# A Tour of the Lanczos Algorithm and its Convergence Guarantees through the Decades

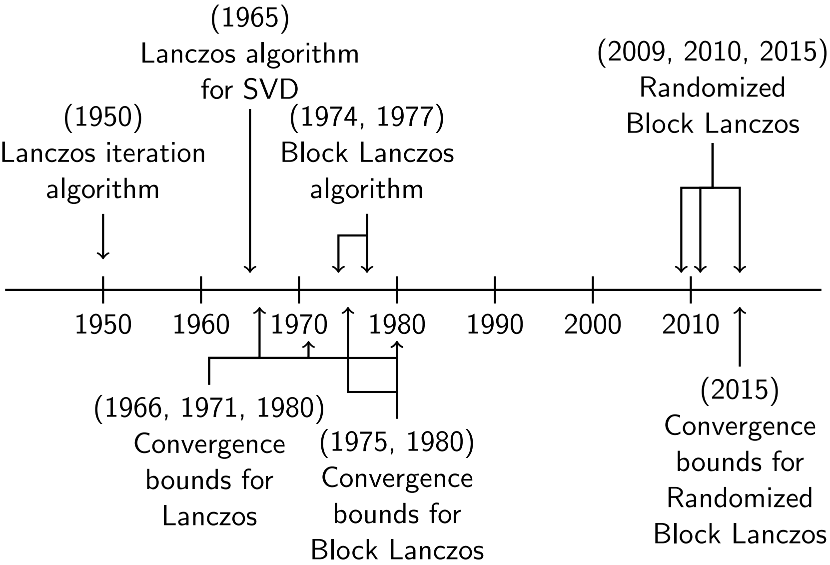

Qiaochu Yuan

Department of Mathematics
UC Berkeley

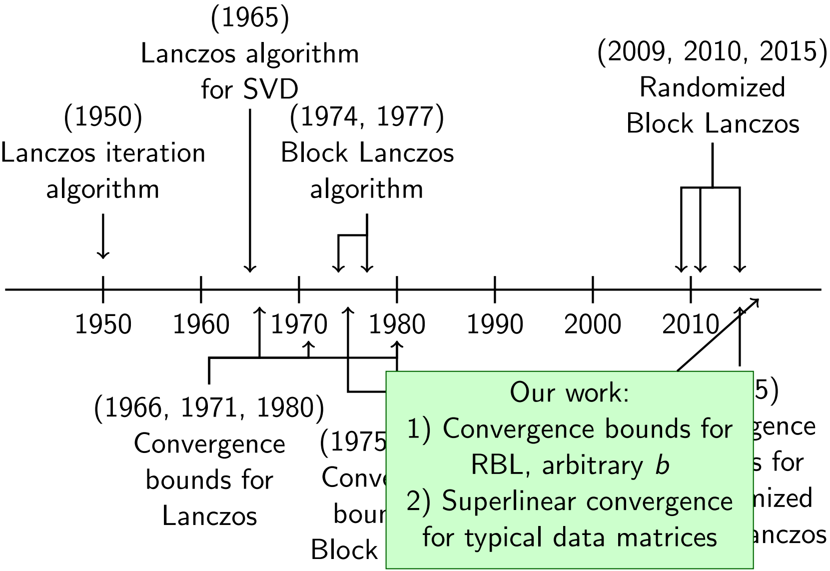
*Joint work with Prof. Ming Gu, Bo Li*


April 17, 2018

# Timeline

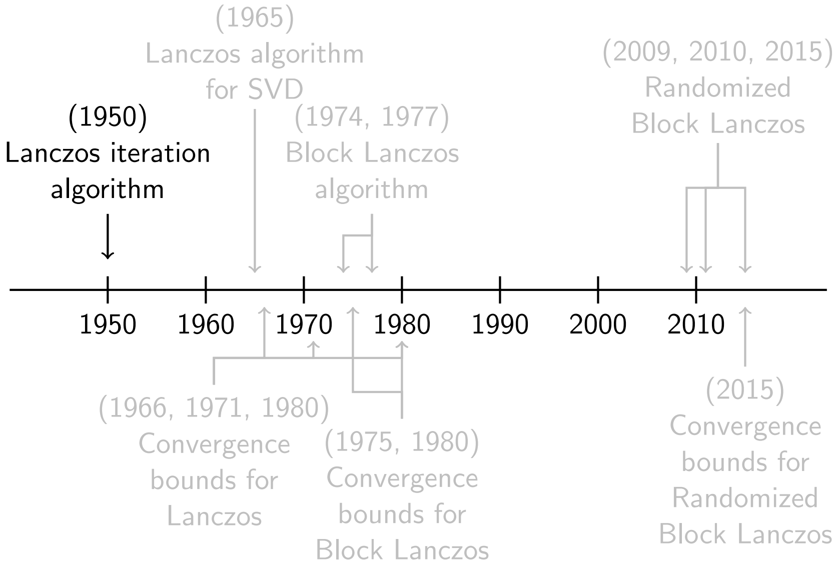

(1965)
Lanczos algorithm for SVD
(2009, 2010, 2015)
Randomized Block Lanczos
(1950)
Lanczos iteration algorithm
(1974, 1977)
Block Lanczos algorithm
1950
1960
1970
1980
1990
2000
2010
(1966, 1971, 1980)
Convergence bounds for Lanczos
(1975, 1980)
Convergence bounds for Block Lanczos
(2015)
Convergence bounds for Randomized Block Lanczos

# Timeline



(1965)
Lanczos algorithm
for SVD
(1950)
Lanczos iteration
algorithm
(1974, 1977)
Block Lanczos
algorithm
(2009, 2010, 2015)
Randomized
Block Lanczos
1950
1960
1970
1980
1990
2000
2010
(1966, 1971, 1980)
Convergence
bounds for
Lanczos
(1975
Conv
bour
Block
5)
gence
s for
nized
anczos
Our work:
1) Convergence bounds for
RBL, arbitrary $b$
2) Superlinear convergence
for typical data matrices

# Timeline


(1965)
Lanczos algorithm
for SVD
(2009, 2010, 2015)
Randomized
Block Lanczos
(1950)
Lanczos iteration
algorithm
(1974, 1977)
Block Lanczos
algorithm
1950
1960
1970
1980
1990
2000
2010
(1966, 1971, 1980)
Convergence
bounds for
Lanczos
(1975, 1980)
Convergence
bounds for
Block Lanczos
(2015)
Convergence
bounds for
Randomized
Block Lanczos

# Lanczos Iteration Algorithm

Developed by Lanczos in 1950 [Lan50].

Widely used iterative algorithm for computing the extremal eigenvalues and corresponding eigenvectors of a large, sparse, symmetric matrix $\mathbf{A}$.

## Goal

Given a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, with eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_n$ and associated eigenvectors $\mathbf{u}_1, \cdots, \mathbf{u}_n$, want to find approximations for

- $\lambda_i$, $i = 1, \cdots, k$, the $k$ largest eigenvalues of $\mathbf{A}$
- $\mathbf{u}_i$, $i = 1, \cdots, k$, the associated eigenvectors

where $k \ll n$.

# Lanczos - Details

General idea:

1. Select an initial vector $\mathbf{v}$.
2. Construct Krylov subspace $\mathcal{K}(\mathbf{A}, \mathbf{v}, k) = \mathrm{span}\{\mathbf{v}, \mathbf{A}\mathbf{v}, \mathbf{A}^2\mathbf{v}, \cdots, \mathbf{A}^{k-1}\mathbf{v}\}$.
3. Restrict and project $\mathbf{A}$ to the Krylov subspace, $\mathbf{T} = \mathrm{proj}_{\mathcal{K}} \mathbf{A}|_{\mathcal{K}}$
4. Use eigen values and vectors of $\mathbf{T}$ as approximations to those of $\mathbf{A}$.

In matrices:

$$\mathbf{K}_k = \begin{bmatrix} \mathbf{v} & \mathbf{A}\mathbf{v} & \cdots & \mathbf{A}^{k-1}\mathbf{v} \end{bmatrix} \in \mathbb{R}^{n \times k}$$

$$\mathbf{Q}_k = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_k \end{bmatrix} \leftarrow \mathrm{qr}(\mathbf{K}_k)$$

$$\mathbf{T}_k = \mathbf{Q}_k^T \mathbf{A} \mathbf{Q}_k \in \mathbb{R}^{k \times k}$$

# Lanczos - Details

$$\mathbf{A}\begin{bmatrix}\mathbf{q}_1 & \cdots & \mathbf{q}_j\end{bmatrix} = \left[\begin{array}{ccc|c}\mathbf{q}_1 & \cdots & \mathbf{q}_j & \mathbf{q}_{j+1}\end{array}\right]\left[\begin{array}{cccc}\alpha_1 & \beta_1 & & \\ \beta_1 & \ddots & \ddots & \\ & \ddots & \ddots & \beta_{j-1} \\ & & \beta_{j-1} & \alpha_j \\ \hline & & & \beta_j\end{array}\right]$$

At each step $j = 1, \cdots, k$ of Lanczos iteration:

$$\mathbf{A}\mathbf{Q}_j = \mathbf{Q}_j\mathbf{T}_j + \beta_j \mathbf{q}_{j+1}\mathbf{e}_{j+1}^T$$

Use the three-term recurrence:

$$\mathbf{A}\mathbf{q}_j = \beta_{j-1}\mathbf{q}_{j-1} + \alpha_j\mathbf{q}_j + \beta_j\mathbf{q}_{j+1}$$

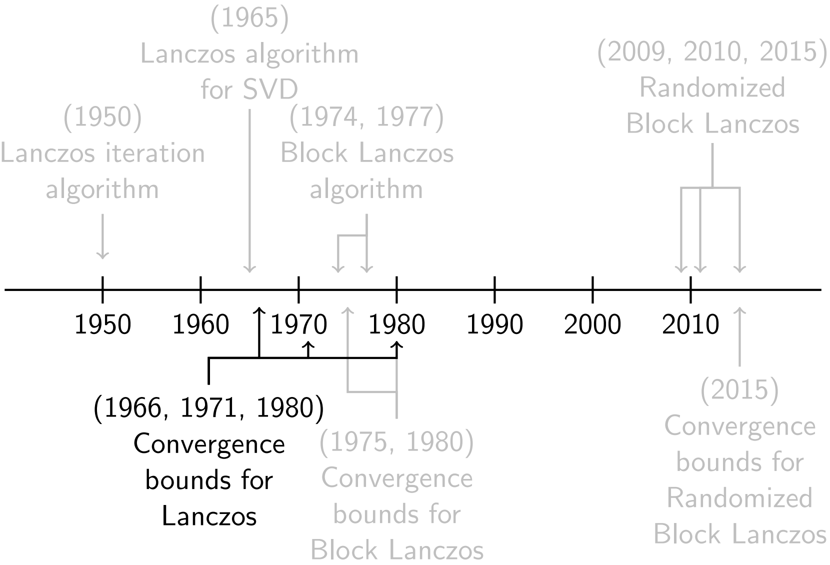
Calculate the $\alpha$s, $\beta$s as:

$$\alpha_j = \mathbf{q}_j^T\mathbf{A}\mathbf{q}_j \tag{1}$$

$$\mathbf{r}_j = (\mathbf{A} - \alpha_j\mathbf{I})\,\mathbf{q}_j - \beta_{j-1}\mathbf{q}_j \tag{2}$$

$$\beta_j = \|\mathbf{r}_j\|_2,\ \mathbf{q}_{j+1} = \mathbf{r}_j/\beta_j \tag{3}$$

# Timeline


(1965)
Lanczos algorithm for SVD
(1950)
Lanczos iteration algorithm
(1974, 1977)
Block Lanczos algorithm
(2009, 2010, 2015)
Randomized Block Lanczos
1950
1960
1970
1980
1990
2000
2010
(1966, 1971, 1980)
Convergence bounds for Lanczos
(1975, 1980)
Convergence bounds for Block Lanczos
(2015)
Convergence bounds for Randomized Block Lanczos

# Convergence of Lanczos

How well does $\lambda_i^{(k)}$, the eigenvalues of $\mathbf{T}_k$, approximate $\lambda_i$, the eigenvalues of $\mathbf{A}$, for $i = 1, \cdots, k$?

First answered by Kaniel in 1966 [Kan66] and Paige in 1971 [Pai71].

**Theorem (Kaniel-Paige Inequality)**

*If* $\mathbf{v}$ *is chosen to be not orthogonal to the eigenspace associated with* $\lambda_1$*, then*

$$0 \leq \lambda_1 - \lambda_1^{(k)} \leq (\lambda_1 - \lambda_n) \frac{\tan^2 \theta(\mathbf{u}_1, \mathbf{v})}{T_{k-1}^2(\gamma_1)} \tag{4}$$

*where* $T_i(x)$ *is the Chebyshev polynomial of degree* $i$*,* $\theta(\cdot, \cdot)$ *is the angle between two vectors, and*

$$\gamma_1 = 1 + 2\frac{\lambda_1 - \lambda_2}{\lambda_2 - \lambda_n}$$

# Convergence of Lanczos

Later generalized by Saad in 1980 [Saa80].

## Theorem (Saad Inequality)

*For* $i = 1, \cdots, k$, *if* $\mathbf{v}$ *is chosen such that* $\mathbf{u}_i^T \mathbf{v} \neq 0$, *then*

$$0 \leq \lambda_i - \lambda_i^{(k)} \leq (\lambda_i - \lambda_n) \left( \frac{L_i^{(k)} \tan \theta (\mathbf{u}_i, \mathbf{v})}{T_{k-i}(\gamma_i)} \right)^2 \tag{5}$$

*where* $T_i(x)$ *is the Chebyshev polynomial of degree* $i$, $\theta(\cdot, \cdot)$ *is the angle between two vectors, and*

$$\gamma_i = 1 + 2 \frac{\lambda_i - \lambda_{i+1}}{\lambda_{i+1} - \lambda_n}$$

$$L_i^{(k)} = \begin{cases} \prod_{j=1}^{i-1} \frac{\lambda_j^{(k)} - \lambda_n}{\lambda_j^{(k)} - \lambda_i} & \textit{if } i \neq 1 \\ 1 & \textit{if } i = 1 \end{cases}$$

# Aside - Chebyshev Polynomials

Recall

$$T_j(x) = \frac{1}{2}\left(\left(x + \sqrt{x^2 - 1}\right)^j + \left(x - \sqrt{x^2 - 1}\right)^j\right) \quad (6)$$

When $j$ is large,

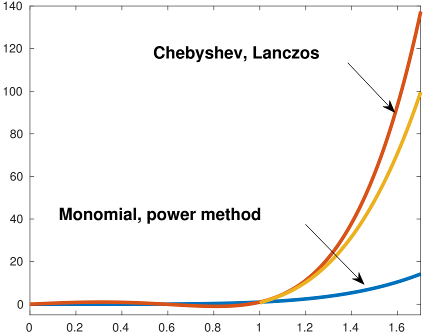
$$T_j(x) \approx \frac{1}{2}\left(x + \sqrt{x^2 - 1}\right)^j$$

and when $g$ is small,

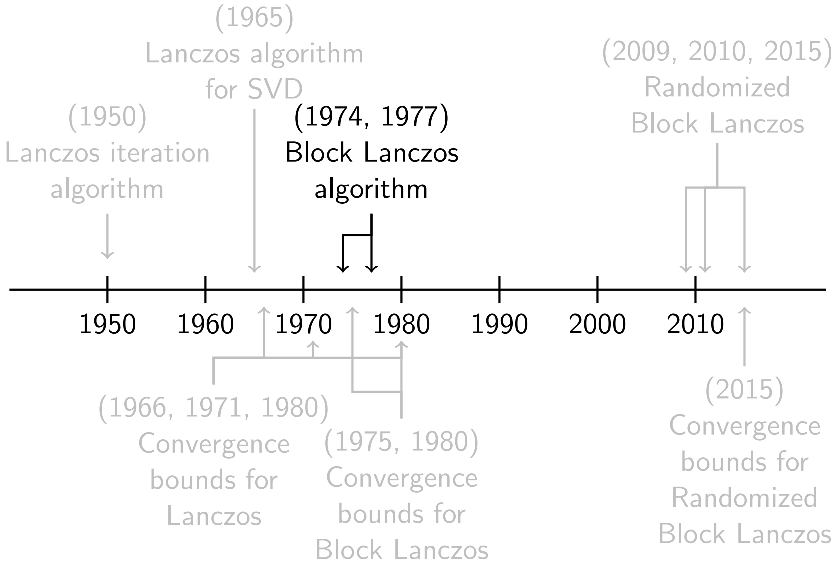
$$T_j(1+g) \approx \frac{1}{2}\left(1 + g + \sqrt{2g}\right)^j$$

determines the convergence of Lanczos with

$$g = \Theta\left(\frac{\lambda_i - \lambda_{i+1}}{\lambda_{i+1} - \lambda_n}\right)$$



# Timeline



(1965)
Lanczos algorithm
for SVD
(2009, 2010, 2015)
Randomized
Block Lanczos
(1950)
Lanczos iteration
algorithm
(1974, 1977)
Block Lanczos
algorithm
1950
1960
1970
1980
1990
2000
2010
(1966, 1971, 1980)
Convergence
bounds for
Lanczos
(1975, 1980)
Convergence
bounds for
Block Lanczos
(2015)
Convergence
bounds for
Randomized
Block Lanczos

# Block Lanczos Algorithm

Introduced by Golub and Underwood in 1977 [GU77] and Cullum and Donath in 1974 [CD74].

The block generalization of the Lanczos method uses, instead of a single initial vector $\mathbf{v}$, a block of $b$ vectors $\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_b \end{bmatrix}$, and builds the Krylov subspace in $q$ iterations as $\mathcal{K}_q(\mathbf{A}, \mathbf{V}, q) = \mathrm{span}\left\{\mathbf{V}, \mathbf{AV}, \cdots, \mathbf{A}^{q-1}\mathbf{V}\right\}$.

$k \leq b$, $bq \ll n$.

Compared to classical Lanczos, block Lanczos

- is more memory and cache efficient, using BLAS3 operations.
- has the ability to converge to eigenvalues with cluster size $> 1$.
- has faster convergence with respect to number of iterations.

# Block Lanczos Algorithm - Details

$$\mathbf{A}\begin{bmatrix}\mathbf{Q}_1 & \cdots & \mathbf{Q}_j\end{bmatrix} = \left[\begin{array}{ccc|c}\mathbf{Q}_1 & \cdots & \mathbf{Q}_j & \mathbf{Q}_{j+1}\end{array}\right]\left[\begin{array}{cccc}\mathcal{A}_1 & \mathcal{B}_1^T & & \\ \mathcal{B}_1 & \ddots & \ddots & \\ & \ddots & \ddots & \mathcal{B}_{j-1}^T \\ & & \mathcal{B}_{j-1} & \mathcal{A}_j \\ \hline & & & \mathcal{B}_j\end{array}\right]$$

At each step $j = 1, \cdots, k$ of Lanczos iteration:

$$\mathbf{AQ} = \mathbf{QT}_j + \mathbf{Q}_{j+1}\begin{bmatrix}\mathbf{0} & \cdots & \mathbf{0} & \mathcal{B}_j\end{bmatrix}$$

Use the three-term recurrence:

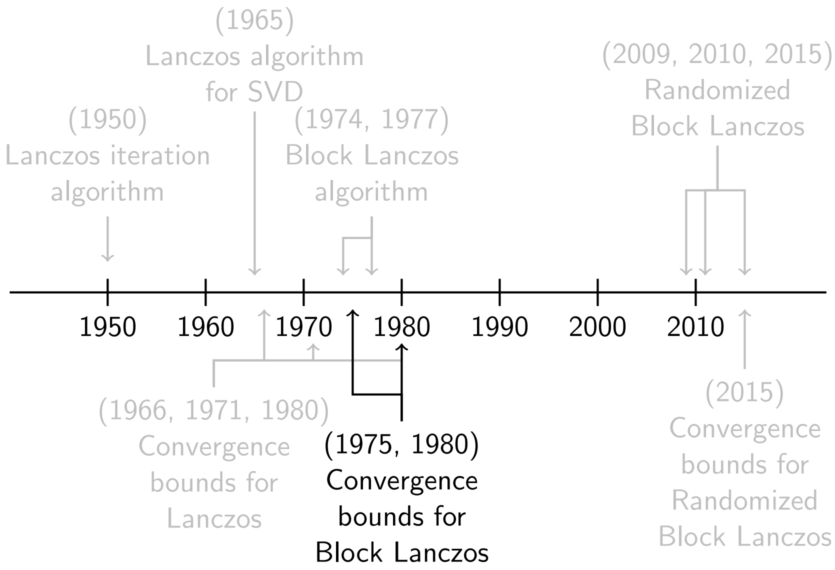
$$\mathbf{AQ}_j = \mathbf{Q}_{j-1}\mathcal{B}_{j-1}^T + \mathbf{Q}_j\mathcal{A}_j + \mathbf{Q}_{j+1}\mathcal{B}_j$$

Calculate the $\mathcal{A}$s, $\mathcal{B}$s as:

$$\mathcal{A}_j = \mathbf{Q}_j^T\mathbf{AQ}_j \tag{7}$$

$$\mathbf{Q}_{j+1}\mathcal{B}_j \leftarrow \mathrm{qr}\left(\mathbf{AQ}_j - \mathbf{Q}_j\mathcal{A}_j - \mathbf{Q}_j\mathcal{B}_{j-1}^T\right) \tag{8}$$

# Timeline


(1965)
Lanczos algorithm for SVD
(2009, 2010, 2015)
Randomized Block Lanczos
(1950)
Lanczos iteration algorithm
(1974, 1977)
Block Lanczos algorithm
1950
1960
1970
1980
1990
2000
2010
(1966, 1971, 1980)
Convergence bounds for Lanczos
(1975, 1980)
Convergence bounds for Block Lanczos
(2015)
Convergence bounds for Randomized Block Lanczos

# Convergence of Block Lanczos

First analyzed by Underwood in 1975 [Und75] with the introduction of the algorithm.

## Theorem (Underwood Inequality)

*Let* $\lambda_i$, $\mathbf{u}_i$, $i = 1, \cdots, n$ *be the eigenvalues and eigenvectors of* $\mathbf{A}$ *respectively. Let* $\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_b \end{bmatrix}$. *If* $\mathbf{U}^T\mathbf{V}$ *is of full rank* $b$, *then for* $i = 1, \cdots, b$

$$0 \leq \lambda_i - \lambda_i^{(q)} \leq (\lambda_1 - \lambda_n) \frac{\tan^2 \Theta(\mathbf{U}, \mathbf{V})}{T_{q-1}^2(\rho_i)} \tag{9}$$

*where* $T_i(x)$ *is the Chebyshev polynomial of degree* $i$, $\cos \Theta(\mathbf{U}, \mathbf{V}) = \sigma_{\min}\left(\mathbf{U}^T\mathbf{V}\right)$, *and*

$$\rho_i = 1 + 2\frac{\lambda_i - \lambda_{b+1}}{\lambda_{b+1} - \lambda_n}$$

# Convergence of Block Lanczos

## Theorem (Saad Inequality [Saa80])

*Let* $\lambda_j$, $\mathbf{u}_j$, $j = 1, \cdots, n$ *be the eigenvalues and eigenvectors of* $\mathbf{A}$ *respectively. Let* $\mathbf{U}_i = \begin{bmatrix} \mathbf{u}_i & \cdots & \mathbf{u}_{i+b-1} \end{bmatrix}$. *For* $i = 1, \cdots, b$, *if* $\mathbf{U}_i^T \mathbf{V}$ *is of full rank* $b$, *then*

$$0 \leq \lambda_i - \lambda_i^{(q)} \leq (\lambda_i - \lambda_n) \left( \frac{L_i^{(q)} \tan \Theta \left( \mathbf{U}_i, \mathbf{V} \right)}{T_{q-i} \left( \hat{\gamma}_i \right)} \right)^2 \tag{10}$$

*where* $T_i(x)$ *is the Chebyshev polynomial of degree* $i$, $\cos \Theta \left( \mathbf{U}, \mathbf{V} \right) = \sigma_{\min} \left( \mathbf{U}^T \mathbf{V} \right)$, *and*

$$\hat{\gamma}_i = 1 + 2 \frac{\lambda_i - \lambda_{i+b}}{\lambda_{i+b} - \lambda_n}$$

$$L_i^{(q)} = \begin{cases} \prod_{j=1}^{i-1} \frac{\lambda_j^{(q)} - \lambda_n}{\lambda_j^{(q)} - \lambda_i} & \textit{if } i \neq 1 \\ 1 & \textit{if } i = 1 \end{cases}$$

# Aside - Block Size

Recall, when $j$ is large and $g$ is small

$$T_j(1+g) \approx \frac{1}{2}\left(1+g+\sqrt{2g}\right)^j \tag{11}$$

$$\text{classical: } g = \Theta\left(\frac{\lambda_i - \lambda_{i+1}}{\lambda_{i+1} - \lambda_n}\right)$$

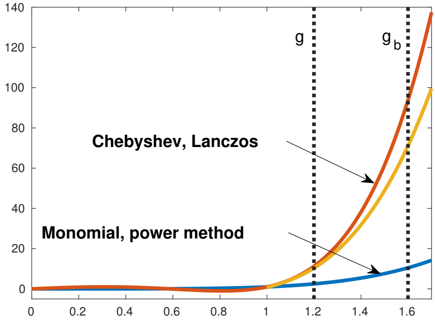
$$\text{block: } g_b = \Theta\left(\frac{\lambda_i - \lambda_{i+b}}{\lambda_{i+b} - \lambda_n}\right)$$

Suppose eigenvalue distributed as $\lambda_j > (1+\epsilon)\lambda_{j+1}$ for all $j$:

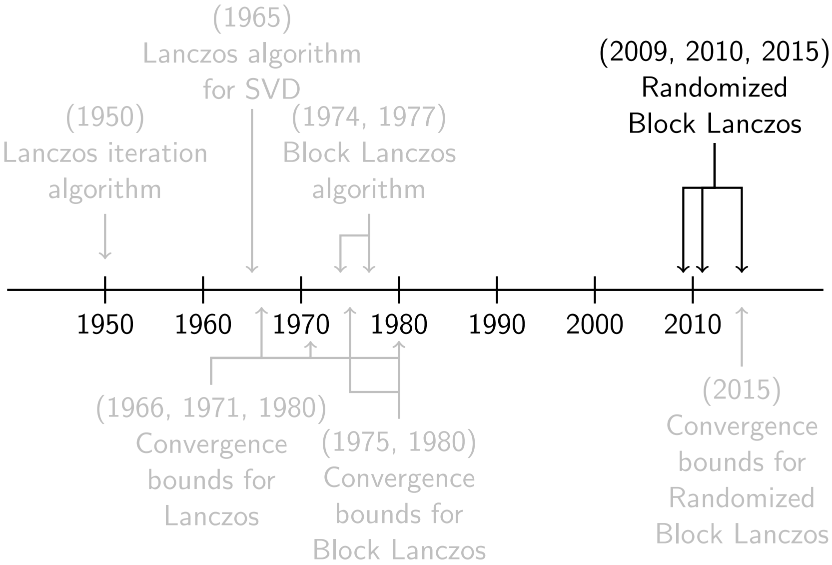
$$g_b \approx \frac{1-(1+\epsilon)^{-b}}{(1+\epsilon)^{-b}} \cdot \lambda_i$$

$$\approx b \cdot \lambda_i \epsilon$$

$$\approx b \cdot g$$



# Timeline


(1965)
Lanczos algorithm for SVD
(2009, 2010, 2015)
Randomized Block Lanczos
(1950)
Lanczos iteration algorithm
(1974, 1977)
Block Lanczos algorithm
1950
1960
1970
1980
1990
2000
2010
(1966, 1971, 1980)
Convergence bounds for Lanczos
(1975, 1980)
Convergence bounds for Block Lanczos
(2015)
Convergence bounds for Randomized Block Lanczos

# Randomized Block Lanczos

In recent years, there has been increased interest in algorithms to compute low-rank approximations of matrices, with

- high computational efficiency requirement, running on large matrices
- low approximation accuracy requirement, 2-3 digits of accuracy

Applications mostly in big-data computations

- compression of data matrices
- matrix processing techniques, e.g. PCA
- optimization of the nuclear norm objective function

# Randomized Block Lanczos

Grew out of work done in randomized algorithms, in particular Randomized Subspace Iteration, by

- Rokhlin, Szlam, and Tygert in 2009 [RST09]
- Halko, Martinsson, and Tropp in 2011 [HMST11]
- Gu [Gu15], and Musco [MM15] in 2015

Idea: Instead of taking any initial set of vectors $\mathbf{V}$, an unfortunate choice of which could result in poor convergence, choose $\mathbf{V} = \mathbf{A\Omega}$, a random projection of the columns of $\mathbf{A}$, to better capture the range space.

# RSI & RBL Pseudocode

**Algorithm 1** Randomized Subspace Iteration (RSI) pseudocode

**Input:** $\mathbf{A} \in \mathbb{R}^{m \times n}$, target rank $k$, block size $b$, number of iter. $q$

**Output:** $\mathbf{B}_k \in \mathbb{R}^{m \times n}$ , a rank-$k$ approximation

1: Draw $\mathbf{\Omega} \in \mathbb{R}^{n \times b}$. $\omega_{ij} \sim \mathcal{N}(0, 1)$.

2: Form $\mathbf{K} = (\mathbf{A}\mathbf{A}^T)^{q-1}\mathbf{A}\mathbf{\Omega}$.

3: Orthogonalize $\mathrm{qr}(\mathbf{K}) \to \mathbf{Q}$.

4: $\mathbf{B}_k = (\mathbf{Q}\mathbf{Q}^T\mathbf{A})_k = \mathbf{Q}(\mathbf{Q}^T\mathbf{A})_k$.

$*$ $(\mathbf{M})_k$ indicates the $k$-truncated SVD of matrix $\mathbf{M}$.

**Algorithm 2** Randomized Block Lanczos (RBL) pseudocode

**Input:** $\mathbf{A} \in \mathbb{R}^{m \times n}$, target rank $k$, block size $b$, number of iter. $q$

**Output:** $\mathbf{B}_k \in \mathbb{R}^{m \times n}$, a rank-$k$ approximation

1: Draw $\mathbf{\Omega} \in \mathbb{R}^{n \times b}$. $\omega_{ij} \sim \mathcal{N}(0, 1)$.

2: Form $\mathbf{K} = [\mathbf{A}\mathbf{\Omega}, (\mathbf{A}\mathbf{A}^T)\mathbf{A}\mathbf{\Omega}, \cdots, (\mathbf{A}\mathbf{A}^T)^{q-1}\mathbf{A}\mathbf{\Omega}]$.

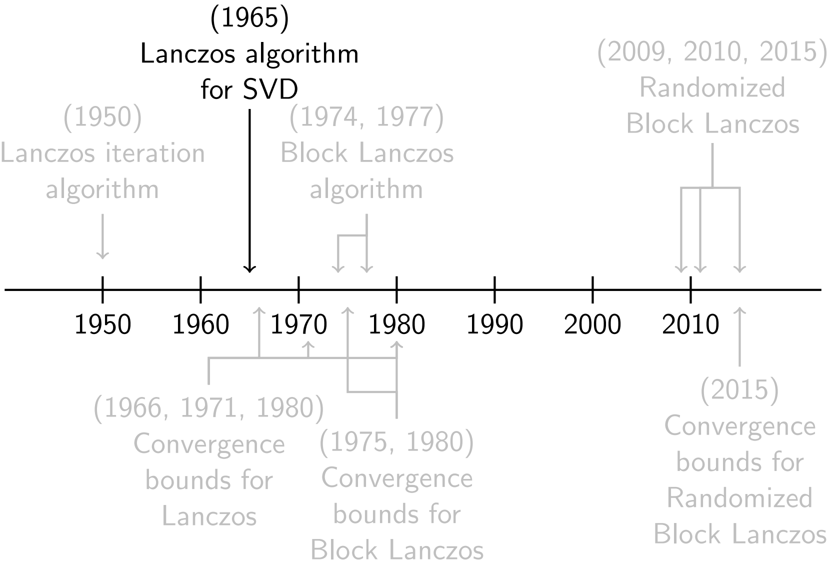
3: Orthogonalize $\mathrm{qr}(\mathbf{K}) \to \mathbf{Q}$.

4: $\mathbf{B}_k = (\mathbf{Q}\mathbf{Q}^T\mathbf{A})_k = \mathbf{Q}(\mathbf{Q}^T\mathbf{A})_k$.

$*$ requires $bq \geq k$.

$**$ numerically unstable.

# Timeline



(1965)
Lanczos algorithm for SVD
(1950)
Lanczos iteration algorithm
(1974, 1977)
Block Lanczos algorithm
(2009, 2010, 2015)
Randomized Block Lanczos
1950
1960
1970
1980
1990
2000
2010
(1966, 1971, 1980)
Convergence bounds for Lanczos
(1975, 1980)
Convergence bounds for Block Lanczos
(2015)
Convergence bounds for Randomized Block Lanczos

# Lanczos for SVD

A modification of the Lanczos algorithm can be used to find the extremal singular value pairs of a non-symmetric matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$.

Utilizing the connections between the singular value decomposition of $\mathbf{A}$ and the eigen decompositions of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$,

1. Select a block of initial vectors $\mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_b \end{bmatrix}$.
2. Construct Krylov subspaces $\mathcal{K}_V\left(\mathbf{A}^T\mathbf{A}, \mathbf{V}, q\right)$ and $\mathcal{K}_U\left(\mathbf{A}\mathbf{A}^T, \mathbf{A}\mathbf{V}, q\right)$.
3. Restrict and project $\mathbf{A}$ to form $\mathbf{B} = \mathrm{proj}_{\mathcal{K}_U}\mathbf{A}|_{\mathcal{K}_V}$
4. Use singular values and vectors of $\mathbf{B}$ as approximations to those of $\mathbf{A}$.

# Golub-Kahan Bidiagonalization [GK65]

$$\mathbf{A}\begin{bmatrix}\mathbf{V}_1 & \cdots & \mathbf{V}_j\end{bmatrix} = \begin{bmatrix}\mathbf{U}_1 & \cdots & \mathbf{U}_j\end{bmatrix}\begin{bmatrix}\mathcal{A}_1 & \mathcal{B}_1^T & & \\ & \ddots & \ddots & \\ & & \ddots & \mathcal{B}_{j-1}^T \\ & & & \mathcal{A}_j\end{bmatrix}$$

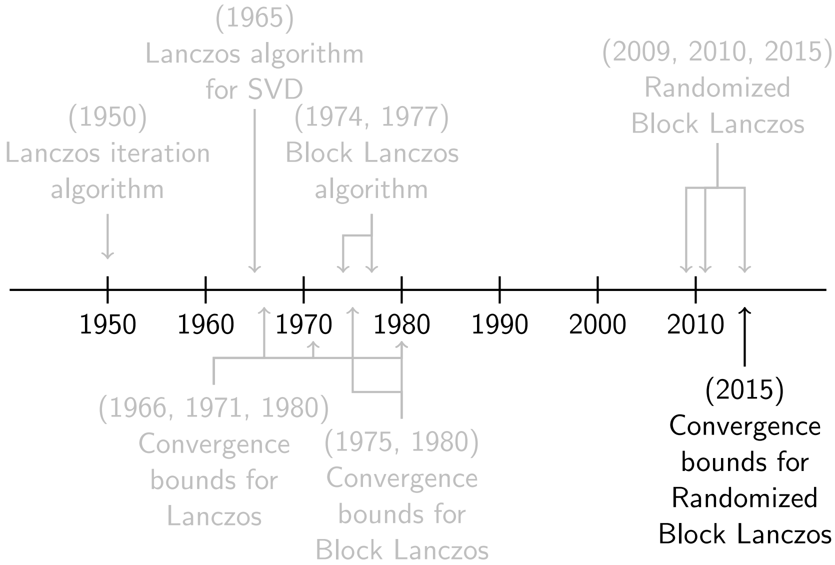
$$\mathbf{A}^T\begin{bmatrix}\mathbf{U}_1 & \cdots & \mathbf{U}_j\end{bmatrix} = \begin{bmatrix}\mathbf{V}_1 & \cdots & \mathbf{V}_j \,|\, \mathbf{V}_{j+1}\end{bmatrix}\begin{bmatrix}\mathcal{A}_1 & & & \\ \mathcal{B}_1 & \ddots & & \\ & \ddots & \ddots & \\ & & \mathcal{B}_{j-1} & \mathcal{A}_j \\ \hline & & & \mathcal{B}_j\end{bmatrix}$$

At each step $j = 1, \cdots, k$ of the bidiagonalization, calculate $\mathcal{A}$s and $\mathcal{B}$s as:

$$\mathbf{U}_j\mathcal{A}_j \leftarrow \mathrm{qr}\left(\mathbf{A}\mathbf{V}_j - \mathbf{U}_{j-1}\mathcal{B}_{j-1}^T\right) \tag{12}$$

$$\mathbf{V}_{j+1}\mathcal{B}_j \leftarrow \mathrm{qr}\left(\mathbf{A}^T\mathbf{U}_j - \mathbf{V}_j\mathcal{A}_j\right) \tag{13}$$

# Timeline


(1965)
Lanczos algorithm for SVD
(1950)
Lanczos iteration algorithm
(1974, 1977)
Block Lanczos algorithm
(2009, 2010, 2015)
Randomized Block Lanczos
1950
1960
1970
1980
1990
2000
2010
(1966, 1971, 1980)
Convergence bounds for Lanczos
(1975, 1980)
Convergence bounds for Block Lanczos
(2015)
Convergence bounds for Randomized Block Lanczos

# Convergence of RBL

Analyzed by Musco and Musco in 2015 [MM15].

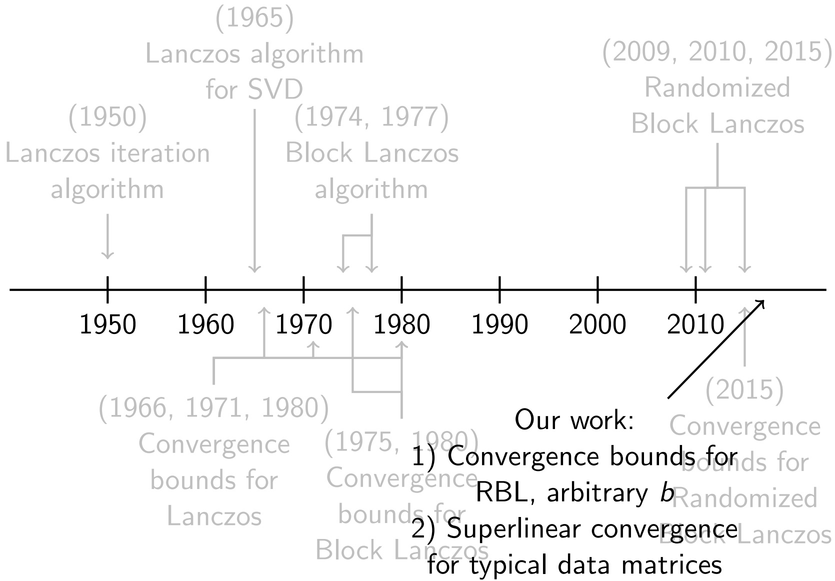
**Theorem**

*For block size* $b \geq k$, *in* $q = \Theta\left(\frac{\log n}{\epsilon}\right)$ *iterations of RSI and* $q = \Theta\left(\frac{\log n}{\sqrt{\epsilon}}\right)$ *iterations of RBL, with constant probability* 99/100, *the following inequalities bounds are satisfied:*

$$|\sigma_i^2 - \sigma_i^2(\mathbf{B}_k)| \leq \epsilon\sigma_{k+1}^2, \ , \textit{for } i = 1, \cdots, k \tag{14}$$

*In the event that* $\sigma_{b+1} \leq c\sigma_k$ *with* $c < 1$, *taking* $q = \Theta\left(\frac{\log(n/\epsilon)}{\min(1, \sigma_k/\sigma_{b+1} - 1)}\right)$ *and* $q = \Theta\left(\frac{\log(n/\epsilon)}{\sqrt{\min(1, \sigma_k/\sigma_{b+1} - 1)}}\right)$ *suffices for RSI and RBL respectively.*

# Timeline



(1965)
Lanczos algorithm
for SVD
(1950)
Lanczos iteration
algorithm
(1974, 1977)
Block Lanczos
algorithm
(2009, 2010, 2015)
Randomized
Block Lanczos
1950
1960
1970
1980
1990
2000
2010
(1966, 1971, 1980)
Convergence
bounds for
Lanczos
(1975, 1980)
Convergence
bounds for
Block Lanczos
(2015)
Convergence
bounds for
Randomized
Block Lanczos
Our work:
1) Convergence bounds for
RBL, arbitrary *b*
2) Superlinear convergence
for typical data matrices

# Aside: Convergence of RSI

Many of our analysis techniques are similar to those used by Gu to analyze RSI [Gu15]. The bounds for RBL resulting from the current work will be similar in form to the bounds for RSI.

**Theorem (RSI convergence)**

*Let $\mathbf{B}_k$ be the approximation returned by the RSI algorithm on $\mathbf{A} = \mathbf{U\Sigma V}^T$, with target rank $k$, block size $b \geq k$, and $q$ iterations. If $\hat{\mathbf{\Omega}}_1$ has full row rank in $\mathbf{V}^T\mathbf{\Omega} = \begin{bmatrix} \hat{\mathbf{\Omega}}_1^T & \hat{\mathbf{\Omega}}_2^T \end{bmatrix}^T$, then*

$$\sigma_j \geq \sigma_j(\mathbf{B}_k) \geq \frac{\sigma_j}{\sqrt{1 + \|\hat{\mathbf{\Omega}}_2\|_2^2 \|\hat{\mathbf{\Omega}}_1^\dagger\|_2^2 \left(\frac{\sigma_{b+1}}{\sigma_j}\right)^{4q+2}}} \tag{15}$$

# Current Work - Core Ideas

The core pieces of our analysis are:

1. the growth behavior of Chebyshev polynomials,
2. the choice of a clever orthonormal basis for Krylov subspace,
3. the creation of a spectrum "gap", by separating the spectrum of $\mathbf{A}$ into those singular values that are "close" to $\sigma_k$, and those that are sufficiently smaller in magnitude.

## Setup & Notation

For block size $b$, random Gaussian matrix $\mathbf{\Omega} \in \mathbb{R}^{n \times b}$, we are interested in the column span of

$$\mathbf{K}_q = \left[\mathbf{A\Omega} \quad \left(\mathbf{AA}^T\right)\mathbf{A\Omega} \quad \cdots \quad \left(\mathbf{AA}^T\right)^{q-1}\mathbf{A\Omega}\right]$$

Let the SVD of $\mathbf{A} \in \mathbb{R}^{m \times n}$ be $\mathbf{U\Sigma V}^T$. For any $0 \leq p \leq q$, define

$$\begin{aligned}\hat{\mathbf{K}}_p :&= \mathbf{U}T_{2p+1}(\mathbf{\Sigma})\left[\hat{\mathbf{\Omega}} \quad \mathbf{\Sigma}^2\hat{\mathbf{\Omega}} \quad \cdots \quad \mathbf{\Sigma}^{2(q-p-1)}\hat{\mathbf{\Omega}}\right] \\ &= \mathbf{U}T_{2p+1}(\mathbf{\Sigma})\mathbf{V}_{q-p}\end{aligned}$$

Note $\mathbf{V}_{q-p} \in \mathbb{R}^{n \times b(q-p)}$. We require $b(q-p) \geq k$, the target rank.

**Lemma**

*With $\hat{\mathbf{K}}_p$ and $\mathbf{K}_q$ as previously defined,*

$$\operatorname{span}\{\mathbf{K}_q\} \supseteq \operatorname{span}\left\{\hat{\mathbf{K}}_p\right\} \tag{16}$$

# Convergence of RBL

**Theorem (Main Result)**

*Let $\mathbf{B}_k$ be the approximation returned by the RBL algorithm. With notation as previously defined, if the random starting $\mathbf{\Omega}$ is initialized such that the block Vandermonde formed by the sub-blocks of $\mathbf{V}_{q-p}$ is invertible, then for all $1 \leq j \leq k$, and all choices[a] of $s, r$,*

$$\sigma_j \geq \sigma_j(\mathbf{B}_k) \geq \frac{\sigma_{j+s}}{\sqrt{1 + \mathcal{C}^2 T_{2p+1}^{-2} \left(1 + 2 \cdot \frac{\sigma_j - \sigma_{j+s+r+1}}{\sigma_{j+s+r+1}}\right)}} \tag{17}$$

*where $p = q - \frac{k+r}{b}$, and $\mathcal{C}$ is a constant independent of $q$.*

[a] $s$ is chosen to be non-zero to handle multiple singular values, and can be set to zero otherwise.

# Summary of Convergence Bounds

Rewriting bounds in comparable forms:

| citation | bound | req. on b |
|---|---|---|
| Saad [Saa80] | $\lambda_j^{(q)} \geq \frac{\lambda_j}{1+L_j^{(q)^2}\tan^2\Theta(\mathbf{U},\mathbf{V})\,T_{q-j}^{-2}\left(1+2\frac{\lambda_j-\lambda_{j+b}}{\lambda_{j+b}}\right)}$ | $b \geq k$ |
| Musco [MM15] spec. indep. | $\sigma_j^{(q)} \geq \frac{\sigma_j}{\sqrt{1+\mathcal{C}_1^2\log^2(n)\,q^{-2}\frac{\sigma_{k+1}^2}{\sigma_j^2}}}$ | $b \geq k$ |
| Musco [MM15] spec. dep. | $\sigma_j^{(q)} \geq \frac{\sigma_j}{\sqrt{1+\mathcal{C}_2 n e^{-q\sqrt{\min(1,\sigma_k/\sigma_{b+1}-1)}}\frac{\sigma_{k+1}^2}{\sigma_j^2}}}$ | $b \geq k$ |
| Current Work | $\sigma_j^{(q)} \geq \frac{\sigma_j}{\sqrt{1+\mathcal{C}_3^2\,T_{2q+1-2(k+r)/b}^{-2}\left(1+2\frac{\sigma_j-\sigma_{j+r+1}}{\sigma_{j+r+1}}\right)}}$ | $b \geq 1$<br>$bq \geq k+r$ |

# Typical Case - Superlinear Convergence

Recall: $\{a_q\}$ convergence *superlinearly* to $a$ if

$$\lim_{q\to\infty} \frac{|a_{q+1} - a|}{|a_q - a|} = 0 \tag{18}$$

- In practice, Lanczos algorithms (classical, block, randomized) often exhibit superlinear convergence behavior.
- It has been shown that classical Lanczos iteration is theoretically superlinearly convergent under certain assumptions about the singular spectrum [saa94, Li10].
- We show this for block Lanczos algorithms, i.e., that under certain assumptions about the singular spectrum, block Lanczos produces rank $k$ approximations $\mathbf{B}_k$ such that $\sigma_j(\mathbf{B}_k) \to \sigma_j$ superlinearly.

# Typical Case - Superlinear Convergence

A typical data matrix might have singular value spectrum decaying to 0, i.e., $\sigma_j \to 0$. In this case our bound suggests that convergence is governed by

$$a_q := \left(\mathcal{C}(r) T_p^{-1}(1+g)\right)^2 \approx \left(\mathcal{C}(r) \cdot \frac{1}{2}\left(1+g+\sqrt{2g}\right)^{-p}\right)^2 \qquad \to 0$$

with

$$g = 2\frac{\sigma_j - \sigma_{j+r+1}}{\sigma_{j+r+1}} = 2\left(\frac{\sigma_j}{\sigma_{j+r+1}} - 1\right) \to \infty$$

$$p = 2\left(q - \frac{k+r}{b}\right) + 1 = 2q + \left(1 - 2\frac{k+r}{b}\right)$$

We argue that $a_{q+1}/a_q \to 0$ as follows: for all $\epsilon > 0$, choose[1] $r$ so that $1 + g \geq \epsilon^{-\frac{1}{2}}$. Then,

$$\frac{a_{q+1}}{a_q} \leq \epsilon$$

[1] Recall 1) our main result holds for all $r$; 2) $k + r = (q-p)b$, and so choosing $r$ amounts to choosing $q$.

# Effect of $r$

There is some optimal value of $r$, typically non-zero, which achieves the best convergence factor. The balance is between larger (smaller) values of $r$, which implies lower (higher) Chebyshev degree but bigger (smaller) gap.

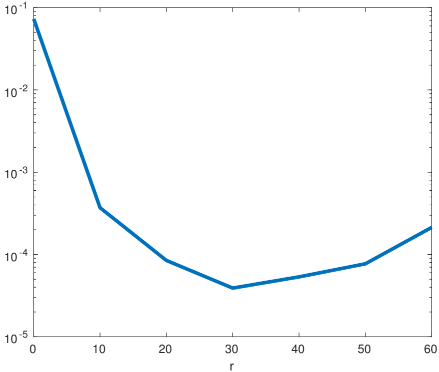
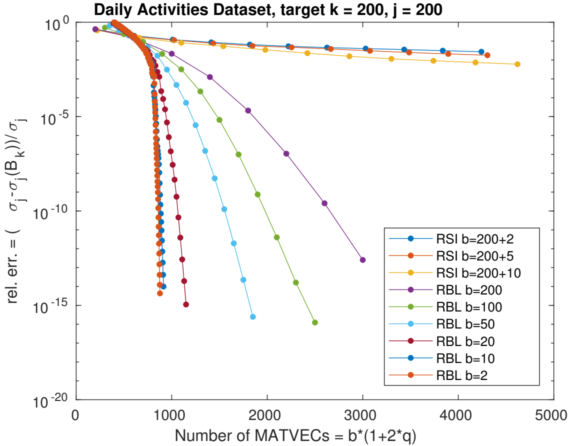



Figure: Value of reciprocal convergence factor $T^{-1}_{2q+1-2((k+r)/b)}\left(1+2\frac{\sigma_j-\sigma_{j+r+1}}{\sigma_{j+r+1}}\right)$ as $r$ varies, for Daily Activities and Sports Dataset, $k=j=100$, $b=10$, $q=20$.

# Numerical Example

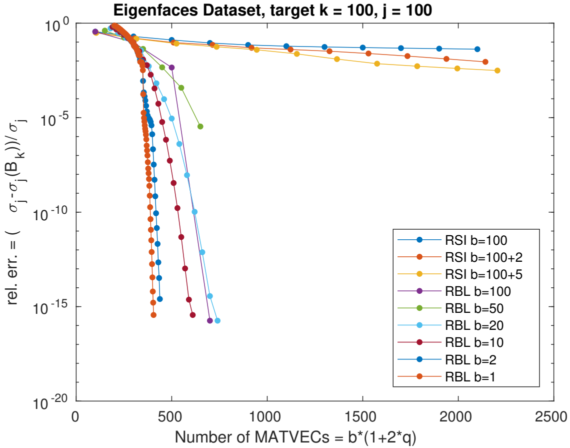
Experimentally, choices of smaller block sizes $1 \leq b < k$ appear favorable with superlinear convergence for all block sizes.

Figure: Daily Activities and Sports Dataset - $\mathbf{A} \in \mathbb{R}^{9120 \times 5625}$.



# Numerical Example

Figure: Eigenfaces Dataset - $\mathbf{A} \in \mathbb{R}^{10304 \times 400}$.



# Conclusions

- Both the theoretical analysis and numerical evidence suggest that, holding the number of matrix vector operations constant, RBL with smaller block size $b$ is better.
- For matrices with decaying spectrum, RBL achieves superlinear convergence.
- However the preference for smaller $b$ must be balanced with the advantages of a larger $b$ for computational efficiency and numerical stability reasons in a practical implementation, and should be further investigated.

# References I


- Jane Cullum and William E Donath, *A block lanczos algorithm for computing the q algebraically largest eigenvalues and a corresponding eigenspace of large, sparse, real symmetric matrices*, Decision and Control including the 13th Symposium on Adaptive Processes, 1974 IEEE Conference on, vol. 13, IEEE, 1974, pp. 505–509.
- Gene Golub and William Kahan, *Calculating the singular values and pseudo-inverse of a matrix*, Journal of the Society for Industrial and Applied Mathematics, Series B: Numerical Analysis **2** (1965), no. 2, 205–224.
- Gene Howard Golub and Richard Underwood, *The block lanczos method for computing eigenvalues*, Mathematical software, Elsevier, 1977, pp. 361–377.

# References II


- Ming Gu, *Subspace iteration randomization and singular value problems*, SIAM Journal on Scientific Computing **37** (2015), no. 3, A1139–A1173.
- Nathan Halko, Per-Gunnar Martinsson, Yoel Shkolnisky, and Mark Tygert, *An algorithm for the principal component analysis of large data sets*, SIAM Journal on Scientific computing **33** (2011), no. 5, 2580–2594.
- Shmuel Kaniel, *Estimates for some computational techniques in linear algebra*, Mathematics of Computation **20** (1966), no. 95, 369–378.
- Cornelius Lanczos, *An iteration method for the solution of the eigenvalue problem of linear differential and integral operators*, United States Governm. Press Office Los Angeles, CA, 1950.

# References III


- Ren-Cang Li, *Sharpness in rates of convergence for the symmetric lanczos method*, Mathematics of Computation **79** (2010), no. 269, 419–435.
- Cameron Musco and Christopher Musco, *Randomized block krylov methods for stronger and faster approximate singular value decomposition*, Advances in Neural Information Processing Systems, 2015, pp. 1396–1404.
- Christopher Conway Paige, *The computation of eigenvalues and eigenvectors of very large sparse matrices.*, Ph.D. thesis, University of London, 1971.
- Vladimir Rokhlin, Arthur Szlam, and Mark Tygert, *A randomized algorithm for principal component analysis*, SIAM Journal on Matrix Analysis and Applications **31** (2009), no. 3, 1100–1124.

# References IV


Yousef Saad, *On the rates of convergence of the lanczos and the block-lanczos methods*, SIAM Journal on Numerical Analysis **17** (1980), no. 5, 687–706.

*Theoretical error bounds and general analysis of a few lanczos-type algorithms*, 1994.

Richard Underwood, *An iterative block lanczos method for the solution of large sparse symmetric eigenproblems*, Tech. report, 1975.