Accuracy of randomized Nyström

Laura Grigori

EPFL and PSI

November 12, 2024

Accuracy of randomized SVD

Consider randomized SVD:

$$A_{\mathrm{RSVD}} = QQ^T A = \mathcal{P}^o A \tag{1}$$

where Q is orthonormal basis of $A\Omega$. For some $d \geq k$ and $\varepsilon^* \leq \frac{1}{2}$, if Ω is an $(\frac{1}{3}, \delta, d)$ OSE and $(\sqrt{\frac{\varepsilon}{d}}, \sqrt{\delta}N, 1) OSE$, with probablity $1 - 2\delta$ we have:

$$\|A - A_{\mathrm{RSVD}}\|_F^2 \leq (1 + \varepsilon^*)\|A - [\![A]\!]_d\|_F^2, \tag{2}$$

where $[\![A]\!]_d$ is rank-d truncated SVD of A.

- SRHT: OSE properties satisfied with $l = O(d \log \frac{n}{\delta} \log \frac{m}{\delta})$ rows
- Gaussian: OSE properties satisfied with $l = O(d \log \frac{m}{\delta})$ rows
- proof not provided, given in [Balabanov et al., 2023]

Accuracy of randomized SVD (contd)

- Under the condition (2), quasi-optimality of A_{RSVD} with respect to Frobenius norm.
- If $(\sum_{i=d+1}^{n} \sigma_i^2)^{1/2} = \|A - [\![A]\!]_d\|_F$ is small compared to the $k+1$-th singular value $\sigma_{k+1} = \|A - [\![A]\!]_k\|_2$ then $[\![A_{\mathrm{RSVD}}]\!]_k$ almost as accurate as $[\![A]\!]_k$ for all three norms.

Accuracy of randomized Nyström

Let Π_X be orthogonal projector onto range of $X = A^{\frac{1}{2}}\Omega^{\mathrm{T}}$. We have:

$$
\begin{aligned}
A_{Nyst} &= (A\Omega)(\Omega^T A\Omega)^+(\Omega^T A) \\
&= A^{1/2}(A^{1/2}\Omega(\Omega^T A^{1/2}A^{1/2}\Omega)^+\Omega^T A^{1/2})A^{1/2} \\
&= A^{1/2}\Pi_X A^{1/2}
\end{aligned}
$$

We obtain [Gittens, 2011]:

$$
\begin{aligned}
A - A_{Nyst} &= A^{1/2}(I - \Pi_X)A^{1/2} = A^{1/2}(I - \Pi_X)^2 A^{1/2} && (3) \\
&= (A^{\frac{1}{2}} - \Pi_X A^{\frac{1}{2}})^T (A^{\frac{1}{2}} - \Pi_X A^{\frac{1}{2}}) && (4)
\end{aligned}
$$

- Since A is SPSD and $I - \Pi_x$ is SPSD, then A_{Nyst} is also SPSD
- Equation (4) allows to use the error bound of randomized SVD

Accuracy of randomized Nyström (contd)

Since A is SPSD, there is a unique $A^{1/2}$ that is SPSD, has same eigenspace as A and $A = (A^{1/2})^2$
By using the identity:

$$A - A_{Nyst} = (A^{\frac{1}{2}} - \Pi_X A^{\frac{1}{2}})^T (A^{\frac{1}{2}} - \Pi_X A^{\frac{1}{2}}),$$

the goal is to show that Π_X captures well the action of $A^{\frac{1}{2}}$. We use the error bound of randomized SVD, for some $d \geq k$ and $\varepsilon^* \leq \frac{1}{2}$ it holds that

$$\|A^{\frac{1}{2}} - \Pi_X A^{\frac{1}{2}}\|_F^2 \leq (1 + \varepsilon^*)\|A^{\frac{1}{2}} - [\![A^{\frac{1}{2}}]\!]_d\|_F^2, \quad (5)$$

where $[\![A^{\frac{1}{2}}]\!]_d$ is rank-d truncated SVD of $A^{\frac{1}{2}}$. We obtain

$$\begin{aligned} \|A - A_{Nyst}\|_* &= \|A^{\frac{1}{2}} - \Pi_X A^{\frac{1}{2}}\|_F^2 & (6) \\ &\leq (1 + \varepsilon^*)\|A^{\frac{1}{2}} - [\![A^{\frac{1}{2}}]\!]_d\|_F^2 = (1 + \varepsilon^*)\|A - [\![A]\!]_d\|_* & (7) \end{aligned}$$

Accuracy of randomized Nyström (contd)

Equation (5) also implies accuracy of truncated approximation $[\![A_{Nyst}]\!]_k$ due to following relation (using triangle inequality and best rank-k approximation of A_{Nyst}, see e.g. instance [Tropp et al., 2017, Proposition 6.1]):

$$\begin{aligned} \|A - [\![A_{Nyst}]\!]_k\|_\xi &\leq \|A - A_{Nyst}\|_\xi + \|A_{Nyst} - [\![A_{Nyst}]\!]_k\|_\xi && (8) \\ &\leq \|A - A_{Nyst}\|_\xi + \|A_{Nyst} - [\![A]\!]_k\|_\xi && (9) \\ &\leq 2\|A - A_{Nyst}\|_\xi + \|A - [\![A]\!]_k\|_\xi && (10) \end{aligned}$$

where $\xi = 2$ or $*$, so that we have by (6),

$$\|A - [\![A_{Nyst}]\!]_k\|_\xi \leq 3\|A - [\![A]\!]_d\|_* + \|A - [\![A]\!]_k\|_\xi \tag{11}$$

Accuracy of randomized Nyström (contd)

- Under the condition (5), quasi-optimality of $[\![A_{Nyst}]\!]_k$ with respect to nuclear norm.
- If $\sum_{i=d+1}^{n} \sigma_i = \|A - [\![A]\!]_d\|_*$ is small compared to the $k+1$-th singular value $\sigma_{k+1} = \|A - [\![A]\!]_k\|_2$ then $[\![A_{Nyst}]\!]_k$ almost as accurate as $[\![A]\!]_k$ for both nuclear and spectral norm.

References (1)

Balabanov, O., Beaupere, M., Grigori, L., and Lederer, V. (2023).
Block subsampled randomized hadamard transform for low-rank approximation on distributed architectures.
In *ICML'23: Proceedings of the 40th International Conference on Machine Learning*, number 66, pages 1564–1576.

Gittens, A. (2011).
The spectral norm error of the naive nystrom extension.
arXiv preprint arXiv:1110.5305.

Tropp, J. A., Yurtsever, A., Udell, M., and Cevher, V. (2017).
Practical sketching algorithms for low-rank matrix approximation.
SIAM Journal on Matrix Analysis and Applications, 38(4):1454–1485.