S^TS-SVD with Applications

Davide Palitta   Valeria Simoncini

Dipartimento di Matematica, Centro AM^2
Alma Mater Studiorum - Università di Bologna

Randomized Numerical Linear Algebra
ICERM, Providence (RI), February 2 – 6, 2026

Sketching for (vector) least squares problem

$$\min \|b - Ax\|_2^2$$

$A \in \mathbb{R}^{m\times n}$, $m \geq n$, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$

- sketching (skecth-and-solve) $\leftarrow$ *standard* practice to speed up the numerical solution of

$$\min \|S(b - Ax)\|_2^2$$

  for $S \in \mathbb{R}^{s\times m}$, $s \ll m$
- S is often chosen as an (ε, δ, k)-**subspace embedding** $\leftarrow$ given a k-dimensional subspace $\mathcal{V} \subset \mathbb{R}^m$ (e.g., Range($[A, b]$)) we want

$$(1-\varepsilon)\|v\|^2 \leq \|Sv\|^2 \leq (1+\varepsilon)\|v\|^2, \quad \forall\, v \in \mathcal{V}$$

  with failure probability $1 - \delta$

[Sarlos (2006)], **[Rokhlin, Tygert (2008)]**, **[Avron, Maymounkov, Toledo (2010)]**

Sketching for (vector) least squares problem

$$\min \|S(b - Ax)\|_2^2$$

$A \in \mathbb{R}^{m \times n}$, $m \geq n$, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$, $S \in \mathbb{R}^{s \times m}$

- Common examples of S are Gaussians ($s = \mathcal{O}(\varepsilon^{-2} \log k \log 1/\delta)$) and SRTTs ($s = \mathcal{O}(\varepsilon^{-2}(k + \log m/\delta) \log k/\delta)$)
- If

$$x_* = \arg\min \|b - Ax\|_2^2, \qquad x_S = \arg\min \|S(b - Ax)\|_2^2$$

then

$$\|b - Ax_S\|_2^2 \leq \frac{1 + \varepsilon}{1 - \varepsilon} \|b - Ax_*\|_2^2$$

[Martinsson, Tropp (2020)]

and when it comes to a **matrix** least squares problem?

Nearest Orthogonal Matrix

$$\min \|A - Q\|_* \quad \text{s.t.} \quad Q^T Q = I, \; * = 2, F$$

- Applications
  - *Noisy rotation matrix correction*: restore orthogonality in noisy data that is supposed to be orthogonal
  - *Image recognition*: make the recognition process more robust to variations like illumination and small sample sizes
  - ...
- Exact solution provided by the orthogonal factor of the polar decomposition of A

$$A = PH, \quad P^T P = I, H \geq 0$$

- If $A = U\Sigma V^T$, then

$$P = UV^T \quad H = V\Sigma V^T$$

[Higham (1994)]

and when it comes to a **matrix** least squares problem?

Can we still accelerate the solution of

$$\min \|A - Q\|_* \quad \text{s.t.} \quad Q^T Q = I, \ * = 2, F$$

by **sketching**?

- $\min \|S(A - Q)\|_*$ (easy part)
- How do we include the orthogonality constraint?

Sketched norms

- S induces a semidefinite inner product (always)
- This is positive definite on the embedded space w/ h.p.

Definition (sketched norms)

Given a matrix $P \in \mathbb{R}^{m \times n}$ and a sketching matrix $S \in \mathbb{R}^{s \times m}$, if S is an $(\varepsilon, \delta, \text{rank}(P))$-subspace embedding for Range(P) then we can define the following matrix norms

$$\|P\|_{S,F}^2 := \text{trace}(P^T S^T S P)$$

and

$$\|P\|_{S,2}^2 := \max_{\|x\|_2=1} x^T P^T S^T S P x$$

Nearest sketched orthogonal matrix

$$\min \|A - Q\|_{S,*} \quad \text{s.t.} \quad Q^T S^T S Q = I, \; * = 2, F$$

Nearest sketched orthogonal matrix

$$\min \|A - Q\|_{S,*} \quad \text{s.t.} \quad Q^T S^T S Q = I, \ * = 2, F$$

How do we solve it?

the singular vectors of A are not $S^T S$-orthogonal!

The $S^T S$-SVD

REMARK: the SVD can be defined by using any inner product!

$S^T S$-SVD

Given $A \in \mathbb{R}^{m \times n}$, $m \geq n$, and an $(\varepsilon, \delta, \text{rank}(A))$-subspace embedding $S \in \mathbb{R}^{s \times m}$ of Range(A), $s \geq \text{rank}(A)$, there exist, with high probability, an $S^T S$-orthogonal matrix $W \in \mathbb{R}^{m \times r}$, an orthogonal matrix $V \in \mathbb{R}^{n \times r}$, $r = \min\{s, n\}$, such that A can be written with high probability as

$$A = W \Theta V^T, \quad \Theta = \begin{bmatrix} \theta_1 & & \\ & \ddots & \\ & & \theta_r \end{bmatrix} \in \mathbb{R}^{r \times r}, \quad \theta_1 \geq \ldots \geq \theta_r \geq 0$$

The $S^T S$-SVD

REMARK: the SVD can be defined by using any inner product!

$S^T S$-SVD

Given $A \in \mathbb{R}^{m \times n}$, $m \geq n$, and an $(\varepsilon, \delta, \text{rank}(A))$-subspace embedding $S \in \mathbb{R}^{s \times m}$ of Range(A), $s \geq$ rank(A), there exist, with high probability, an $S^T S$-orthogonal matrix $W \in \mathbb{R}^{m \times r}$, an orthogonal matrix $V \in \mathbb{R}^{n \times r}$, $r = \min\{s, n\}$, such that A can be written with high probability as

$$A = W \Theta V^T, \quad \Theta = \begin{bmatrix} \theta_1 & & \\ & \ddots & \\ & & \theta_r \end{bmatrix} \in \mathbb{R}^{r \times r}, \quad \theta_1 \geq \ldots \geq \theta_r \geq 0$$

IT'S AN EQUALITY

(w/ high probability)

How to compute the $S^T S$-SVD?

Variant I: Compute the randomized QR of A

$$A = QR, \quad Q^T S^T S Q = I, \; R \text{ upper triang.}$$

and (deterministic, standard) SVD of R

$$R = U \Theta V^T$$

then, the $S^T S$-SVD of A is given by

$$A = (QU) \Theta V^T$$

How to compute the $S^T S$-SVD?

Variant II: Compute (deterministic, standard) SVD of SA

$$SA = U\Theta V^T$$

then, the $S^T S$-SVD of A is given by

$$A = (AV\Theta^{\dagger})\Theta V^T$$

- Variant I: possibly more robust but definitely more expensive
- Variant II: exploit the sparsity of A but small θ_i's may be troublesome

On the S^TS-SVD (once again)

Theorem [Gilbert, Park, Wakin (2013)]

Let S be an $(\varepsilon, \delta, \text{rank}(A))$-subspace embedding for Range(A), and let $A = W\Theta V^T$ be the S^TS-SVD of $A \in \mathbb{R}^{m\times n}$, $m \geq n$, and $A = U\Sigma Y^T$, $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n)$, its standard SVD. Then with high probability,

$$\sqrt{1-\varepsilon}\cdot\sigma_k \leq \theta_k \leq \sqrt{1+\varepsilon}\cdot\sigma_k, \quad \text{for all } k = 1, \ldots, n$$

REMARK: Gilbert et al defined the *sketched SVD* of A as the SVD of SA, namely

$$SA = U\Theta V^T$$

No relation between U and A (no notion of S^TS-orthogonality)

Let's see it with an example

- $A \in \mathbb{R}^{m\times m}$, $m = 5\,000$, Cauchy matrix

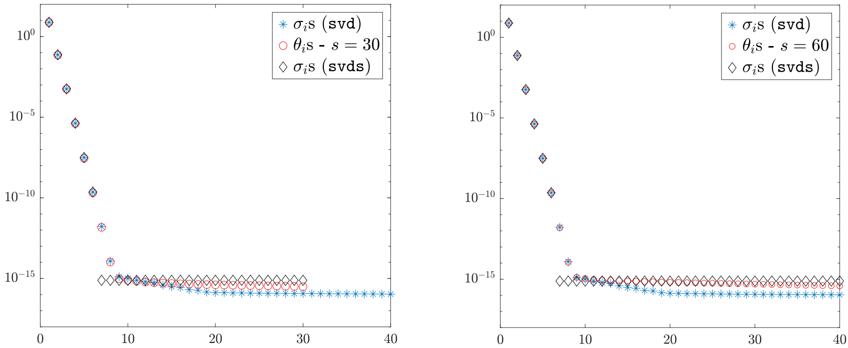
$$A_{i,j} = \frac{1}{x_i + y_j}$$

  $x_i \in [2, 100]$, $y_j \in [-1\,000, -500]$
- $S \in \mathbb{R}^{s\times m}$ SRTT

$$S = \sqrt{\frac{m}{s}} DCF$$

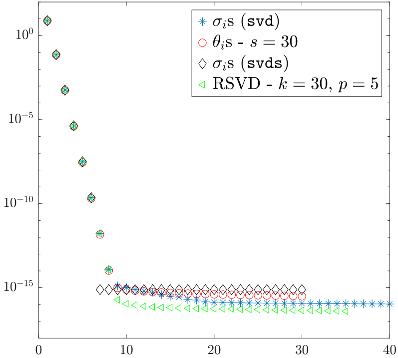
  - $E \in \mathbb{R}^{m\times m}$ diagonal w/ Radamacher entries
  - $C \in \mathbb{R}^{m\times m}$ discrete cosine transform
  - $D \in \mathbb{R}^{s\times m}$ random selection of s entries

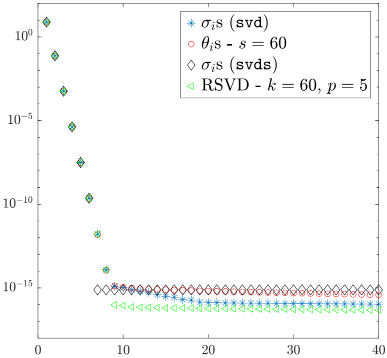
Let's see it with an example



- svd: 3.3 secs
- svds: 1.82 ($s = 30$), 3.6 ($s = 60$) secs
- $S^T S$-SVD: **0.14** ($s = 30$), **0.15** ($s = 60$) secs

Let's see it with an example





- `svd`: 3.3 secs
- `svds`: 1.82 ($s = 30$), 3.6 ($s = 60$) secs
- $S^T S$-SVD: **0.14** ($s = 30$), **0.15** ($s = 60$) secs
- RSVD: 0.43 ($k = 30$, $p = 5$), 0.46 ($k = 60$, $p = 5$) secs

Randomized SVD

$$U\Sigma V^T \approx A$$

- If σ_i's are the **true** singular values of A and $\widehat{\sigma}_i$'s those given by RSVD, then

$$\frac{\sigma_j}{\sqrt{1+\gamma_j^2\|\Omega_2\Omega_1^\dagger\|_2^2}} \leq \widehat{\sigma}_j \leq \sigma_j, \quad \text{for all } j=1,\ldots,k,$$

where $\gamma_j := \sigma_{j+1}/\sigma_j$ $\Omega_1 := W_1^T\Omega$, $\Omega_2 := W_2^T\Omega$, $W = [W_1, W_2]$ **true** right singular vectors of A. Main assumptions **[Saibaba (2019)]**

  - $\text{rank}(\Omega_1) = k$
  - $\gamma_k := \sigma_{k+1}/\sigma_k < 1$

Randomized SVD

$$U\Sigma V^T \approx A$$

- If σ_i's are the **true** singular values of A and $\widehat{\sigma}_i$'s those given by RSVD, then

$$\frac{\sigma_j}{\sqrt{1+\gamma_j^2\|\Omega_2\Omega_1^\dagger\|_2^2}} \leq \widehat{\sigma}_j \leq \sigma_j, \quad \text{for all } j=1,\dots,k,$$

where $\gamma_j := \sigma_{j+1}/\sigma_j$ $\Omega_1 := W_1^T\Omega$, $\Omega_2 := W_2^T\Omega$, $W=[W_1, W_2]$ **true** right singular vectors of A. Main assumptions **[Saibaba (2019)]**

  - $\text{rank}(\Omega_1) = k$
  - $\gamma_k := \sigma_{k+1}/\sigma_k < 1$

Randomized SVD

$$U\Sigma V^T \approx A$$

- If σ_i's are the **true** singular values of A and $\widehat{\sigma}_i$'s those given by RSVD, then

$$\frac{\sigma_j}{\sqrt{1+\gamma_j^2\|\Omega_2\Omega_1^\dagger\|_2^2}} \leq \widehat{\sigma}_j \leq \sigma_j, \quad \text{for all } j = 1, \ldots, k,$$

  where $\gamma_j := \sigma_{j+1}/\sigma_j$ $\Omega_1 := W_1^T\Omega$, $\Omega_2 := W_2^T\Omega$, $W = [W_1, W_2]$ **true** right singular vectors of A. Main assumptions **[Saibaba (2019)]**
  - $\text{rank}(\Omega_1) = k$
  - $\gamma_k := \sigma_{k+1}/\sigma_k < 1$

Can we further improve the singular values bound?

- RSVD

$$\frac{\sigma_j}{\sqrt{1+\gamma_j^2\|\Omega_2\Omega_1^\dagger\|_2^2}} \leq \widehat{\sigma}_j \leq \sigma_j, \quad \text{for all } j=1,\dots,k,$$

- S^TS-SVD

$$\sqrt{1-\varepsilon}\cdot\sigma_k \leq \theta_k \leq \sqrt{1+\varepsilon}\cdot\sigma_k, \quad \text{for all } k=1,\dots,n$$

Can we get something like

$$\sqrt{1-\varepsilon}\cdot\sigma_k \leq \theta_k \leq \sigma_k, \quad \text{for all } k=1,\dots,n \textbf{ ?}$$

Sketched norms

Definition (sketched norms)

Given a matrix $P \in \mathbb{R}^{m\times n}$ and a sketching matrix $S \in \mathbb{R}^{s\times m}$, if S is an $(\varepsilon, \delta, \text{rank}(P))$-subspace embedding for Range(P) and $P = W\Theta V^T$ is its S^TS-SVD, then

$$\|P\|_{S,F}^2 := \text{trace}(P^TS^TSP) = \sum_{i=1}^{n} \theta_i^2$$

and

$$\|P\|_{S,2}^2 := \max_{\|x\|_2=1} x^TP^TS^TSPx = \theta_1$$

We can define the sketched version of any norm based on singular values: Schatten-p norm

$$\|P\|_p = \left(\sum_{i=1}^{n} \sigma_i^p\right)^{1/p}, \qquad \|P\|_{S,p} = \left(\sum_{i=1}^{n} \theta_i^p\right)^{1/p}$$

Sketched norms

Definition (sketched norms)

Given a matrix $P \in \mathbb{R}^{m\times n}$ and a sketching matrix $S \in \mathbb{R}^{s\times m}$, if S is an $(\varepsilon, \delta, \text{rank}(P))$-subspace embedding for Range(P) and $P = W\Theta V^T$ is its S^TS-SVD, then

$$\|P\|_{S,F}^2 := \text{trace}(P^TS^TSP) = \sum_{i=1}^{n} \theta_i^2$$

and

$$\|P\|_{S,2}^2 := \max_{\|x\|_2=1} x^TP^TS^TSPx = \theta_1$$

We can define the sketched version of any norm based on singular values:

Nuclear norm

$$\|P\|_* = \sum_{i=1}^{n} \sigma_i, \qquad \|P\|_{S,*} = \sum_{i=1}^{n} \theta_i$$

Back to the (sketched) nearest orthogonal matrix

$$\min \|A - Q\|_{S,*} \quad \text{s.t.} \quad Q^T S^T S Q = I, \; * = 2, F$$

Let the $S^T S-$SVD of A be given by

$$A = W \Theta V^T, \quad (W^T S^T S W = I)$$

then

$$P = WV^T$$

is s.t.

$$P = \arg\min \|A - Q\|_{S,*} \quad \text{s.t.} \quad Q^T S^T S Q = I, \; * = 2, F \quad (w/h.p.)$$

REMARK: P is the $S^T S$-orthogonal factor of the *randomized polar decomposition* of A ($A = PH$, $H = V \Theta V^T$)

Back to the (sketched) nearest orthogonal matrix

$$\min_{Q^TQ=I} \|A-Q\|_*, \quad \min_{Q^TS^TSQ=I} \|A-Q\|_{S,*} \quad * = 2, F$$

Back to the (sketched) nearest orthogonal matrix

$$\min_{Q^TQ=I} \|A-Q\|_*, \qquad \min_{Q^TS^TSQ=I} \|A-Q\|_{S,*} \qquad * = 2, F$$

Theorem [P., Simoncini (2025)]

Let

$$P := \underset{Q^TS^TSQ=I}{\arg\min} \|A-Q\|_{S,2}$$

and let

$$T := \underset{Q^TQ=I}{\arg\min} \|A-Q\|_2$$

then, with high probability it holds

$$\|A-T\|_2 - \frac{\varepsilon}{1-\varepsilon} \leq \|A-P\|_2 \leq \frac{1+\varepsilon}{1-\varepsilon}\|A-T\|_2 + \frac{\varepsilon}{1-\varepsilon}$$

A numerical example

- $A \in \mathbb{R}^{m \times n}$, $m = 37\,932$, $n = 331$ (`abtaha2` in the SuiteSparse Matrix Collection Repo.)
- $S \in \mathbb{R}^{s \times m}$ Gaussian

$$P := \operatorname*{arg\,min}_{Q^T S^T S Q = I} \|A - Q\|_{S,2} \quad T := \operatorname*{arg\,min}_{Q^T Q = I} \|A - Q\|_2$$

A numerical example

- $A \in \mathbb{R}^{m\times n}$, $m = 37\,932$, $n = 331$ (`abtaha2` in the SuiteSparse Matrix Collection Repo.)
- $S \in \mathbb{R}^{s\times m}$ Gaussian

$$P := \underset{Q^TS^TSQ=I}{\arg\min} \|A-Q\|_{S,2} \quad T := \underset{Q^TQ=I}{\arg\min} \|A-Q\|_2$$

s	S^TS-orth $\|A-P\|_2$	Distance $\|P-T\|_2$	Time (secs)
$2n$	24.99	3.69	0.46
$4n$	24.80	2.57	0.46
$6n$	24.77	2.36	0.47
$8n$	24.76	2.27	0.48
$10n$	24.75	2.21	0.48
$12n$	24.74	2.18	0.49

$\|A-T\|_2 = 24.77$, 1.09 secs to compute T (we are 2+X faster)

The $S^T S$-SVD helps a lot!

Theorem [P., Simoncini (2025)]

Let $P \in \mathbb{R}^{m\times n}$ have $S^T S$-orthonormal columns with S being an $(\varepsilon, \delta, \mathrm{rank}(P))$-subspace embedding for Range(P). Then, with high probability,

$$\|P^T P - I\|_F \leq \frac{\varepsilon}{1-\varepsilon}\sqrt{n} \quad \text{and} \quad \|P^T P - I\|_2 \leq \frac{\varepsilon}{1-\varepsilon}$$

The S^TS-SVD helps a lot!

Theorem [P., Simoncini (2025)]

Let $P \in \mathbb{R}^{m\times n}$ have S^TS-orthonormal columns with S being an $(\varepsilon, \delta, \text{rank}(P))$-subspace embedding for Range(P). Then, with high probability,

$$\|P^TP - I\|_F \leq \frac{\varepsilon}{1-\varepsilon}\sqrt{n} \quad \text{and} \quad \|P^TP - I\|_2 \leq \frac{\varepsilon}{1-\varepsilon}$$

Theorem [P., Simoncini (2025)]

Let $T \in \mathbb{R}^{m\times n}$ have orthonormal columns. Let S be an $(\varepsilon, \delta, \text{rank}(T))$-subspace embedding for Range(T). Then, with high probability,

$$\|T^TS^TST - I\|_F \leq \varepsilon\sqrt{n} \quad \text{and} \quad \|T^TS^TST - I\|_2 \leq \varepsilon$$

Conclusions

What I showed you

- Formalization of the S^TS-SVD
- Characterization of the nearest sketched orthogonal matrix
- Relation between sketched and original problem

The S^TS-SVD may provide

- Efficient detection of singular value decays
- Fast computation of low-rank approximations w/ (S^TS-)orthogonal factors
- Full characterization of skecthing for (vector) least squares problems

Reference: *S^TS-SVD via Sketching and the Nearest S^TS-Orthogonal Matrix*
D. Palitta and V. Simoncini
BIT, Vol. 66, no. 7, (2026)