU) & C. Boutsikas, G. Dexter, L. Ma (Purdue)

PURDUE UNIVERSITY®

Motivation

Iron Law

All numbers used in a computer shall have a fixed number of digits. Therefore, the output of (almost) all primitive operations executed in a computer are wrong.

- **Major concern:** These *roundoff* errors accumulate and could be catastrophic[1].
- Turing Award (1970) to J. H. Wilkinson for his work in linear algebraic computations and backward error analysis.

[1] Anecdotally, a *very* promiment numerical analyst was hesitant to fly after they found out that computers (and, therefore, roundoff errors) were involved in aircraft design and flight planning...

Motivation, cont'd

Iron Law

All numbers used in a computer shall have a fixed number of digits. Therefore, the output of (almost) all primitive operations executed in a computer are wrong.

- We need to *round* numbers in order to be stored/represented/used by a computer.
- We think of this *rounding* process as a *deficiency*, since it leads to errors.

Could rounding be a *blessing* for 21st century computing?

Computing in the 21st century

Data Science, Machine Learning, and Artificial Intelligence *dominate modern computing*.

- Data are noisy and highly accurate computations could result in overfitting[2].
- *Regularization* is fundamental in DS/ML/AI algorithms.
- **Rounding is a form of implicit regularization!**

[2] ...to irrelevancies, according to Michael W. Mahoney.

Research Topic

Rounding and the smallest singular value of a matrix

Given a matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$ (exact representation), what happens to its smallest singular value after rounding $\mathbf{A}$ to $\tilde{\mathbf{A}} \in \mathcal{F}^{n \times d}$?

- Here $\mathcal{F}$ could be, for example, the set of all *double*, *single*, or *half* precision numbers.

Prior knowledge

Large singular values remain unharmed, but small singular values tend to increase.

See, for example, [Stewart & Sun, 1990, pg. 266]

> *"...small singular values tend to increase" [under small perturbations]*

and [Rump, 2009, pg. 261]

> *"...even an approximate inverse of an arbitrarily ill-conditioned matrix does, in general, contain useful information. This is due to a kind of regularization by rounding to working precision."*

Rounding as a perturbation

A straight-forward approach

- Model rounding error as a perturbation $\mathbf{E}$
- Formally, $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{E}$
- Use perturbation theory to get bounds

What does Weyl's inequality reveal about the small singular values?

- If the largest singular value of $\mathbf{E}$ ("noise" due to rounding) is larger than the smallest singular value of $\mathbf{A}$, not much...

$$\underbrace{\sigma_{\min}(\mathbf{A}) - \|\mathbf{E}\|_2}_{\text{trivial if } \leq 0} \leq \sigma_{\min}(\underbrace{\mathbf{A} + \mathbf{E}}_{\tilde{\mathbf{A}}})$$

A higher order expansion

[Boutsikas, Drineas, and Ipsen SIMAX '24]

(Building upon [G. W. Stewart LAA '84])

- Partition the $n \times d$ matrices $\mathbf{A}$ and $\mathbf{E}$
- ($\mathbf{\Sigma}_1$ is $(d-1) \times (d-1)$)

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

A higher order expansion

[Boutsikas, Drineas, and Ipsen SIMAX '24]

- Assume
  1. Large singular values are large: $\sigma_{d-1} > 4\|\mathbf{E}\|_2$
  2. A single small singular value: $\sigma_d < \|\mathbf{E}\|_2$
- We prove[3]

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

- r_3, r_4 contains terms of $\mathcal{O}(\|\mathbf{E}\|_2^3)$ or higher

$$r_3 = 2\mathbf{e}_{12}^T \underbrace{(\mathbf{\Sigma}_1 + \mathbf{E}_{11})^{-T} \begin{pmatrix} \mathbf{e}_{21} & \mathbf{E}_{31}^T \end{pmatrix} \begin{pmatrix} e_{22} + \sigma_d \\ \mathbf{e}_{32} \end{pmatrix}}_{\mathbf{w}}$$

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

[3] We also prove a generalized version of this result for clusters of small singular values.

Pros & Cons

Pros

- True lower bound (beyond second order)
- Assumes a small gap between σ_{d-1}, σ_d
- Numerical experiments confirm our theory

Cons

- The higher order terms are challenging to interpret

Pros & Cons

Pros

- True lower bound (beyond second order)
- Assumes a small gap between σ_{d-1}, σ_d
- Numerical experiments confirm our theory

Cons

- The higher order terms are challenging to interpret

Let's use *Randomized Algorithms*, specifically *Stochastic Rounding (SR)*.

Normalized FP numbers

FP Model

- Given a basis β and a precision p

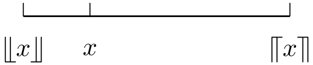
$$x = (-1)^s \cdot m \cdot \beta^{e-p}$$

- s is the sign bit
- e is the exponent
- The significand m is an integer in

$$\beta^{p-1} \leq m \leq \beta^p$$

Properties

- Let $\mathcal{F}$ be the set of normalized FP numbers and let $x \in \mathbb{R} - \mathcal{F}$
- The two FP numbers enclosing x are denoted by $\lfloor\!\lfloor x \rfloor\!\rfloor$, $\lceil\!\lceil x \rceil\!\rceil$



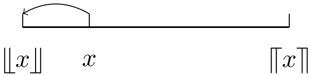
- The following inequality holds:

$$\max\{x - \lfloor\!\lfloor x \rfloor\!\rfloor, \lceil\!\lceil x \rceil\!\rceil - x\} \leq \beta^{1-p}|x|$$

Deterministic vs Stochastic Rounding (SR)

Deterministic

- Round-to-Nearest (RN)

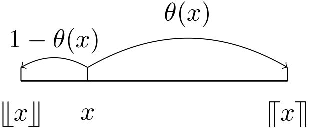


- For RN

$$\max\left\{x - \llfloor x \rrfloor, \llceil x \rrceil - x\right\} \leq {}^{1}\!/_{2}\,\beta^{1-p}|x|$$

Stochastic

- $\theta(x) = \dfrac{x - \llfloor x \rrfloor}{\llceil x \rrceil - \llfloor x \rrfloor}$
- SR - *nearness*:

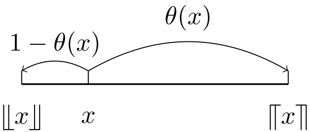


- Property: $\mathbb{E}\left[\mathtt{SR}(x)\right] = x$

Stochastic Rounding (SR)

Stochastic Rounding

- $\theta(x) = \dfrac{x - \lfloor\!\lfloor x \rfloor\!\rfloor}{\lceil\!\lceil x \rceil\!\rceil - \lfloor\!\lfloor x \rfloor\!\rfloor}$
- SR - *nearness*:



- Property: $\mathbb{E}\left[\mathrm{SR}(x)\right] = x$

History:

- Can be traced back to Forsythe 1950
- Also von Neumann & Goldstine 1947
- Recent resurgence: increasing interest for low-precision FP arithmetic for ML and DNNs [Gupta *et al.* 2015]
- Many patents held by (GPU) chip designers
- Review: Croci *et al.* 2022

SR: A simple example

Why SR?

- Let $\mathcal{F} = \{0, 1\}$ and consider the rank one matrix

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \\ \vdots & \vdots \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \xrightarrow{\texttt{RN}(\mathbf{A})} \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix}$$

- Any deterministic rounding will result to a rounded matrix $\tilde{\mathbf{A}}$ that is also rank one.

SR: A simple example

[Dexter, Boutsikas, Ma, Ipsen, and Drineas ArXiv '24]

This is not the case for SR

- Let $\mathcal{F} = \{0, 1\}$ and consider the rank one matrix

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \\ \vdots & \vdots \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \xrightarrow{\texttt{SR}(\mathbf{A})} \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix} = \tilde{\mathbf{A}}$$

- We can prove that for such $n \times 2$ matrices (with probability at least 0.997)

$$\sigma_{\min}\left(\widetilde{\mathbf{A}}\right) \gtrapprox {}^{1}/_{2}\sqrt{n}$$

Our bound

[Dexter, Boutsikas, Ma, Ipsen, and Drineas ArXiv '24]

For simplicity, assume $\mathbf{A} \in [-1, 1]^{n \times d}$ and let $\tilde{\mathbf{A}}$ be the stochastically rounded $\mathbf{A}$.

$$\sigma_{\min}\left(\widetilde{\mathbf{A}}\right) \geq$$

Model

- $\mathbf{A} \in^{n \times d}$ with $n \gg d$
- SR to FP numbers
- $\mathbf{E} = \widetilde{\mathbf{A}} - \mathbf{A}$
- $\mathbb{E}[\mathbf{E}] = \mathbf{0}$

Ingredients

- β is the basis of our FP arithmetic
- p is the working precision

Our bound

[Dexter, Boutsikas, Ma, Ipsen, and Drineas ArXiv '24]

For simplicity, assume $\mathbf{A} \in [-1,1]^{n\times d}$ and let $\tilde{\mathbf{A}}$ be the stochastically rounded $\mathbf{A}$.

$$\sigma_{\min}\left(\widetilde{\mathbf{A}}\right) \geq \beta^{1-p}\sqrt{n}\left(\sqrt{\nu} - \varepsilon_{n,d}\right)$$

Model

- $\mathbf{A} \in^{n\times d}$ with $n \gg d$
- SR to FP numbers
- $\mathbf{E} = \widetilde{\mathbf{A}} - \mathbf{A}$
- $\mathbb{E}[\mathbf{E}] = \mathbf{0}$

Ingredients

- β is the basis of our FP arithmetic
- p is the working precision
- ν measures the amount of available randomness during the rounding process
- $\varepsilon_{n,d}$ captures *lower-order* terms

Our bound: ν

[Dexter, Boutsikas, Ma, Ipsen, and Drineas ArXiv '24]

For simplicity, assume $\mathbf{A} \in [-1,1]^{n\times d}$ and let $\tilde{\mathbf{A}}$ be the stochastically rounded $\mathbf{A}$.

$$\sigma_{\min}\left(\widetilde{\mathbf{A}}\right) \geq \beta^{1-p}\sqrt{n}\left(\sqrt{\nu}-\varepsilon_{n,d}\right)$$

Model

- $\mathbf{A} \in^{n\times d}$ with $n \gg d$
- SR to FP numbers
- $\mathbf{E} = \widetilde{\mathbf{A}} - \mathbf{A}$
- $\mathbb{E}[\mathbf{E}] = \mathbf{0}$

Understanding ν

- Consider a matrix with, say, two identical columns whose entries are FPs: $\sigma_{\min}(\mathbf{A}) = 0$.
- SR will not modify those columns: $\sigma_{\min}(\tilde{\mathbf{A}}) = 0$.
- Intuitively: no randomness for SR to exploit.
- This lack of randomness is captured by ν, which, in this case, is equal to zero.

Our bound: ν

[Dexter, Boutsikas, Ma, Ipsen, and Drineas ArXiv '24]

For simplicity, assume $\mathbf{A} \in [-1,1]^{n\times d}$ and let $\tilde{\mathbf{A}}$ be the stochastically rounded $\mathbf{A}$.

$$\sigma_{\min}\left(\widetilde{\mathbf{A}}\right) \geq \beta^{1-p}\sqrt{n}\left(\sqrt{\nu} - \varepsilon_{n,d}\right)$$

Model

- $\mathbf{A} \in^{n\times d}$ with $n \gg d$
- SR to FP numbers
- $\mathbf{E} = \widetilde{\mathbf{A}} - \mathbf{A}$
- $\mathbb{E}[\mathbf{E}] = \mathbf{0}$

Understanding ν

- Formally[a]: $\nu \propto \min_{\text{all columns } j} \sum_{i=1}^{n} \mathrm{Var}\left(\mathbf{E}_{ij}\right)$
- $0 \leq \nu \leq 1$

[a]Skipping a normalization factor

Interpreting our bound

For simplicity, assume $\mathbf{A} \in [-1,1]^{n\times d}$ and let $\tilde{\mathbf{A}}$ be the stochastically rounded $\mathbf{A}$.

$$\sigma_{\min}\left(\widetilde{\mathbf{A}}\right) \geq \beta^{1-p}\sqrt{n}\left(\sqrt{\nu} - \varepsilon_{n,d}\right)$$

- As n grows, $\sigma_{\min}\left(\widetilde{\mathbf{A}}\right)$ increases
- β^{1-p} captures the parameters of FP arithmetic
- ν captures the amount of available *stochasticity* in $\mathtt{SR}(\mathbf{A})$
- $\varepsilon_{n,d}$ depends only on n, d:
  - $\rightarrow$ If n is $\omega(d^4)$, then $\lim_{n\to\infty} \varepsilon_{n,d} = 0$.
  - $\rightarrow$ (hiding $\log$ factors)

Our main result: A perturbation theory bound

Main Theorem

Let $\mathbf{A}$ and $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{E}$ be real $n \times d$ matrices. Here $\mathbf{E}$ models a zero-mean random perturbation matrix with minimal (normalized) column variance ν and $\max_{i,j} |\mathbf{E}_{ij}| \leq \mathbf{R}$.

If $n \geq 836$, then with probability at least $1 - {}^{1}/_{n^c} - {}^{2d^2}/_{n^2}$,

$$\sigma_{\min}(\widetilde{\mathbf{A}}) \geq \mathbf{R}\sqrt{n}(\sqrt{\nu} - \varepsilon_{n,d}),$$

where

$$\varepsilon_{n,d} \equiv \sqrt{\frac{d}{n}} + 2d^2\sqrt{\frac{\log n}{n}} + \frac{C(\log n)^{2/3}}{n^{1/30}} \cdot \left(\frac{d}{n}\right)^{\frac{1}{54}},$$

and c and C are absolute constants.

Our bound is tight

[Dexter, Boutsikas, Ma, Ipsen, and Drineas ArXiv '24]

Tightness of our bound

Let $\mathbf{A}$ and $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{E}$ be real $n \times d$ matrices. Here $\mathbf{E}$ models a zero-mean random perturbation matrix with minimal (normalized) column variance ν and $\max_{i,j} |\mathbf{E}_{ij}| \leq \mathbf{R}$.

Our main bound is that, with high probability,

$$\sigma_{\min}(\widetilde{\mathbf{A}}) \gtrapprox \mathbf{R}\sqrt{n\nu}.$$

We exhibit $n \times d$ matrices $\mathbf{A}$ for which SR returns the matrix $\widetilde{\mathbf{A}}$ such that

$$\sigma_{\min}\left(\widetilde{\mathbf{A}}\right) \leq \left(1 + \sqrt{1/(d-1)}\right) \cdot \mathbf{R}\sqrt{n\nu}.$$

Proof outline

Steps:

1. We introduce the orthogonal projector $\mathbf{P}_{\mathbf{A}}$ onto the left column space of $\mathbf{A}$. This allows us to focus on $\mathbf{P}_{\mathbf{A}}\mathbf{E}$.
2. Weyl's inequality yields a lower bound on the smallest singular value of $(\mathbf{I} - \mathbf{P}_{\mathbf{A}})\mathbf{E}$ by lower bounding the smallest singular value of $\mathbf{E}$ and upper bounding the largest singular value of $\mathbf{P}_{\mathbf{A}}\mathbf{E}$.
3. Application of a Random Matrix Theory bound from [Dumitriu & Zhu '23] shows that the smallest singular value of $\mathbf{E}$ is sufficiently large.
4. The largest singular value of the projection $\mathbf{P}_{\mathbf{A}}\mathbf{E}$ is small, because $\mathbf{P}_{\mathbf{A}}$ projects $\mathbf{E}$ on the low-dimensional subspace of dimension d.
5. Standard measure concentration bounds show that $\mathbf{E}$ *does not concentrate* in any low-dimensional subspace.
6. Finally, we combine the bounds for the smallest singular value of $\mathbf{E}$ and the largest singular value of $\mathbf{P}_{\mathbf{A}}\mathbf{E}$.

Experiments (1)

Our universe

- $\mathbf{A} \in [-1, 1]^{n \times d}$
- All elements of $\mathtt{SR}(\mathbf{A}) \in \mathcal{F}^{\{p\}}$, $1 \leq p \leq 5$

$$\mathcal{F}^{\{p\}} = \{\pm\, {}^{m}/_{10^p}, \quad \text{for all integers } m = \underbrace{0, 1, 2, \ldots, 10^p - 1\}}_{\leq p \text{ digits}} \;\cup\; \{\pm 1\}$$

Setting (1)

- $\sigma_{\min}(\mathbf{A}) = 0$
- $n = 10^4;\ 10^5;\ 10^6$ and $d = 10;\ 100;\ 1000$
- For a fixed d, all $\mathbf{A}$ have the same singular values

Experiments (1) [recall: $n \times d$ matrix A and $\sigma_{\min}(\mathrm{A}) = 0$]

Each entry in the tables shows the pair of values $\left(\sigma_{\min}(\widetilde{\mathbf{A}}), \mathbf{R}\sqrt{n\nu}\right)$

Precision $\mathbf{p} = \mathbf{1}$

$\mathbf{d}$; $\mathbf{n}$	$\mathbf{10^4}$	$\mathbf{10^6}$
$\mathbf{10}$	$\mathbf{(4.11, 4.08)}$	$\mathbf{(34.11, 32.96)}$
$\mathbf{10^2}$	$\mathbf{(4.08, 4.07)}$	$\mathbf{(32.76, 32.46)}$
$\mathbf{10^3}$	$\mathbf{(3.84, 4.05)}$[a]	$\mathbf{(33.26, 33.03)}$

[a] Square-ish matrix

Precision $\mathbf{p} = \mathbf{3}$

$\mathbf{d}$; $\mathbf{n}$	$\mathbf{10^4}$	$\mathbf{10^6}$
$\mathbf{10}$	$\mathbf{(0.04, 0.04)}$	$\mathbf{(0.41, 0.41)}$
$\mathbf{10^2}$	$\mathbf{(0.04, 0.04)}$	$\mathbf{(0.41, 0.41)}$
$\mathbf{10^3}$	$\mathbf{(0.039, 0.04)}$[a]	$\mathbf{(0.41, 0.41)}$

[a] Square-ish matrix

Experiments (2)

Our universe

- $\mathbf{A} \in [-1, 1]^{n \times d}$
- All elements of $\texttt{SR}(\mathbf{A}) \in \mathcal{F}^{\{p\}}$, $1 \leq p \leq 5$

$$\mathcal{F}^{\{p\}} = \{\pm\, {}^{m}/_{10^p}, \quad \text{for all integers } m = \underbrace{0, 1, 2, \ldots, 10^p - 1}_{\leq p \text{ digits}}\} \;\cup\; \{\pm 1\}$$

Setting (2)

- $\mathbf{A}^h$ with $\nu \approx 1$ *(high value)*
- $\mathbf{A}^l$ with $\nu \approx 5 \cdot 10^{-4}$ *(low value)*
- $\sigma_{\min}(\mathbf{A}^h) = \sigma_{\min}(\mathbf{A}^l) = 0$
- Fixed $n = 10^4$ and $d = 10;\ 100;\ 1000$

Experiments (2) [recall: $10^4 \times d$ matrix A and $\sigma_{\min}(\mathrm{A}) = 0$]

Each entry in the tables shows the pair of values $\left(\sigma_{\min}(\widetilde{\mathbf{A}}), \mathbf{R}\sqrt{n\nu}\right)$; $n = 10^4$ **fixed**

Precision $\mathbf{p} = \mathbf{1}$

$\mathbf{d}$; ν	(high) $\mathbf{1}$	(low) $\mathbf{5 \cdot 10^{-4}}$
$\mathbf{10}$	$(\mathbf{5.01}, \mathbf{5})$	$(\mathbf{2.34}, \mathbf{2.31})$
$\mathbf{10^2}$	$(\mathbf{4.95}, \mathbf{5})$	$(\mathbf{2.27}, \mathbf{2.30})$
[a] $\mathbf{10^3}$	$(\mathbf{4.79}, \mathbf{5})$	$(\mathbf{2.20}, \mathbf{2.29})$

[a] Square-ish matrix

Precision $\mathbf{p} = \mathbf{3}$

$\mathbf{d}$; ν	(high) $\mathbf{1}$	(low) $\mathbf{5 \cdot 10^{-4}}$
$\mathbf{10}$	$(\mathbf{0.05}, \mathbf{0.05})$	$(\mathbf{0.023}, \mathbf{0.023})$
$\mathbf{10^2}$	$(\mathbf{0.05}, \mathbf{0.05})$	$(\mathbf{0.023}, \mathbf{0.023})$
[a] $\mathbf{10^3}$	$(\mathbf{0.047}, \mathbf{0.05})$	$(\mathbf{0.022}, \mathbf{0.023})$

[a] Square-ish matrix

Experiments (3)

Our universe

- $\mathbf{A} \in [-1,1]^{n\times d}$
- All elements of $\mathtt{SR}(\mathbf{A}) \in \mathcal{F}^{\{p\}}$, $1 \le p \le 5$

$$\mathcal{F}^{\{p\}} = \{\pm\, {}^{m}\!/_{10^p}, \quad \text{for all integers } m = \underbrace{0,1,2,\ldots,10^p-1}_{\le p \text{ digits}}\} \;\cup\; \{\pm 1\}$$

Setting (3)

- $\sigma_{\min}(\mathbf{A}) = 10^{-2}$
- $n = 10^4;\ 10^5;\ 10^6$ and $d = 10;\ 100;\ 1000$
- For a fixed d, all $\mathbf{A}$ have the same singular values

Experiments (3) [recall: $n \times d$ matrix A and $\sigma_{\min}(\mathrm{A}) = 10^{-2}$]

Each entry in the tables shows the pair of values $\left(\sigma_{\min}(\widetilde{\mathbf{A}}), \mathbf{R}\sqrt{n\nu}\right)$

Precision $\mathbf{p} = 1$

$\mathbf{d}$; $\mathbf{n}$	$\mathbf{10^4}$	$\mathbf{10^6}$
$\mathbf{10}$	$(\mathbf{4.11}, \mathbf{4.07})$	$(\mathbf{31.87}, \mathbf{30.85})$
$\mathbf{10^2}$	$(\mathbf{4.07}, \mathbf{4.06})$	$(\mathbf{34.59}, \mathbf{34.09})$
$\mathbf{10^3}$	$(\mathbf{3.86}, \mathbf{4.05})$[a]	$(\mathbf{33.24}, \mathbf{33.01})$

[a] Square-ish matrix

Precision $\mathbf{p} = 4$

$\mathbf{d}$; $\mathbf{n}$	[a] $\mathbf{10^4}$	[b] $\mathbf{10^6}$
$\mathbf{10}$	$(\mathbf{0.01}, \mathbf{0.004})$	$(\mathbf{0.04}, \mathbf{0.04})$
$\mathbf{10^2}$	$(\mathbf{0.01}, \mathbf{0.004})$	$(\mathbf{0.04}, \mathbf{0.04})$
$\mathbf{10^3}$	$(\mathbf{0.01}, \mathbf{0.004})$	$(\mathbf{0.04}, \mathbf{0.04})$

[a] n is "small"$\rightarrow$ smaller singular value does not increase much; bounds are tight

[b] n is "large"$\rightarrow$ smaller singular increases more; bounds are tight

Future work

Theory

- New Random Matrix Theory bounds for matrices whose entries are independent, but not identically distributed random variables.
  1. Can be used to prove similar bounds for square-ish matrices.
  2. Can be used to remove or reduce the $\epsilon_{n,d}$ factor.
- Effect of stochastic rounding in downstream applications[a].

Experiments

- Experimental evaluation in GPUs/IPUs that support stochastic rounding, e.g., GraphCore IPU.
- Effect of stochastic rounding in downstream applications[a].

[a] From simple regression problems to DNN training.

References

C. Boutsikas, P. Drineas, and I. C. F. Ipsen, Small singular values can increase in lower precision, *SIAM Journal on Matrix Analysis and Applications*, 2024.

G. Dexter, C. Boutsikas, L. Ma, I. C. F. Ipsen, and P. Drineas, Stochastic rounding implicitly regularizes tall-and-thin matrices, *arXiv:2403.12278*, 2024.

References

J. V. Neumann, and H. H. Goldstine, Numerical inverting of matrices of high order, *Bulletin of the American Mathematical Society*, 1947.

G. E. Forsythe, Round-off errors in numerical integration on automatic machinery, *Bulletin of the American Mathematical Society*, 1950.

G. W. Stewart, A second order perturbation expansion for small singular values, *Linear Algebra and its Applications*, 1984.

G. W. Stewart, and J. G. Sun, Matrix perturbation theory, *Academic press*, 1990.

S. M. Rump, Inversion of extremely ill-conditioned matrices in floating-point, *Japan Journal of Industrial and Applied Mathematics*, 2009.

S. Gupta, A. Agrawal, K. Gopalakrishnan, and P. Narayanan, Deep learning with limited numerical precision *International Conference on Machine Learning*, 2015.

M. Croci, M. Fasi, N. J. Higham, T. Mary, and M. Mikaitis, Stochastic rounding: implementation, error analysis and applications, *Royal Society Open Science*, 2022.

I. Dumitriu and Y. Zhu, Extreme singular values of inhomogeneous sparse random rectangular matrices, *arXiv:2209.12271*, 2023.