Improved analysis of randomized SVD for top-eigenvector approximation

Ruo-Chun Tzeng[1] Po-An Wang[2] Florian Adriaens[1] Aristides Gionis[1] Chi-Jen Lu[3]

[1]Division of Theoretical Computer Science
[2]Division of Decision and Control Systems
KTH Royal Institute of Technology, Sweden

[3]Institute of Information Science
Academia Sinica, Taiwan

Top-eigenvector approximation

Given $\mathcal{T} \subseteq \mathbb{R}^n \backslash \{\mathbf{0}\}$ and a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, find

$$\operatorname{argmax}_{\mathbf{x}\in\mathcal{T}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}.$$

A computational efficient way to solve these problem is

1. Find the top-eigenvector $\mathbf{u}_1$ of $\mathbf{A}$
2. Round u_1 into a vector in $\mathcal{T}$ (if needed)

However, what we practically obtain is the approximated top-eigenvector $\hat{\mathbf{u}}$ of $\mathbf{A}$ by numerical solvers, not $\mathbf{u}_1$.

Characterizing the gap between $\mathbf{u}_1$ and $\hat{\mathbf{u}}$

Let $(\lambda_i, \mathbf{u}_i)$ be the i-th largest eigenpair of $\mathbf{A}$, $\lambda_1 > 0$ and define

$$R(\hat{\mathbf{u}}) = \lambda_1^{-1}\frac{\hat{\mathbf{u}}^T\mathbf{A}\hat{\mathbf{u}}}{\hat{\mathbf{u}}^T\hat{\mathbf{u}}}. \tag{1}$$

We consider $\mathcal{O}(nd)$-space and $\mathcal{O}(q)$-pass algorithms with $d, q \in \mathbb{N}$.

Prior analysis of $R(\hat{\mathbf{u}})$ are all additive bounds. For these additive bounds to be meaningulf, [4] showed that $q = \Omega(\ln n)$ is required.

State-of-the-art: $R(\hat{\mathbf{u}}) \geq 1 - \mathcal{O}(\ln n/q)$ achieved by Randomized SVD [1], shown by [3], for any $\mathbf{A} \succcurlyeq 0$.

Q: Is $q = \Omega(\ln n)$ necessary or an artifact of analysis?

A: We provide the first non-trivial guarantee of $R(\hat{\mathbf{u}})$ in the regime of $q = o(\ln n)$ by analyzing of Randomized SVD [1]:

Algorithm: RSVD($\mathbf{A}, q, d$)

```
Y ← A^q S where S ~ N(0,1)^{n×d};
Y = QR;
B ← Q^T A Q;
û = Q u_1(B);
return û;
```

Our techniques: reduction to projection length

Our core technique is a reduction from

$$R(\hat{\mathbf{u}}) = \max_{\mathbf{a}\in\mathbb{S}^{d-1}} \frac{\sum_{i\in[n]} \alpha_i^{2q+1}\langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2}{\sum_{i\in[n]} \alpha_i^{2q}\langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2} \text{ with } \alpha_i = \frac{\lambda_i}{\lambda_1}, \forall i \in [n],$$

for any $\mathbf{A} \succcurlyeq 0$, to $\cos^2\theta(\mathbf{e}_1, \mathbf{S})$ which is well-known to be $\Theta(\frac{d}{n})$ w.h.p.:

(Lemma by [2]) Let $\mathbf{S} \sim \mathcal{N}(0,1)^{n\times d}$ with $d \ll n$. Then,

$$\forall \mathbf{v} \neq \mathbf{0}, \cos^2\theta(\mathbf{v}, \mathbf{S}) = \Theta\left(\frac{d}{n}\right) \text{ w.p. } 1 - e^{-\Omega(d)},$$

where $\theta(\mathbf{v}, \mathbf{X}) = \cos^{-1}\left(\max_{\mathbf{x}\in\text{range}(\mathbf{X})} \frac{\langle \mathbf{v},\mathbf{x}\rangle}{\|\mathbf{v}\|_2\|\mathbf{x}\|_2}\right)$.

Our technique generalizes to indefinite $\mathbf{A}$ under a mild assumption.

Our improved analysis of RSVD

Positive semidefinite matrices:

(Theorem 1) $\forall \mathbf{A} \succcurlyeq 0$, $R(\hat{\mathbf{u}}) = \left(\Omega\left(\frac{d}{n}\right)\right)^{\frac{1}{2q+1}}$ w.p. $1 - e^{-\Omega(d)}$.

(Theorem 2) $\exists \mathbf{A} \succcurlyeq 0$ s.t. $R(\hat{\mathbf{u}}) = \mathcal{O}\left(\left(\frac{d}{n}\right)^{\frac{1}{2q+1}}\right)$ w.p. $1 - e^{-\Omega(d)}$.

(Theorem 3) For $\mathbf{A} \succcurlyeq 0$ with (i_0, γ)-power-law decay, $i_0 \in [n]$ and $\gamma > 1/2q$, $R(\hat{\mathbf{u}}) = \Omega\left(\left(\frac{d}{d+i_0}\right)^{\frac{1}{2q+1}}\right)$ w.p. $1 - e^{-\Omega(d)}$.

Indefinite matrices:

Assume $\exists \kappa \in (0, 1]$ s.t. $\sum_{i=2}^n \alpha_i^{2q+1} \geq \kappa \sum_{i=2}^n |\alpha_i|^{2q+1}$.

(Theorem 4) For $\mathbf{A}$ with (i_0, γ)-power-law decay, $i_0 \in [n]$ and $\gamma > 1/2q$, $\exists c_\kappa > 0$ s.t.

$$R(\hat{\mathbf{u}}) = \Omega\left(c_\kappa\left(\frac{d}{d+i_0}\right)^{\frac{1}{2q+1}}\right) \text{ with prob. } \geq 1 - e^{-\Omega(\sqrt{d}\kappa^2)}.$$

Extension: exploiting prior knowledge of large $\langle \mathbf{u}_1, \mathbf{1}\rangle^2$

Remind that $\mathbf{Y}_{:,j} = \mathbf{A}^q\mathbf{S}_{:,j} = \sum_{i=1}^n \lambda_i^q(\mathbf{u}_i^T\mathbf{S}_{:,j})\mathbf{u}_i$ for any $j \in [d]$.
For large $\langle \mathbf{u}_1, \mathbf{1}\rangle^2$, it is possible to make $\mathbf{Y}_{:,j}$ align to $\mathbf{u}_1$ faster by sampling entries of $\mathbf{S}$ i.i.d. from non-centered distributions.

Algorithm: RandSum($\mathbf{A}, q, d, p$)

```
S_1 ~ N(0,1)^{n×⌈d/2⌉}, S_2 ~ Bernoulli(p)^{n×⌊d/2⌋};
S ← [S_1  S_2];
return RSVD(A,S,q,d);
```

Positive semidefinite matrices:

(Theorem 5) For $\mathbf{A} \succcurlyeq 0$, $\hat{\mathbf{u}}$ =RandSum($\mathbf{A}$,q,d,p) satisfies

$$R(\hat{\mathbf{u}}) = \left(\Omega\left(\frac{\max\{d, \langle \mathbf{u}_1, \mathbf{1}_n\rangle^2\}}{n}\right)\right)^{\frac{1}{2q+1}} \text{ w.p. } 1 - e^{-\Omega(d)}.$$

Indefinite matrices:

Under one additional assumption, the guarantee of RSVD and RandSum for p.s.d. matrices generalize to indefinite matrices.

References

[1] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM review*, 2011.

[2] Moritz Hardt and Eric Price. The noisy power method: a meta algorithm with applications. In *Proc. of NeurIPS*, 2014.

[3] Cameron Musco and Christopher Musco. Randomized block krylov methods for stronger and faster approximate singular value decomposition. In *Proc. of NeurIPS*, 2015.

[4] Max Simchowitz, Ahmed El Alaoui, and Benjamin Recht. Tight query complexity lower bounds for pca via finite sample deformed wigner law. In *Proc. of STOC*, 2018.

European Research Council
Established by the European Commission

Social Mining & Big Data Ecosystem
SoBigData
RESEARCH INFRASTRUCTURE

WASP
WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM