Towards some RandNLA Techniques for Determinant Approximation, Sparse PCA and Analysis of Krylov Subspace Methods

Eugenia-Maria Kontopoulou & Petros Drineas

Computer Science, Purdue University

LogDet Problem

Given an SPD matrix $A \in \mathbb{R}^{n\times n}$, compute (exactly or approximately) $\log\det(A)$.

Additive Error Approximation

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be an SPD matrix. For any α with $\lambda_1(\mathbf{A}) < \alpha$, define $\mathbf{B} = \mathbf{A}/\alpha$ and $\mathbf{C} = \mathbf{I}_n - B$. Then,

$$\log\det(A) = n\log(\alpha) - \sum_{k=1}^{\infty} \frac{\mathbf{Tr}\left(\log\left(\mathbf{C}^k\right)\right)}{k}.$$

Algorithm 1

Input: $\mathbf{A} \in \mathbb{R}^{n\times n}$, accuracy parameter $\epsilon > 0$, integer $m > 0$.

1. Compute an estimate to the largest eigenvalue of $\mathbf{A}$,$\lambda_1(\tilde{\mathbf{A}})$, using the Power Method.
2. $\mathbf{C} = \mathbf{I}_n - \mathbf{A}/(7\lambda_1(\tilde{\mathbf{A}}))$
3. Create $p = \lceil 20\log(2/\delta)/\epsilon^2 \rceil$ i.i.d random Gaussian vectors, $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_p$.
4. Estimate $\sum_{k=1}^{\infty} \frac{\mathbf{Tr}\left(\log\left(\mathbf{C}^k\right)\right)}{k}$ with a truncated Taylor Series type randomized trace estimator that computes $\sum_{k=1}^{m}\left(\frac{1}{p}\sum_{i=1}^{p} \mathbf{g}_i^\top \mathbf{C}^k \mathbf{g}_i\right)$

Let $\widehat{\log\det}(\mathbf{A})$ be the $\log\det$ approximation of the above procedure. Then, we **prove** that with probability at least $1-2\delta$,

$$|\widehat{\log\det}(\mathbf{A}) - \log\det(A)| \leq 2\epsilon\Gamma$$

where $\Gamma = \sum_{i=1}^{n} \log\left(7 \cdot \frac{\lambda_1(\mathbf{A})}{\lambda_i(\mathbf{A})}\right)$ and $m \geq \lceil 7\kappa(\mathbf{A})\log(\frac{1}{\epsilon}) \rceil$.

Relative Error Approximation

Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be an SPD matrix whose eigenvalues lie in the interval $(\theta_1, 1)$, for some $0 < \theta_1 < 1$. Let $\mathbf{C} = \mathbf{I}_n - A$. Then,

$$\log\det(A) = -\sum_{k=1}^{\infty} \frac{\mathbf{Tr}\left(\log\left(\mathbf{C}^k\right)\right)}{k}.$$

Algorithm 2

Input: $\mathbf{A} \in \mathbb{R}^{n\times n}$, accuracy parameter $\epsilon > 0$, integer $m > 0$.

1. $\mathbf{C} = \mathbf{I}_n - \mathbf{A}$
2. Create $p = \lceil 20\log(2/\delta)/\epsilon^2 \rceil$ i.i.d random Gaussian vectors, $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_p$.
3. Estimate $\sum_{k=1}^{\infty} \frac{\mathbf{Tr}\left(\log\left(\mathbf{C}^k\right)\right)}{k}$ with a truncated Taylor Series type randomized trace estimator that computes $\sum_{k=1}^{m}\left(\frac{1}{p}\sum_{i=1}^{p} \mathbf{g}_i^\top \mathbf{C}^k \mathbf{g}_i\right)$

Let $\widehat{\log\det}(\mathbf{A})$ be the $\log\det$ approximation of the above procedure on inputs $\mathbf{A}$ and ϵ. Then, we **prove** that with probability at least $1-\delta$,

$$|\widehat{\log\det}(\mathbf{A}) - \log\det(A)| \leq 2\epsilon \cdot |\log\det(\mathbf{A})|.$$

Citation

C. Boutsidis, P. Drineas, P. Kambadur, E. Kontopoulou, A. Zouzias (2016), *A Randomized Algorithm for Approximating the Log Determinant of a Symmetric Positive Definite Matrix*, under review at Journal of Linear Algebra and its Applications.
ArXiv: **https://arxiv.org/abs/1503.00374**

Sparse PCA

Given a centered matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$ (the mean of its columns is zero), we seek for a vector $\mathbf{w}_{opt}$ that solves the optimization problem:

$$\begin{aligned}&\underset{\mathbf{w}}{\text{maximize}}\ \ \mathbf{w}^\top\mathbf{X}^\top\mathbf{X}\ \mathbf{w}\\ &\text{subject to } \|\mathbf{w}\|_0 \leq k,\ \|\mathbf{w}\|_2 \leq 1,\ \mathbf{w} \in \mathbb{R}^n.\end{aligned}$$

This problem is **NP-hard** $\rightarrow$ relax to a **problem with convex constraints (but non-convex objective)**:

$$\begin{aligned}&\underset{\mathbf{w}}{\text{maximize}}\ \ \mathbf{w}^\top\mathbf{X}^\top\mathbf{X}\ \mathbf{w}\\ &\text{subject to } \|\mathbf{w}\|_1 \leq \sqrt{k},\ \|\mathbf{w}\|_2 \leq 1,\ \mathbf{w} \in \mathbb{R}^n.\end{aligned}$$

Algorithm

Phase 1: Compute a stationary point $\tilde{\mathbf{w}}_{opt}$

1. Compute the gradient and make a gradient step.
2. Project onto the l_1 ball with radius $\sqrt{k}$ $(\|\mathbf{w}\|_1)$.
3. Repeat until a threshold for the relative error is exceeded.

Phase 2: Invoke a **randomized rounding strategy**.

1. Create a Bernoulli distribution and randomly round the entries of $\mathbf{w}$.
2. Repeat the experiment 10 times and keep the best sparsification.

We prove the following:

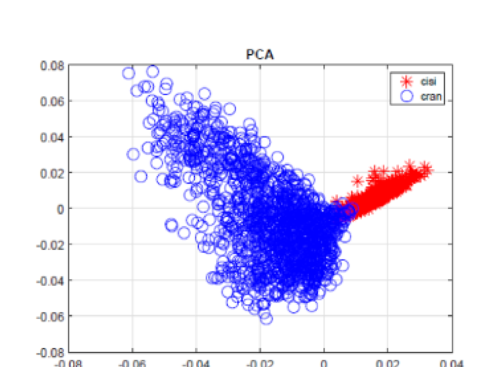
Let $\mathbf{w}_{opt}$ be the optimal solution of the Sparse PCA problem (1) satisfying $\|\mathbf{w}_{opt}\|_2 = 1$ and $\|\mathbf{w}_{opt}\|_0 \leq k$. Let $\hat{\mathbf{w}}_{opt}$ be the vector returned when the rounding sparsification strategy is applied on the optimal solution $\tilde{\mathbf{w}}_{opt}$ of the optimization problem (1), with $s = 200k/\epsilon^2$, where $\epsilon \in (0,1]$ is an accuracy parameter. Then, $\hat{\mathbf{w}}_{opt}$ has the following properties:

1. $\mathbb{E}\|\hat{\mathbf{w}}_{opt}\|_0 \leq s$.
2. With probability at least 3/4,
$$\|\hat{\mathbf{w}}_{opt}\|_2 \leq 1 + 0.15\epsilon.$$
3. With probability at least 3/4,
$$\hat{\mathbf{w}}_{opt}^\top\mathbf{A}\hat{\mathbf{w}}_{opt} \geq \mathbf{w}_{opt}^\top\mathbf{A}\mathbf{w}_{opt} - \epsilon.$$

Citation

K. Fountoulakis, A. Kundu, E. Kontopoulou, P. Drineas (2016), *A Randomized Rounding Algorithm for Sparse PCA*, under review at ACM Transactions on Knowledge Discovery from Data.
ArXiv: **https://arxiv.org/abs/1508.03337**

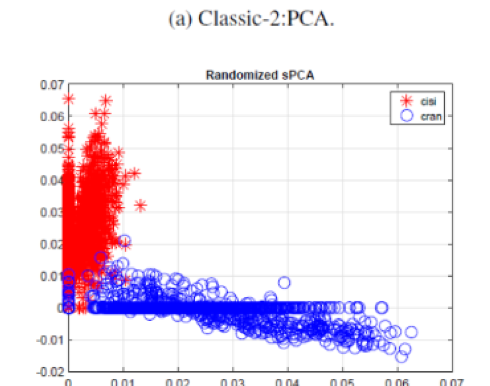
Experiments



(a) Classic-2:PCA.

Datasets

- **Synthetic:** $m = 2^7$, $n = 2^{12}$
- **Classic-2:** $m = 2,858$ documents , $n = 12,427$ terms
  1. CISI collection (1,460 information retrieval abstracts)
  2. CRANFIELD collection (1,398 aeronautical systems abstracts)



(b) Classic-2: randomized sPCA

Table 1: Variance captured by the principal components

	pca	cvx	rspca
Top Principal Comp.	41.9%	41.8%	40.8%
Top two Principal Comp.	61.7%	61.5%	59.7%

Table 2: Sparsity of the principal components

	cvx	rspca
Top Principal Comp.	65%	91%
2nd Top Principal Comp.	9%	89%

Krylov Methods

Given a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ and a starting guess matrix $\mathbf{X} \in \mathbb{R}^{n\times s}$, we want to use the block Krylov space $\mathcal{K}_q(\mathbf{A}\mathbf{A}^\top, \mathbf{A}\mathbf{X})$ to approximate the **left singular vector space** of $\mathbf{A}$.

We prove:

- Spectral & Frobenius bounds for the distance between the approximate and the actual space.
- Quality measurements of the bounds relative to the best low-rank approximation.

Theorem

Let $\phi(x)$ be a polynomial of degree $2q+1$ with odd powers only, such that $\phi(\mathbf{\Sigma}_k)$ is nonsingular. If $\text{rank}(\mathbf{V}_k^\top\mathbf{X}) = k$ then

$$\|\sin\mathbf{\Theta}(\mathcal{K}_q, \mathbf{U}_k)\|_{2,F} \leq \|\phi(\mathbf{\Sigma}_{k,\perp})\|_2\|\phi(\mathbf{\Sigma}_k)^{-1}\|_2\|\mathbf{V}_{k,\perp}^\top\mathbf{X}(\mathbf{V}_k^\top\mathbf{X})^\dagger\|_{2,F}.$$

If, in addition, $\mathbf{X}$ has orthonormal or linearly independent columns, then

$$\|\mathbf{V}_{k,\perp}^\top\mathbf{X}(\mathbf{V}_k^\top\mathbf{X})^\dagger\|_{2,F} = \|\tan\mathbf{\Theta}(\mathbf{X}, \mathbf{V}_k)\|_{2,F}$$

and

$$\|\sin\mathbf{\Theta}(\mathcal{K}_q, \mathbf{U}_k)\|_{2,F} \leq \|\phi(\mathbf{\Sigma}_{k,\perp})\|_2\|\phi(\mathbf{\Sigma}_k)^{-1}\|_2\tan\mathbf{\Theta}(\mathbf{X}, \mathbf{V}_k)\|_{2,F}.$$

where $\mathbf{\Theta}(\mathcal{K}_q, \mathbf{U}_k) \in \mathbb{R}^{k\times k}$ is the diagonal matrix of principal angles between $\mathcal{K}_q$ and range$(\mathbf{U}_k)$.

Theorem

Let $\phi(x)$ be a polynomial of degree $2q+1$ with odd powers only, such that $\phi(\mathbf{\Sigma}_k)$ is nonsingular and $\phi(\sigma_i) \geq \sigma_i$, for $1 \leq i \leq k$. If $\text{rank}(\mathbf{V}_k^\top\mathbf{X}) = k$ then for $1 \leq i \leq k$,

$$\|A - \hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^\top A\|_F \leq \|\mathbf{A} - \mathbf{A}_i\|_F + \Delta$$

$$\|A - \hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^\top A\|_2 \leq \|\mathbf{A} - \mathbf{A}_i\|_2 + \Delta$$

$$\sigma_i - \Delta \leq \|\hat{u}_i^\top A\|_2 \leq \sigma_i.$$

If, in addition, $\mathbf{X}$ has orthonormal columns, then:

$$\Delta = \|\phi(\mathbf{\Sigma}_{k,\perp})\|_2\|\tan\mathbf{\Theta}(\mathbf{X}, \mathbf{V}_k)\|_F$$

Citation

I. Ipsen, P. Drineas, E. Kontopoulou, M. Magdon-Ismail (2016), *Structural Convergence Results for Low-Rank Approximations from Block Krylov Spaces*, submitted to SIAM Journal on Matrix Analysis and Applications.

LogDet Experiments

name	n	nnz	area of origin
thermal2	1228045	8580313	Thermal
ecology2	999999	4995991	2D/3D
ldoor	952203	42493817	Structural
thermomech_TC	102158	711558	Thermal
boneS01	127224	5516602	Model reduction

logdet $(\mathbf{A})$			time (sec)		m
exact	approx		exact	approx	
	mean	std		mean	
1.3869e6	1.3928e6	964.79	31.28	31.24	149
3.3943e6	3.403e6	1212.8	18.5	10.47	125
1.4429e7	1.4445e7	1683.5	117.91	17.60	33
-546787	-546829.4	553.12	57.84	2.58	77
1.1093e6	1.106e6	247.14	130.4	8.48	125