Structural Convergence Results for Low-Rank Approximations from Block Krylov Spaces

Eugenia-Maria Kontopoulou

in collaboration with

P. Drineas, I. Ipsen, & M. Magdon-Ismail

Department of Computer Science
Purdue University

Householder Symposium XX

Blacksburg, June 2017

Given an arbitrary matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and a starting random guess $\mathbf{X} \in \mathbb{R}^{n \times s}$, we build the Krylov space in $\mathbf{AA}^\top$ and $\mathbf{AX}$:

$$\mathcal{K}_q \equiv \mathcal{K}_q(\mathbf{AA}^\top, \mathbf{AX}) = \mathrm{range}(\mathbf{AX} \quad (\mathbf{AA}^\top)\mathbf{AX} \quad \dots \quad (\mathbf{AA}^\top)^q \mathbf{AX}).$$

Assumptions

1. We assume exact arithmetic (there are no issues of numerical stability).
2. The dimension of the Krylov Space is maximal: $\dim(\mathcal{K}_q) = (q+1)s$.
3. $\sigma_k > \sigma_{k+1} > 0$, where k is the number of singular vectors we seek to approximate and σ_k (σ_{k+1}) is the k-th ($k+1$-st) singular value of $\mathbf{A}$.

Assume $\mathbf{A} \in \mathbb{R}^{m \times n}$ and a positive integer $k < \mathrm{rank}(\mathbf{A})$. Let $\mathbf{U}_k$ be the top-k left singular vectors of $\mathbf{A}$. The objective is to construct approximations $\widehat{\mathbf{U}}_k \in \mathbb{R}^{m \times k}$ for $\mathbf{U}_k$.

Dominant Subspace Reconstruction

We are interested in the angles between $\mathrm{range}(\mathbf{U}_k)$ and $\mathrm{range}(\widehat{\mathbf{U}}_k)$. This metric is well defined only if $\mathbf{U}_k$ is unique.

Low Rank Approximation

We are interested in the approximation error between $\mathbf{A}$ and its projection into the space spanned by $\widehat{\mathbf{U}}_k$:

$$\|\mathbf{A} - \widehat{\mathbf{U}}_k \widehat{\mathbf{U}}_k^\top \mathbf{A}\|_{2,F}.$$

This metric is well-defined even if $\mathbf{U}_k$ is not unique.

Prior Work & Motivation

(MM15) Gap-dependent bounds and gap-independent bounds for subspace iteration and low-rank approximations from block Krylov spaces.

- ✓ The gap-dependent bounds can be seen as a special case of our results.
- ✓ In the case of gap-independent bounds, they can only prove bounds for $\left\|\mathbf{A} - \widehat{\mathbf{U}}_k \widehat{\mathbf{U}}_k^\top \mathbf{A}\right\|_{2,F}$.

(HMT11) & (Woo14) Analysis of the subspace iteration. There are no bounds for the angles between $\mathrm{range}(\mathbf{U}_k)$ and $\mathrm{range}(\widehat{\mathbf{U}}_k)$.

(WZZ15) Similar in spirit to (MM15) but they also include gap-dependent bounds for the angles between $\mathrm{range}(\mathbf{U}_k)$ and $\mathrm{range}(\widehat{\mathbf{U}}_k)$ with proof techniques similar to ours.

Prior Work & Motivation

(MM15) Gap-dependent bounds and gap-independent bounds for subspace iteration and low-rank approximations from block Krylov spaces.

- ✓ The gap-dependent bounds can be seen as a special case of our results.
- ✓ In the case of gap-independent bounds, they can only prove bounds for $\left\|\mathbf{A} - \widehat{\mathbf{U}}_k \widehat{\mathbf{U}}_k^\top \mathbf{A}\right\|_{2,F}$.

(HMT11) & (Woo14) Analysis of the subspace iteration. There are no bounds for the angles between $\mathrm{range}(\mathbf{U}_k)$ and $\mathrm{range}(\widehat{\mathbf{U}}_k)$.

(WZZ15) Similar in spirit to (MM15) but they also include gap-dependent bounds for the angles between $\mathrm{range}(\mathbf{U}_k)$ and $\mathrm{range}(\widehat{\mathbf{U}}_k)$ with proof techniques similar to ours.

Our primary focus is to understand the quality of the approximation to the top-k left singular vectors of $\mathbf{A}$ from a block Krylov space.

Basic Notation

Dominant and Subdominant Spaces

Let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ be the full SVD of $\mathbf{A}$ with $\mathbf{U} \in \mathbb{R}^{m\times m}$, $\mathbf{\Sigma} \in \mathbb{R}^{m\times n}$ and $\mathbf{V} \in \mathbb{R}^{n\times n}$, then for an integer $0 < k < \mathrm{rank}(\mathbf{A})$ we can perform the following partitioning:

$$\mathbf{A} = \underbrace{\mathbf{U}_k\mathbf{\Sigma}_k\mathbf{V}_k^\top}_{\text{dominant spaces}} + \underbrace{\mathbf{U}_{k,\perp}\mathbf{\Sigma}_{k,\perp}\mathbf{V}_{k,\perp}^\top}_{\text{sub-dominant spaces}}$$

Principal Angles Matrix

Assume $\mathbf{U}_k \in \mathbb{R}^{m\times k}$ and $\mathbf{X} \in \mathbb{R}^{m\times s}$, with orthonormal columns:

Principal Angles:

$$\theta_i = \cos^{-1}(\sigma_i(\mathbf{U}_k^\top \mathbf{X}))$$

Principal Angles Matrix:

$$\mathbf{\Theta}(\mathbf{U}_k, \mathbf{X}) = \mathrm{diag}(\theta_1, \theta_2, \ldots, \theta_k) \in \mathbb{R}^{k\times k}.$$

Theorem 1

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

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

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

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

and

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

Selecting the Starting Guess X

The starting guess $\mathbf{X}$

The starting guess $\mathbf{X}$ can be any random matrix, e.g. random Gaussian, random sign, sub-sampled randomized Hadamard transform.

RandNLA: bounds for $\| \tan \Theta(\mathbf{X}, \mathbf{V}_k) \|_{2,F}$

Much work on RandNLA has been focused on bounding $\| \tan \Theta(\mathbf{X}, \mathbf{V}_k) \|_{2,F}$ using matrix concentration inequalities (e.g. matrix Chernoff, matrix Bernstein, matrix Hoeffding inequalities).

Full rank of $V_k^\top X$

It guarantees that $\mathrm{range}(V_k)$ and $\mathrm{range}(X)$ are sufficiently close and all principal angles between them are less than $\pi/2$.

Exact Arithmetic Algorithm

to constuct approximations for $\mathbf{U}_k$ from $\mathcal{K}_q$

Input: $\mathbf{A} \in \mathbb{R}^{m \times n}$, starting guess $\mathbf{X} \in \mathbb{R}^{n \times s}$
Target rank $k < \mathrm{rank}(\mathbf{A})$, and assume $\sigma_k > \sigma_{k+1}$
Block dimension $q \geq 1$ with $k \leq (q+1)s \leq m$

Output: $\hat{\mathbf{U}}_k \in \mathbb{R}^{m \times k}$ with orthonormal columns

1: Set $\mathbf{K}_q = \begin{pmatrix} \mathbf{AX} & (\mathbf{AA}^\top)\mathbf{AX} & \cdots & (\mathbf{AA}^\top)^q \mathbf{AX} \end{pmatrix} \in \mathbb{R}^{m \times (q+1)s}$, and assume that $\mathrm{rank}(\mathbf{K}_q) = (q+1)s$.

2: Run an exact arithmetic Rayleigh-Ritz procedure to find the approximation $\mathbf{U}_{W,k}$ of the top k left singular vectors of $\mathbf{W} \in \mathbb{R}^{(q+1)s \times k}$ i.e. the projection of $\mathbf{A}$ into the orthonormal basis, $\mathbf{U}_{K_q}$, of $\mathrm{range}(K_q)$.

3: Return $\hat{\mathbf{U}}_k = \mathbf{U}_{K_q} \mathbf{U}_{W,k} \in \mathbb{R}^{m \times k}$.

Theorem 2

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 $\mathrm{rank}(\mathbf{V}_k^T\mathbf{X}) = k$,

$$\|\mathbf{A} - \hat{\mathbf{U}}_k\hat{\mathbf{U}}_k^T\mathbf{A}\|_{2,F} \quad \leq \quad \|\mathbf{A} - \mathbf{U}_k\mathbf{U}_k^\top\mathbf{A}\|_{2,F} + \|\phi(\mathbf{\Sigma}_{k,\perp})\|_2 \, \|\tan\mathbf{\Theta}(\mathbf{X}, \mathbf{V}_k)\|_F.$$

Selecting the Polynomial $\phi(x)$

Gap-amplifying polynomials

A gap-amplifying polynomial satisfies the following three properties:

- ✓ the small values remain small,
- ✓ the large values are amplified, and
- ✓ the large values are growing super-linearly.

We use rescaled Chebyshev-based gap-amplifying polynomials of the form:

$$\phi(x) = \frac{(1+\gamma)\alpha}{\psi_{q'}(1+\gamma)} \psi_{q'}(x/\alpha),$$

where

$$\gamma = \frac{\sigma_k - \sigma_{k+1}}{\sigma_{k+1}},$$

$q' = 2q + 1$, $x \in [0, \alpha]$ and $\psi_{q'}(x)$ is the Chebyshev polynomial of first kind.

Obtaining a Relative Error

Choice of the degree q

Let $\epsilon > 0$ be an accuracy parameter. If we select

$$q \geq \frac{1}{2\sqrt{\gamma}} \log_2 \frac{4\|\tan \boldsymbol{\Theta}(\mathbf{X}, \mathbf{V}_k)\|_2}{\epsilon},$$

where $\gamma = \frac{\sigma_k - \sigma_{k+1}}{\sigma_{k+1}}$, then the bounds of Theorem 2 become relative:

$$\|\mathbf{A} - \hat{\mathbf{U}}_k \hat{\mathbf{U}}_k^T \mathbf{A}\|_{2,F} \quad \leq \quad (1+\epsilon)\sigma_{k+1}.$$

Remember that:

$$\sigma_{k+1} = \|\mathbf{A} - \mathbf{U}_k {\mathbf{U}_k}^T \mathbf{A}\|_2 \leq \|\mathbf{A} - \mathbf{U}_k {\mathbf{U}_k}^T \mathbf{A}\|_F$$

Proof Techniques

We combine:

- ✓ traditional Lanczos convergence analysis (Saa11), with
- ✓ optimal low-rank approximations via least squares problems (BDM11; BDM14).

Theorem 1 We connect principal angles with least squares residuals.

Theorem 2 We use least squares residuals to interpret orthogonal projections.

Open Problems

- Is it possible to drop the assumption that $\mathbf{V}_k^\top \mathbf{X}$ is full-rank?
- Are our bounds tight enough to be informative?
- Can our bounds be useful in implementing block-Lanczos type methods?

Thank you!

Questions?

References I

(BDM11) C. Boutsidis, P. Drineas, and M. Magdon-Ismail. ``Near-optimal column-based matrix reconstruction''. In: *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science--FOCS 2011*. IEEE Computer Soc., Los Alamitos, CA, 2011, pp. 305-314.

(BDM14) C. Boutsidis, P. Drineas, and M. Magdon-Ismail. ``Near-optimal column-based matrix reconstruction''. In: *SIAM J. Comput.* 43.2 (2014), pp. 687-717.

(HMT11) N. Halko, P. G. Martinsson, and J. A. Tropp. ``Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions''. In: *SIAM Rev.* 53.2 (2011), pp. 217-288.

(MM15) C. Musco and C. Musco. ``Randomized Block Krylov Methods for Stronger and Faster Approximate Singular Value Decomposition''. In: *Advances in Neural Information Processing Systems 28*. Ed. by C. Cortes et al. Curran Associates, Inc., 2015, pp. 1396-1404.

(Saa11) Y. Saad. *Numerical Methods for Large Eigenvalue Problems*. Revised. Classics in Applied Mathematics. Philadelphia: SIAM, 2011.

References II

(Woo14) D. Woodruff. ``Sketching As a Tool for Numerical Linear Algebra''. In: *Found. Trends Theor. Comput. Sci.* 10 (2014), pp. 1-157.

(WZZ15) S. Wang, Z. Zhang, and T. Zhang. ``Improved Analyses of the Randomized Power Method and Block Lanczos Method''. In: *ArXiv* (2015), pp. 1-22.