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

Outline

Motivation

There are many problems of the form:

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

- 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 (CG) detection [1, 10]:
  - $\mathbf{A}$: 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 [6] or Bethe-Hessian matrix [7, 8]
  - $\mathcal{T} = \{\pm 1\}^n \backslash \{\mathbf{0}\}$

Motivation

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 u_1 into a vector in $\mathcal{T}$ (if needed)

Motivation

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 approximated top-eigenvector $\hat{\mathbf{u}}$ of $\mathbf{A}$ by numerical solvers
2. Round $\hat{\mathbf{u}}$ into a vector in $\mathcal{T}$ (if needed)

Motivation

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 approximated top-eigenvector $\hat{\mathbf{u}}$ of $\mathbf{A}$ by numerical solvers
2. Round $\hat{\mathbf{u}}$ into a vector in $\mathcal{T}$ (if needed)

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

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

Motivation

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

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

- For 2-CG [1], using $\hat{\mathbf{u}}$ (resp. $\mathbf{u}_1$) results in $\sqrt{n}/R(\hat{\mathbf{u}})$-approx (resp. $\sqrt{n}$-approx) algorithm.

Motivation

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.

Motivation

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.

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

- Randomized SVD yielding $R(\hat{\mathbf{u}}) \geq 1 - \mathcal{O}(\ln n/q)$ for any $\mathbf{A} \succcurlyeq 0$ w.h.p., shown by [5, 9], is the state-of-the-art.

Motivation

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.

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

Question

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

Randomized SVD [3] for top-eigenvector computation

Notation let $(\lambda_i(\cdot), \mathbf{u}_i(\cdot))$ be the i-th largest eigenpair of the given matrix

Algorithm given a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ with $\lambda_1 > 0$ and $q, d \in \mathbb{N}$

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 $\hat{\mathbf{u}}$;

Randomized SVD [3] for top-eigenvector computation

Notation let $(\lambda_i(\cdot), \mathbf{u}_i(\cdot))$ be the i-th largest eigenpair of the given matrix

Algorithm given a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ with $\lambda_1 > 0$ and $q, d \in \mathbb{N}$

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;
```

$\mathbf{Y} \leftarrow \mathbf{A}^q\mathbf{S}$ 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 $\hat{\mathbf{u}}$;

Step 1: random projection $\mathbf{Y} = \mathbf{A}^q\mathbf{S}$

- Effect of the powering: $\mathbf{Y}_{:,j} = \mathbf{A}^q\mathbf{S}_{:,j} = \sum_{i=1}^n \lambda_i^q(\mathbf{u}_i^T\mathbf{S}_{:,j})\mathbf{u}_i,\ \forall j \in [d]$
- Find the best unit vector $\hat{\mathbf{u}} \in \text{range}(\mathbf{Y})$

Randomized SVD [3] for top-eigenvector computation

Notation let $(\lambda_i(\cdot), \mathbf{u}_i(\cdot))$ be the i-th largest eigenpair of the given matrix

Algorithm given a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ with $\lambda_1 > 0$ and $q, d \in \mathbb{N}$

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

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

Step 1: random projection $\mathbf{Y} = \mathbf{A}^q \mathbf{S}$

- Effect of the powering: $\mathbf{Y}_{:,j} = \mathbf{A}^q \mathbf{S}_{:,j} = \sum_{i=1}^{n} \lambda_i^q (\mathbf{u}_i^T \mathbf{S}_{:,j}) \mathbf{u}_i, \ \forall j \in [d]$
- Find $\hat{\mathbf{u}} = \operatorname{argmax}\{\mathbf{v}^T \mathbf{A} \mathbf{v} : \mathbf{v} \in \operatorname{range}(\mathbf{Y}) \cap \mathbb{S}^{n-1}\}$

Randomized SVD [3] for top-eigenvector computation

Notation let $(\lambda_i(\cdot), \mathbf{u}_i(\cdot))$ be the i-th largest eigenpair of the given matrix

Algorithm given a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ with $\lambda_1 > 0$ and $q, d \in \mathbb{N}$

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

1 $\mathbf{Y} \leftarrow \mathbf{A}^q \mathbf{S}$ where $\mathbf{S} \sim \mathcal{N}(0,1)^{n \times d}$;
2 $\mathbf{Y} = \mathbf{Q}\mathbf{R}$;
3 $\mathbf{B} \leftarrow \mathbf{Q}^T \mathbf{A} \mathbf{Q}$;
4 $\hat{\mathbf{u}} = \mathbf{Q}\, \mathbf{u}_1(\mathbf{B})$;
5 return $\hat{\mathbf{u}}$;

Step 2: compute $\hat{\mathbf{u}} = \operatorname{argmax}\{\mathbf{v}^T \mathbf{A} \mathbf{v} : \mathbf{v} \in \operatorname{range}(\mathbf{Y}) \cap \mathbb{S}^{n-1}\}$

As $\forall \mathbf{v} \in \operatorname{range}(Y) \cap \mathbb{S}^{n-1}$ can be written as $\mathbf{v} = \mathbf{Q}\mathbf{a}$ for some $\mathbf{a} \in \mathbb{S}^{d-1}$, it follows that

$$\hat{\mathbf{u}}^T \mathbf{A} \hat{\mathbf{u}} = \max_{\mathbf{v} \in \operatorname{range}(\mathbf{Y}) \cap \mathbb{S}^{n-1}} \mathbf{v}^T \mathbf{A} \mathbf{v} = \max_{\mathbf{a} \in \mathbb{S}^{d-1}} \mathbf{a}^T \mathbf{B} \mathbf{a} = \lambda_1(\mathbf{B})$$

Challenges

- Goal: analyze the guarantee of RSVD w.r.t. (1)

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 $\hat{\mathbf{u}}$;

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

Challenges

- Goal: analyze the guarantee of RSVD w.r.t. (1)

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 $\hat{\mathbf{u}}$;

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

- Prior works [3, 5] are based on sharp estimations of $\|\mathbf{A} - \mathbf{QQ}^T\mathbf{A}\|_2$ which unfortunately leads to additive bounds of (1), and limited to $\mathbf{A} \succcurlyeq 0$.

Challenges

- Goal: analyze the guarantee of RSVD w.r.t. (1)

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 $\hat{\mathbf{u}}$;

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

- Prior works [3, 5] are based on sharp estimations of $\|\mathbf{A} - \mathbf{QQ}^T\mathbf{A}\|_2$ which unfortunately leads to additive bounds of (1), and limited to $\mathbf{A} \succcurlyeq 0$.
- Converting classical metric to (1) by matrix perturbation theory yields not only additive but also eigengap-dependent bounds.

Challenges

- Goal: analyze the guarantee of RSVD w.r.t. (1)

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 $\hat{\mathbf{u}}$;

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

$\Rightarrow$ This probably suggests that to derive a tight analysis of (1) we should avoid matrix subtractions.

Outline

Our starting point

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

$\mathbf{Y} \leftarrow \mathbf{A}^q\mathbf{S}$ 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 $\hat{\mathbf{u}}$;

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

- Recall that $\hat{\mathbf{u}} = \text{argmax}\{\mathbf{v}^T\mathbf{A}\mathbf{v} : \mathbf{v} \in \text{range}(\mathbf{Y}) \cap \mathbb{S}^{n-1}\}$

Our starting point

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

$\mathbf{Y} \leftarrow \mathbf{A}^q\mathbf{S}$ 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 $\hat{\mathbf{u}}$;

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

- Recall that $\hat{\mathbf{u}} = \text{argmax}\{\mathbf{v}^T\mathbf{A}\mathbf{v} : \mathbf{v} = \mathbf{Y}\mathbf{a}, \forall \mathbf{a} \in \mathbb{S}^{d-1}\}$

Our starting point

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

$\mathbf{Y} \leftarrow \mathbf{A}^q\mathbf{S}$ 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 $\hat{\mathbf{u}}$;

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

- Recall that $\hat{\mathbf{u}} = \operatorname{argmax}\{\mathbf{v}^T\mathbf{A}\mathbf{v} : \mathbf{v} = \mathbf{A}^q\mathbf{S}\mathbf{a}, \forall \mathbf{a} \in \mathbb{S}^{d-1}\}$

Our starting point

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

$\mathbf{Y} \leftarrow \mathbf{A}^q\mathbf{S}$ 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 $\hat{\mathbf{u}}$;

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

- Recall that $\hat{\mathbf{u}} = \operatorname{argmax}\{\mathbf{v}^T\mathbf{A}\mathbf{v} : \mathbf{v} = \mathbf{A}^q\mathbf{S}\mathbf{a}, \forall \mathbf{a} \in \mathbb{S}^{d-1}\}$ and use it to rewrite

$$R(\hat{\mathbf{u}}) = \lambda_1^{-1} \max_{\mathbf{a}\in\mathbb{S}^{d-1}} \frac{(\mathbf{Sa})^T\mathbf{A}^{2q+1}(\mathbf{Sa})}{(\mathbf{Sa})^T\mathbf{A}^{2q}(\mathbf{Sa})}$$

Our starting point

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

---

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

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

- Recall that $\hat{\mathbf{u}} = \text{argmax}\{\mathbf{v}^T\mathbf{A}\mathbf{v} : \mathbf{v} = \mathbf{A}^q\mathbf{S}\mathbf{a}, \forall \mathbf{a} \in \mathbb{S}^{d-1}\}$ and use it to rewrite

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

where $\alpha_i = \lambda_i/\lambda_1$, $\forall i \in [n]$.

Our starting point

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

```
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;
```

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

Question

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

Random projection

Definition (projection length)

The projection length of $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}_n\}$ onto a non-empty $\mathcal{X} \subseteq \mathbb{R}^n$ is $\cos\theta(\mathbf{v}, \mathcal{X})$, where

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

For a matrix $\mathbf{X}$, we use $\theta(\mathbf{v}, \mathbf{X})$ to denote $\theta(\mathbf{v}, \mathrm{range}(\mathbf{X}))$.

Random projection

Definition (projection length)

The projection length of $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}_n\}$ onto a non-empty $\mathcal{X} \subseteq \mathbb{R}^n$ is $\cos\theta(\mathbf{v}, \mathcal{X})$, where

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

For a matrix $\mathbf{X}$, we use $\theta(\mathbf{v}, \mathbf{X})$ to denote $\theta(\mathbf{v}, \mathrm{range}(\mathbf{X}))$.

Lemma (Gaussian random projection)

Let $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}_n\}$ and $\mathbf{S} \sim \mathcal{N}(0,1)^{n\times d}$, $d \ll n$. Then,

$$\cos^2\theta(\mathbf{v}, \mathbf{S}) = \Theta\left(\frac{d}{n}\right) \text{ with probability } 1 - e^{-\Omega(d)}.$$

Random projection

Lemma (Gaussian random projection)

Let $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}_n\}$ and $\mathbf{S} \sim \mathcal{N}(0,1)^{n \times d}$, $d \ll n$. Then,

$$\cos^2 \theta(\mathbf{v}, \mathbf{S}) = \Theta\left(\frac{d}{n}\right) \text{ with probability } 1 - e^{-\Omega(d)}.$$

Intuition due to [2]

Let $\mathbf{z}_1, \cdots, \mathbf{z}_d$ sampled uniformly from d-dimensional orthonormal basis. Observe that $\cos^2 \theta(\mathbf{v}, \mathbf{S}) = \sum_{i \in [d]} \langle \mathbf{v}, \mathbf{z}_i \rangle^2$. Then,

- by $\mathbb{E}[\cos^2 \theta(\mathbf{v}, \mathbf{S})] = \sum_{i \in [d]} \mathbb{E}[\langle \mathbf{v}, \mathbf{z}_i \rangle^2]$, and $\mathbb{E}[\langle \mathbf{v}, \mathbf{z}_i \rangle^2] = \frac{1}{n}$, $\forall i \in [d]$,
- we know $\mathbb{E}[\cos^2 \theta(\mathbf{v}, \mathbf{S})] = \frac{d}{n}$.

It remains to show that $\cos^2 \theta(\mathbf{v}, \mathbf{S})$ concentrates tightly around the mean.

Random projection

Lemma (Gaussian random projection)

Let $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}_n\}$ and $\mathbf{S} \sim \mathcal{N}(0,1)^{n \times d}$, $d \ll n$. Then,

$$\cos^2 \theta(\mathbf{v}, \mathbf{S}) = \Theta\left(\frac{d}{n}\right) \text{ with probability } 1 - e^{-\Omega(d)}.$$

Proof sketch (simplified from [4])

For simplicity, assume $\|\mathbf{v}\|_2 = 1$. It follows from $d \ll n$ that $\text{rank}(\mathbf{S}) = d$ a.s.
(i) $\sigma_1(\mathbf{S}) = \Theta(\sqrt{n})$ with prob. $\geq 1 - e^{-\Omega(n)}$,
(ii) $\sigma_d(\mathbf{S}) = \Omega(\sqrt{n} - \sqrt{d-1})$ with prob. $\geq 1 - e^{-\Omega(n-d)}$

Random projection

Lemma (Gaussian random projection)

Let $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}_n\}$ and $\mathbf{S} \sim \mathcal{N}(0,1)^{n \times d}$, $d \ll n$. Then,

$$\cos^2 \theta(\mathbf{v}, \mathbf{S}) = \Theta\left(\frac{d}{n}\right) \text{ with probability } 1 - e^{-\Omega(d)}.$$

Proof sketch (simplified from [4])

For simplicity, assume $\|\mathbf{v}\|_2 = 1$. It follows from $d \ll n$ that $\operatorname{rank}(\mathbf{S}) = d$ a.s.
(i) $\sigma_1(\mathbf{S}) = \Theta(\sqrt{n})$ with prob. $\geq 1 - e^{-\Omega(n)}$,
(ii) $\sigma_d(\mathbf{S}) = \Omega(\sqrt{n} - \sqrt{d-1})$ with prob. $\geq 1 - e^{-\Omega(n-d)}$

$$\cos \theta(\mathbf{v}, \mathbf{S}) = \max_{\mathbf{v} \in \mathbb{S}^{d-1}} \frac{\langle \mathbf{v}, \mathbf{Sa} \rangle}{\|\mathbf{Sa}\|_2}$$

Random projection

Lemma (Gaussian random projection)

Let $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}_n\}$ and $\mathbf{S} \sim \mathcal{N}(0,1)^{n \times d}$, $d \ll n$. Then,

$$\cos^2 \theta(\mathbf{v}, \mathbf{S}) = \Theta\left(\frac{d}{n}\right) \text{ with probability } 1 - e^{-\Omega(d)}.$$

Proof sketch (simplified from [4])

For simplicity, assume $\|\mathbf{v}\|_2 = 1$. It follows from $d \ll n$ that $\text{rank}(\mathbf{S}) = d$ a.s.

(i) $\sigma_1(\mathbf{S}) = \Theta(\sqrt{n})$ with prob. $\geq 1 - e^{-\Omega(n)}$,

(ii) $\sigma_d(\mathbf{S}) = \Omega(\sqrt{n} - \sqrt{d-1})$ with prob. $\geq 1 - e^{-\Omega(n-d)}$

$$\frac{\sigma_1(\mathbf{S}^T\mathbf{v})}{\sigma_1(\mathbf{S})} \overset{(a)}{\leq} \cos\theta(\mathbf{v}, \mathbf{S}) = \max_{\mathbf{v} \in \mathbb{S}^{d-1}} \frac{\langle \mathbf{v}, \mathbf{S}\mathbf{a} \rangle}{\|\mathbf{S}\mathbf{a}\|_2}$$

(a): setting $\mathbf{a} = \mathbf{S}^T\mathbf{v}/\|\mathbf{S}^T\mathbf{v}\|_2$ and that $\sigma_1(\mathbf{S}) \geq \|\mathbf{S}\mathbf{a}\|_2$ for all $\mathbf{a} \in \mathbb{S}^{d-1}$

Random projection

Lemma (Gaussian random projection)

Let $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}_n\}$ and $\mathbf{S} \sim \mathcal{N}(0,1)^{n \times d}$, $d \ll n$. Then,

$$\cos^2 \theta(\mathbf{v}, \mathbf{S}) = \Theta\left(\frac{d}{n}\right) \text{ with probability } 1 - e^{-\Omega(d)}.$$

Proof sketch (simplified from [4])

For simplicity, assume $\|\mathbf{v}\|_2 = 1$. It follows from $d \ll n$ that $\text{rank}(\mathbf{S}) = d$ a.s.

(i) $\sigma_1(\mathbf{S}) = \Theta(\sqrt{n})$ with prob. $\geq 1 - e^{-\Omega(n)}$,

(ii) $\sigma_d(\mathbf{S}) = \Omega(\sqrt{n} - \sqrt{d-1})$ with prob. $\geq 1 - e^{-\Omega(n-d)}$

$$\frac{\sigma_1(\mathbf{S}^T\mathbf{v})}{\sigma_1(\mathbf{S})} \overset{(a)}{\leq} \cos\theta(\mathbf{v}, \mathbf{S}) = \max_{\mathbf{v} \in \mathbb{S}^{d-1}} \frac{\langle \mathbf{v}, \mathbf{Sa} \rangle}{\|\mathbf{Sa}\|_2} \overset{(b)}{\leq} \frac{\sigma_1(\mathbf{S}^T\mathbf{v})}{\sigma_d(\mathbf{S})}$$

(b): $\langle \mathbf{v}, \mathbf{Sa} \rangle \leq \|\mathbf{S}^T\mathbf{v}\|_2 \|\mathbf{a}\|_2$ and that $\sigma_d(\mathbf{S}) \leq \|\mathbf{Sa}\|_2$ for all $\mathbf{a} \in \mathbb{S}^{d-1}$

Random projection

Lemma (Gaussian random projection)

Let $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}_n\}$ and $\mathbf{S} \sim \mathcal{N}(0,1)^{n \times d}$, $d \ll n$. Then,

$$\cos^2 \theta(\mathbf{v}, \mathbf{S}) = \Theta\left(\frac{d}{n}\right) \text{ with probability } 1 - e^{-\Omega(d)}.$$

Proof sketch (simplified from [4])

For simplicity, assume $\|\mathbf{v}\|_2 = 1$. It follows from $d \ll n$ that $\operatorname{rank}(\mathbf{S}) = d$ a.s.
(i) $\sigma_1(\mathbf{S}) = \Theta(\sqrt{n})$ with prob. $\geq 1 - e^{-\Omega(n)}$,
(ii) $\sigma_d(\mathbf{S}) = \Omega(\sqrt{n} - \sqrt{d-1})$ with prob. $\geq 1 - e^{-\Omega(n-d)}$

$$\frac{\sigma_1(\mathbf{S}^T\mathbf{v})}{\sigma_1(\mathbf{S})} \overset{(a)}{\leq} \cos\theta(\mathbf{v}, \mathbf{S}) = \max_{\mathbf{v} \in \mathbb{S}^{d-1}} \frac{\langle \mathbf{v}, \mathbf{S}\mathbf{a} \rangle}{\|\mathbf{S}\mathbf{a}\|_2} \overset{(b)}{\leq} \frac{\sigma_1(\mathbf{S}^T\mathbf{v})}{\sigma_d(\mathbf{S})}$$

Invoking a union bound of (i)(ii) yields the desired.

Random projection

Lemma (Gaussian random projection)

Let $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}_n\}$ and $\mathbf{S} \sim \mathcal{N}(0,1)^{n \times d}$, $d \ll n$. Then,

$$\cos^2 \theta(\mathbf{v}, \mathbf{S}) = \Theta\left(\frac{d}{n}\right) \text{ with probability } 1 - e^{-\Omega(d)}.$$

Remark

Many interesting results are derived from (i)(ii).

(i) $\sigma_1(\mathbf{S}) = \Theta(\sqrt{n})$ w.h.p. (ii) $\sigma_d(\mathbf{S}) = \Omega(\sqrt{n} - \sqrt{d-1})$ w.h.p.

For example, let $\mathbf{T} = \sqrt{\frac{1}{d}}\mathbf{S}$

Random projection

Lemma (Gaussian random projection)

Let $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}_n\}$ and $\mathbf{S} \sim \mathcal{N}(0,1)^{n\times d}$, $d \ll n$. Then,

$$\cos^2\theta(\mathbf{v},\mathbf{S}) = \Theta\left(\frac{d}{n}\right) \text{ with probability } 1 - e^{-\Omega(d)}.$$

Remark

Many interesting results are derived from (i)(ii).

(i) $\sigma_1(\mathbf{S}) = \Theta(\sqrt{n})$ w.h.p. (ii) $\sigma_d(\mathbf{S}) = \Omega(\sqrt{n} - \sqrt{d-1})$ w.h.p.

For example, let $\mathbf{T} = \sqrt{\frac{1}{d}}\mathbf{S}$ and $\mathbf{T}$ satisfies

- *restricted isometry property*, i.e., $\|\mathbf{T}^T\mathbf{v}\|_2 = (1 \pm \epsilon)\|\mathbf{v}\|_2$ w.h.p.

Random projection

Lemma (Gaussian random projection)

Let $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}_n\}$ and $\mathbf{S} \sim \mathcal{N}(0,1)^{n \times d}$, $d \ll n$. Then,

$$\cos^2 \theta(\mathbf{v}, \mathbf{S}) = \Theta\left(\frac{d}{n}\right) \text{ with probability } 1 - e^{-\Omega(d)}.$$

Remark

Many interesting results are derived from (i)(ii).

(i) $\sigma_1(\mathbf{S}) = \Theta(\sqrt{n})$ w.h.p. (ii) $\sigma_d(\mathbf{S}) = \Omega(\sqrt{n} - \sqrt{d-1})$ w.h.p.

For example, let $\mathbf{T} = \sqrt{\frac{1}{d}}\mathbf{S}$ and $\mathbf{T}$ satisfies

- *restricted isometry property*, i.e., $\|\mathbf{T}^T\mathbf{v}\|_2 = (1 \pm \epsilon)\|\mathbf{v}\|_2$ w.h.p.
- *Johnson-Lindenstrauss Lemma*, i.e., $\|\mathbf{T}^T\mathbf{v}_i - \mathbf{T}^T\mathbf{v}_j\|_2 = (1 \pm \epsilon)\|\mathbf{v}_i - \mathbf{v}_j\|_2$ w.h.p. for any fixed set of N unit vectors $\{\mathbf{v}_i\}_{i \in [N]} \subseteq \mathbb{R}^n$ and $d = \Omega(\epsilon^{-2} \ln N)$

Positive semidefinite matrices

Assume $\mathbf{A} \succcurlyeq 0$, i.e., $\{\alpha_i\}_{i\in[n]}$ are nonnegative, and $d \ll n$.

Lemma (Gaussian random projection)

For any $\mathbf{v} \in \mathbb{R}^n \backslash \{\mathbf{0}_n\}$, $\cos^2\theta(\mathbf{v}, \mathbf{S}) = \Theta\left(\frac{d}{n}\right)$ with prob. $1 - e^{-\Omega(d)}$.

Question

How can we use the lemma to analyze $R(\hat{\mathbf{u}}) = \max_{\mathbf{a}\in\mathbb{S}^{d-1}} \frac{\sum_{i\in[n]} \alpha_i^{2q+1} \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2}{\sum_{i\in[n]} \alpha_i^{2q} \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2}$?

Hint: $\cos^2\theta(\mathbf{v}, \mathbf{S}) = \max_{\mathbf{a}\in\mathbb{S}^{d-1}} \frac{\langle \mathbf{S}^T\mathbf{v}, \mathbf{a}\rangle^2}{\sum_{i\in[n]} \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2}$ for any fixed $\mathbf{v} \in \mathbb{S}^{n-1}$.

Positive semidefinite matrices

Goal: analyze $R(\hat{\mathbf{u}}) = \max_{\mathbf{a}\in\mathbb{S}^{d-1}} \frac{\sum_{i\in[n]} \alpha_i^{2q+1} \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2}{\sum_{i\in[n]} \alpha_i^{2q} \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2}$ given p.s.d. $\mathbf{A}$ and $d \ll n$

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

Fix any $\mathbf{a} \in \mathbb{S}^{d-1}$. Applying Cauchy inequality repeatedly yields:

$$R_{\mathbf{a}} := \frac{\sum_{i\in[n]} \alpha_i^{2q+1} \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2}{\sum_{i\in[n]} \alpha_i^{2q} \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2} \geq \cdots \geq \frac{\sum_{i\in[n]} \alpha_i \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2}{\sum_{i\in[n]} \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2}. \tag{2}$$

Positive semidefinite matrices

Goal: analyze $R(\hat{\mathbf{u}}) = \max_{\mathbf{a}\in\mathbb{S}^{d-1}} \frac{\sum_{i\in[n]} \alpha_i^{2q+1}\langle \mathbf{S}^T\mathbf{u}_i,\mathbf{a}\rangle^2}{\sum_{i\in[n]} \alpha_i^{2q}\langle \mathbf{S}^T\mathbf{u}_i,\mathbf{a}\rangle^2}$ given p.s.d. $\mathbf{A}$ and $d \ll n$

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

Fix any $\mathbf{a} \in \mathbb{S}^{d-1}$. Applying Cauchy inequality repeatedly yields:

$$R_{\mathbf{a}} := \frac{\sum_{i\in[n]} \alpha_i^{2q+1}\langle \mathbf{S}^T\mathbf{u}_i,\mathbf{a}\rangle^2}{\sum_{i\in[n]} \alpha_i^{2q}\langle \mathbf{S}^T\mathbf{u}_i,\mathbf{a}\rangle^2} \geq \cdots \geq \frac{\sum_{i\in[n]} \alpha_i\langle \mathbf{S}^T\mathbf{u}_i,\mathbf{a}\rangle^2}{\sum_{i\in[n]}\langle \mathbf{S}^T\mathbf{u}_i,\mathbf{a}\rangle^2}. \tag{2}$$

Rearranging (2) repeatedly leads to

$$\sum_{i\in[n]}\langle \mathbf{S}^T\mathbf{u}_i,\mathbf{a}\rangle^2 \geq R_{\mathbf{a}}^{-1}\sum_{i\in[n]}\alpha_i\langle \mathbf{S}^T\mathbf{u}_i,\mathbf{a}\rangle^2 \geq \cdots \geq R_{\mathbf{a}}^{-(2q+1)}\sum_{i\in[n]}\alpha_i^{2q+1}\langle \mathbf{S}^T\mathbf{u}_i,\mathbf{a}\rangle^2$$

Positive semidefinite matrices

Goal: analyze $R(\hat{\mathbf{u}}) = \max_{\mathbf{a}\in\mathbb{S}^{d-1}} \frac{\sum_{i\in[n]} \alpha_i^{2q+1} \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2}{\sum_{i\in[n]} \alpha_i^{2q} \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2}$ given p.s.d. $\mathbf{A}$ and $d \ll n$

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

Fix any $\mathbf{a} \in \mathbb{S}^{d-1}$. Rearranging (2) repeatedly leads to

$$\sum_{i\in[n]} \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2 \geq R_{\mathbf{a}}^{-1} \sum_{i\in[n]} \alpha_i \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2 \geq \cdots \geq R_{\mathbf{a}}^{-(2q+1)} \sum_{i\in[n]} \alpha_i^{2q+1} \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2$$

which implies that

$$R_{\mathbf{a}}^{2q+1} \geq \frac{\sum_{i\in[n]} \alpha_i^{2q+1} \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2}{\sum_{i\in[n]} \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2} \geq \frac{\langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2}{\sum_{i\in[n]} \langle \mathbf{S}^T\mathbf{u}_i, \mathbf{a}\rangle^2}$$

Positive semidefinite matrices

Goal: analyze $R(\hat{\mathbf{u}}) = \max_{\mathbf{a}\in\mathbb{S}^{d-1}} \frac{\sum_{i\in[n]} \alpha_i^{2q+1}\langle \mathbf{S}^T\mathbf{u}_i,\mathbf{a}\rangle^2}{\sum_{i\in[n]} \alpha_i^{2q}\langle \mathbf{S}^T\mathbf{u}_i,\mathbf{a}\rangle^2}$ given p.s.d. $\mathbf{A}$ and $d \ll n$

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

Fix any $\mathbf{a}\in\mathbb{S}^{d-1}$. Rearranging (2) repeatedly leads to

$$\sum_{i\in[n]}\langle \mathbf{S}^T\mathbf{u}_i,\mathbf{a}\rangle^2 \geq R_{\mathbf{a}}^{-1}\sum_{i\in[n]}\alpha_i\langle \mathbf{S}^T\mathbf{u}_i,\mathbf{a}\rangle^2 \geq \cdots \geq R_{\mathbf{a}}^{-(2q+1)}\sum_{i\in[n]}\alpha_i^{2q+1}\langle \mathbf{S}^T\mathbf{u}_i,\mathbf{a}\rangle^2$$

which implies that

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

Our results: positive semidefinite matrices

(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. $\geq 1 - e^{-\Omega(d)}$.

Our results: positive semidefinite matrices

(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. $\geq 1 - e^{-\Omega(d)}$.

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

- Proof: consider $\alpha_i = \left(\frac{d}{n}\right)^{1/(2q+1)}$, $\forall i \geq 2$ and use Gaussian projection lemma

Our results: positive semidefinite matrices

(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. $\geq 1 - e^{-\Omega(d)}$.

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

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

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

[a] (i_0, γ)-power-law decay implies there exists constant $C > 0$ such that $\frac{\sigma_i}{\sigma_1} \leq C \cdot i^{-\gamma}$ for all $i \geq i_0$.

Our results: indefinite matrices

Assumption 1

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

(Theorem 4) For **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: RandSum

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

If you know $\langle \mathbf{u}_1, \mathbf{1} \rangle^2 = \Theta(n)$, is there a better choice of $\mathbf{S}$?

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

Extension: RandSum

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

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

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

Extension: RandSum

Assumption 2

There exists a constant $\kappa' \in (0,1]$ such that $\sum_{i=2}^{n} \alpha_i^{2q+1} \xi_i \geq \kappa' \sum_{i=2}^{n} |\alpha_i|^{2q+1} \xi_i$ and $\langle \mathbf{u}_1, \mathbf{1}_n \rangle^2 = \Omega(1)$, where $\xi_i = \mathbb{E}\left[\langle \mathbf{S}^T \mathbf{u}_i, \frac{\mathbf{1}_d}{\sqrt{d}} \rangle^2\right]$, $\forall i \in [n]$.

(Theorem 6) For $\mathbf{A}$ with (i_0, γ)-power-law decay, $i_0 \in [n]$ and $\gamma > 1/2q$, and satisfying Assumption 1, and 2, RandSum($\mathbf{A}$,q,d,p) returns $\hat{\mathbf{u}}$ satisfying

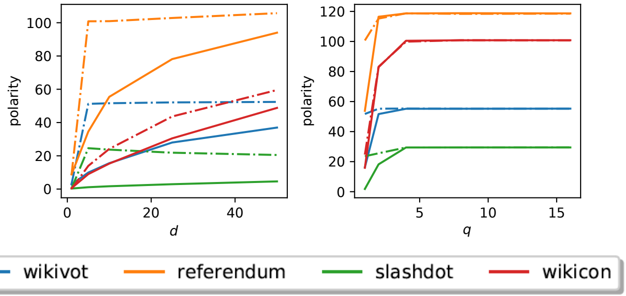
$$R(\hat{\mathbf{u}}) = \Omega\left(\left(\max\left\{\frac{d}{d+i_0}, \frac{\langle \mathbf{u}_1, \mathbf{1}_n \rangle^2}{n}\right\}\right)^{\frac{1}{2q+1}}\right) \text{ with prob. } \geq 1 - e^{-\Omega(\sqrt{d})}.$$

(the dependency on κ, κ' are hidden here for simplicity).

Outline

Experiment: 2-conflicting group detection [1, 10]

	WikiVot	Referendum	Slashdot	WikiCon
$\lvert V\rvert$	7 115	10 884	82 140	116 717
$\lvert E\rvert$	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 (1) for indefinite matrices.
- Study the property of Bernoulli random projection and demonstrate its usefulness to the task of conflicting group detection [1, 10].

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] Sanjoy Dasgupta and Anupam Gupta.
An elementary proof of a theorem of johnson and lindenstrauss.
Random Structures & Algorithms, 2003.

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

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

Reference II

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

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

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

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

Reference III

[9] 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.

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