Improved analysis of randomized SVD for top-eigenvector approximation

Ruo-Chun Tzeng[1]   Po-An Wang[1]   Florian Adriaens[1]

Aristides Gionis[1]   Chi-Jen Lu[2]

[1]KTH   [2]Academic Sinica

International Conference on ArtificialIntelligence and Statistics (AISTATS) 2022

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

- PCA: $\mathbf{A} = \mathbf{X}\mathbf{X}^T$ where $\mathbf{X} \in \mathbb{R}^{n \times m}$ and $\mathcal{T} = \mathbb{R}^n \backslash \{\mathbf{0}\}$
- k-conflicting group detection [1, 8]:
  - $\mathbf{A}$: undirected signed adjacency matrix
  - $\mathcal{T} = \{q, 0, -1\}^n \backslash \{\mathbf{0}\}$ for $q \in [k-1]$
- 2-community detection:
  - $\mathbf{A}$: modularity matrix [4] or Bethe-Hessian matrix [5, 6]
  - $\mathcal{T} = \{\pm 1\}^n \backslash \{\mathbf{0}\}$

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 $\underset{\mathbf{x} \in \mathcal{T}}{\operatorname{argmax}} \dfrac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$.

A computational efficient way to solve these problem is

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

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 $\mathbf{u}_1$ into $\mathcal{T}$ (if needed)~~

1. Find the approximated $\hat{\mathbf{u}}$ of $\mathbf{A}$
2. Round $\hat{\mathbf{u}}$ into $\mathcal{T}$ (if needed)

Top-eigenvector approximation

To characterize the gap, let $(\lambda_1, \mathbf{u}_1)$ of $\mathbf{A}$ be the top-eigenpair of $\mathbf{A}$, $\lambda_1 > 0$ and define

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

In this paper, we aim to study the performance of numerical solvers, w.r.t. (1), under $\mathcal{O}(nd)$-space and $\mathcal{O}(q)$-pass setting.

Top-eigenvector approximation

To characterize the gap, let $(\lambda_1, \mathbf{u}_1)$ of $\mathbf{A}$ be the top-eigenpair of $\mathbf{A}$, $\lambda_1 > 0$ and define

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

In this paper, we aim to study the performance of numerical solvers, w.r.t. (1), under $\mathcal{O}(nd)$-space and $\mathcal{O}(q)$-pass setting.

However, all prior works are additive bounds and require $q = \Omega(\ln n)$ to be meaningful.

- State-of-the-art [3, 7]: $R(\hat{\mathbf{u}}) \geq 1 - \mathcal{O}(\ln n / q)$ for any $\mathbf{A} \succcurlyeq 0$ w.h.p.

Top-eigenvector approximation

To characterize the gap, let $(\lambda_1, \mathbf{u}_1)$ of $\mathbf{A}$ be the top-eigenpair of $\mathbf{A}$, $\lambda_1 > 0$ and define

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

In this paper, we aim to study the performance of numerical solvers, w.r.t. (1), under $\mathcal{O}(nd)$-space and $\mathcal{O}(q)$-pass setting.

However, all prior works are additive bounds and require $q = \Omega(\ln n)$ to be meaningful.

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

Our result: improved analysis of Randomized SVD

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

$\mathbf{Y} \leftarrow \mathbf{A}^q \mathbf{S}$ where $\mathbf{S} \sim \mathcal{N}(0,1)^{n \times d}$;
$\mathbf{Y} = \mathbf{QR}$;
$\mathbf{B} \leftarrow \mathbf{Q}^T \mathbf{AQ}$;
$\hat{\mathbf{u}} = \mathbf{Q}\,\mathbf{u}_1(\mathbf{B})$;
return $\hat{\mathbf{u}}$;

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

(Remark) $R(\hat{\mathbf{u}}) = e^{-\mathcal{O}(\ln n/(2q+1))} \geq 1 - \mathcal{O}(\ln n/q)$ subsumes the state-of-the-art [3].

Our technique: a reduction to random projection length

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

Proof idea:

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

where $\alpha_i = \frac{\lambda_i}{\lambda_1}, \forall i \in [n]$.

$$\Rightarrow \cos^2\theta(\mathbf{u}_1, \mathbf{S}) = \max_{\mathbf{a}\in\mathbb{S}^{d-1}} \frac{\langle \mathbf{S}^T\mathbf{u}_1, \mathbf{a}\rangle^2}{\sum_{i\in[n]} \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2}$$

(Random projection lemma) For $\mathbf{v} \in \mathbb{S}^{n-1}$ and $d \ll n$, with prob. $1 - e^{-\Omega(d)}$,

$$\cos^2\theta(\mathbf{v}, \mathbf{S}) = \Theta\left(\frac{d}{n}\right).$$

Our results: improved analysis of Randomized SVD

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

(Theorem 3) For $\mathbf{A} \succcurlyeq 0$ with (i_0, γ)-power-law decay, $i_0 \in [n]$ and $\gamma > 1/2q$,

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

(Assumption 1) $\exists \kappa \in (0,1]$ such that $\sum_{i=2}^{n} \alpha_i^{2q+1} \geq \kappa \sum_{i=2}^{n} |\alpha_i|^{2q+1}$.

(Theorem 4) For $\mathbf{A}$ with (i_0, γ)-power-law decay, $i_0 \in [n]$ and $\gamma > 1/2q$, and satisfying Assumption 1, there exists a constant $c_\kappa > 0$ such that

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

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

Question: If $\langle \mathbf{u}_1, \mathbf{1} \rangle^2 = \Theta(n)$, is there a better choice of $\mathbf{S}$ other than $\mathcal{N}(0,1)^{n \times d}$?

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;
```

Hint: $\mathbf{Y}_{:,j} = \mathbf{A}^q \mathbf{S}_{:,j} = \sum_{i=1}^{n} \lambda_i^q (\mathbf{u}_i^T \mathbf{S}_{:,j}) \mathbf{u}_i, \ \forall j \in [d]$

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

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

1 $\mathbf{S}_1 \sim \mathcal{N}(0,1)^{n \times \lceil \frac{d}{2} \rceil}$, $\mathbf{S}_2 \sim \text{Bernoulli}(p)^{n \times \lfloor \frac{d}{2} \rfloor}$;

2 $\mathbf{S} \leftarrow [\mathbf{S}_1 \quad \mathbf{S}_2]$;

3 return RSVD($\mathbf{A}$,$\mathbf{S}$,q,d);

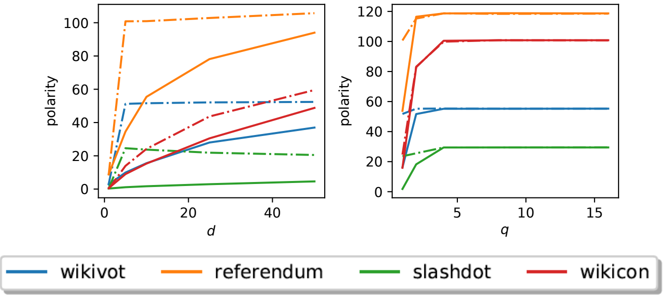
(Theorem 5) For $\mathbf{A} \succcurlyeq 0$, RandSum($\mathbf{A}$,$q$,$d$,$p$) returns $\hat{\mathbf{u}}$ satisfying

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

Theorem 5 generalizes to indefinite $\mathbf{A}$ under an assumption similar to **Assumption 1**.

Experiment: 2-conflicting group detection [1, 8]

	WikiVot	Referendum	Slashdot	WikiCon
$\|V\|$	7 115	10 884	82 140	116 717
$\|E\|$	100 693	251 406	500 481	2 026 646
(γ, i_0)	(4.6, 15)	(4.5, 16)	(5.3, 17)	(2.8, 22)
κ	0.397	0.620	0.204	0.034
$\cos\theta(\mathbf{u}_1, \mathbf{1}_n)$	0.378	0.399	0.194	0.193



- RSVD: solid line
- RandSum: dashed line

Summary

Contributions:

- Improve the analysis of RSVD, especially in the regime of $o(\ln n)$ passes, and provides the first analysis of $R(\cdot)$ for indefinite matrices.
- Study the property of Bernoulli random projection and demonstrate its usefulness to the task of conflicting group detection [1, 8].

Future works:

- It is an open problem to characterize the fundamental limit of $R(\hat{\mathbf{u}})$ for any q-pass $\mathcal{O}(nd)$-space algorithm.
- It would be useful to extend our results to (row/column)-stochastic matrices and to top-k eigenvectors approximations.

Reference I

[1] Francesco Bonchi, Edoardo Galimberti, Aristides Gionis, Bruno Ordozgoiti, and Giancarlo Ruffo.
Discovering polarized communities in signed networks.
In *Proc. of CIKM*, 2019.

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

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

Reference II

[4] Mark EJ Newman.
Modularity and community structure in networks.
Proc. of NAS, 2006.

[5] Alaa Saade, Florent Krzakala, and Lenka Zdeborová.
Spectral clustering of graphs with the bethe hessian.
In *Proc. of NeurIPS*, 2014.

[6] Alaa Saade, Marc Lelarge, Florent Krzakala, and Lenka Zdeborová.
Spectral detection in the censored block model.
In *Proc. of ISIT*, 2015.

[7] 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.

Reference III

[8] Ruo-Chun Tzeng, Bruno Ordozgoiti, and Aristides Gionis.
Discovering conflicting groups in signed networks.
In *Proc. of NeurIPS*, 2020.