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

(1950) Lanczos iteration algorithm
(1965) Lanczos algorithm for SVD
(1974, 1977) Block Lanczos algorithm
(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
Timeline

- (1950) Lanczos iteration algorithm
- (1965) Lanczos algorithm for SVD
- (1974, 1977) Block Lanczos algorithm
- (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 $A$.

**Goal**

Given a symmetric matrix $A \in \mathbb{R}^{n \times n}$, with eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_n$ and associated eigenvectors $u_1, \cdots, u_n$, want to find approximations for

- $\lambda_i$, $i = 1, \cdots, k$, the $k$ largest eigenvalues of $A$
- $u_i$, $i = 1, \cdots, k$, the associated eigenvectors

where $k \ll n$. 
Lanczos - Details

General idea:

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

In matrices:

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

$$Q_k = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_k \end{bmatrix} \leftarrow \text{qr}(K_k)$$

$$T_k = Q_k^T A Q_k \in \mathbb{R}^{k \times k}$$
Lanczos - Details

\[ A [q_1 \cdots q_j] = [ q_1 \cdots q_j | q_{j+1} ] \]

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

\[ AQ_j = Q_j T_j + \beta_j q_{j+1} e_{j+1}^T \]

Use the three-term recurrence:

\[ Aq_j = \beta_{j-1} q_{j-1} + \alpha_j q_j + \beta_j q_{j+1} \]

Calculate the \( \alpha \)s, \( \beta \)s as:

\[ \alpha_j = q_j^T Aq_j \]  \hspace{1cm} (1)

\[ r_j = (A - \alpha_j I) q_j - \beta_{j-1} q_j \]  \hspace{1cm} (2)

\[ \beta_j = \|r_j\|_2, \quad q_{j+1} = r_j / \beta_j \]  \hspace{1cm} (3)
Timeline

(1950) Lanczos iteration algorithm

(1965) Lanczos algorithm for SVD

(1974, 1977) Block Lanczos algorithm


(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 \( T_k \), approximate \( \lambda_i \), the eigenvalues of \( A \), for \( i = 1, \ldots, k \)?


**Theorem (Kaniel-Paige Inequality)**

*If \( v \) is chosen to be not orthogonal to the eigenspace associated with \( \lambda_1 \), then*

\[
0 \leq \lambda_1 - \lambda_1^{(k)} \leq (\lambda_1 - \lambda_n) \frac{\tan^2 \theta (u_1, v)}{T_{k-1}^2 (\gamma_1)}
\]

*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

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

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

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

$$L_i^{(k)} = \begin{cases} 1 & \text{if } i = 1 \\ \prod_{j=1}^{i-1} \frac{\lambda_j^{(k)} - \lambda_n}{\lambda_j^{(k)} - \lambda_i} & \text{if } i \neq 1 \end{cases}$$
Recall
\[ T_j(x) = \frac{1}{2} \left( (x + \sqrt{x^2 - 1})^j + (x - \sqrt{x^2 - 1})^j \right) \] (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

- (1950) Lanczos iteration algorithm
- (1965) Lanczos algorithm for SVD
- (1974, 1977) Block Lanczos algorithm
- (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 $v$, a block of $b$ vectors $V = [v_1 \cdots v_b]$, and builds the Krylov subspace in $q$ iterations as $\mathcal{K}_q (A, V, q) = \text{span} \{ V, AV, \cdots, A^{q-1}V \}$.

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

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

\[
A \begin{bmatrix} Q_1 & \cdots & Q_j \end{bmatrix} = \begin{bmatrix} Q_1 & \cdots & Q_j & Q_{j+1} \end{bmatrix}
\]

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

\[
AQ = QT_j + Q_{j+1} \begin{bmatrix} 0 & \cdots & 0 & B_j \end{bmatrix}
\]

Use the three-term recurrence:

\[
AQ_j = Q_{j-1}B_{j-1}^T + Q_jA_j + Q_{j+1}B_j
\]

Calculate the \( A \)s, \( B \)s as:

\[
A_j = Q_j^T AQ_j \quad (7)
\]

\[
Q_{j+1}B_j \leftarrow \text{qr} \left( AQ_j - Q_jA_j - Q_jB_{j-1}^T \right) \quad (8)
\]
Timeline

(1950) Lanczos iteration algorithm
(1965) Lanczos algorithm for SVD
(1974, 1977) Block Lanczos algorithm
(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 $\mathbf{U} = [\mathbf{u}_1 \cdots \mathbf{u}_b]$. If $\mathbf{U}^T \mathbf{V}$ is of full rank $b$, then for $i = 1, \cdots, b$

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

where $T_i(x)$ is the Chebyshev polynomial of degree $i$, $\cos \Theta (\mathbf{U}, \mathbf{V}) = \sigma_{\min} (\mathbf{U}^T \mathbf{V})$, 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 $\lambda_j, u_j, j = 1, \cdots, n$ be the eigenvalues and eigenvectors of $A$ respectively. Let $U_i = [u_i \cdots u_{i+b-1}]$. For $i = 1, \cdots, b$, if $U_i^T V$ is of full rank $b$, then

$$0 \leq \lambda_i - \lambda_i^{(q)} \leq (\lambda_i - \lambda_n) \left( \frac{L_i^{(q)} \tan \Theta(U_i, V)}{T_{q-i}(\gamma_i)} \right)^2$$  \hspace{1cm} (10)

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

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

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

Recall, when \( j \) is large and \( g \) is small

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

(11)

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

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

(1950) Lanczos iteration algorithm

(1965) Lanczos algorithm for SVD


(1974, 1977) Block Lanczos algorithm

(1975, 1980) Convergence bounds for Block Lanczos


(2015) Convergence bounds for 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 $V$, an unfortunate choice of which could result in poor convergence, choose $V = A\Omega$, a random projection of the columns of $A$, to better capture the range space.
Algorithm 1 Randomized Subspace Iteration (RSI) pseudocode

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

**Output:** \( B_k \in \mathbb{R}^{m \times n} \), a rank-\( k \) approximation

1: Draw \( \Omega \in \mathbb{R}^{n \times b} \). \( \omega_{ij} \sim \mathcal{N}(0, 1) \).
2: Form \( K = (AA^T)^{q-1}A\Omega \).
3: Orthogonalize \( qr(K) \to Q \).
4: \( B_k = (QQ^TA)_k = Q(Q^TA)_k \).

\( (M)_k \) indicates the \( k \)-truncated SVD of matrix \( M \).

Algorithm 2 Randomized Block Lanczos (RBL) pseudocode

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

**Output:** \( B_k \in \mathbb{R}^{m \times n} \), a rank-\( k \) approximation

1: Draw \( \Omega \in \mathbb{R}^{n \times b} \). \( \omega_{ij} \sim \mathcal{N}(0, 1) \).
2: Form 
   \[ K = [A\Omega, (AA^T)A\Omega, \ldots, (AA^T)^{q-1}A\Omega] \]
3: Orthogonalize \( qr(K) \to Q \).
4: \( B_k = (QQ^TA)_k = Q(Q^TA)_k \).

\* requires \( bq \geq k \).
\** numerically unstable.
Timeline

(1950) Lanczos iteration algorithm
(1950)

(1965) Lanczos algorithm for SVD
(1965)

(1974, 1977) Block Lanczos algorithm
(1974, 1977)


(1975, 1980) Convergence bounds for Block Lanczos
(1975, 1980)

(2015) Convergence bounds for Randomized Block Lanczos
(2015)
A modification of the Lanczos algorithm can be used to find the extremal singular value pairs of a non-symmetric matrix $A \in \mathbb{R}^{m \times n}$.

Utilizing the connections between the singular value decomposition of $A$ and the eigen decompositions of $AA^T$ and $A^TA$,

1. Select a block of initial vectors $V = [v_1 \cdots v_b]$.
2. Construct Krylov subspaces $K_V (A^T A, V, q)$ and $K_U (AA^T, AV, q)$.
3. Restrict and project $A$ to form $B = \text{proj}_{K_U} A|_{K_V}$
4. Use singular values and vectors of $B$ as approximations to those of $A$. 
Golub-Kahan Bidiagonalization [GK65]

\[
\begin{bmatrix} V_1 & \cdots & V_j \end{bmatrix} = \begin{bmatrix} U_1 & \cdots & U_j \end{bmatrix} \begin{bmatrix} A_1 & B_1^T \\ \vdots & \ddots & \vdots \\ \vdots & & B_{j-1}^T \\ A_j & & \end{bmatrix}
\]

\[
\begin{bmatrix} \end{bmatrix} = \begin{bmatrix} U_1 & \cdots & U_j \end{bmatrix} \begin{bmatrix} A_1 & B_1 \\ \vdots & \ddots & \vdots \\ \vdots & & B_{j-1} \\ B_j & A_j & \end{bmatrix}
\]

At each step \( j = 1, \cdots, k \) of the bidiagonalization, calculate \( A_j \) and \( B_j \) as:

\[
U_j A_j \leftarrow \text{qr} \left( A V_j - U_{j-1} B_{j-1}^T \right)
\]

\[
V_{j+1} B_j \leftarrow \text{qr} \left( A^T U_j - V_j A_j \right)
\]
Timeline

- (1950) Lanczos iteration algorithm
- (1965) Lanczos algorithm for SVD
- (1974, 1977) Block Lanczos algorithm
- (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:

$$|\sigma_i^2 - \sigma_i^2(B_k)| \leq \epsilon \sigma_{k+1}^2, \text{ for } i = 1, \cdots, k$$

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.
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 $B_k$ be the approximation returned by the RSI algorithm on $A = U\Sigma V^T$, with target rank $k$, block size $b \geq k$, and $q$ iterations. If $\hat{\Omega}_1$ has full row rank in $V^T\Omega = [\hat{\Omega}_1^T \; \hat{\Omega}_2^T]^T$, then

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

(15)
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 $A$ into those singular values that are “close” to $\sigma_k$, and those that are sufficiently smaller in magnitude.
For block size $b$, random Gaussian matrix $\Omega \in \mathbb{R}^{n \times b}$, we are interested in the column span of

$$K_q = \begin{bmatrix} A\Omega & (AA^T)A\Omega & \cdots & (AA^T)^{q-1}A\Omega \end{bmatrix}$$

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

$$\hat{K}_p := UT_{2p+1}(\Sigma) \begin{bmatrix} \hat{\Omega} & \Sigma^2\hat{\Omega} & \cdots & \Sigma^{2(q-p-1)}\hat{\Omega} \end{bmatrix} = UT_{2p+1}(\Sigma)V_{q-p}$$

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

**Lemma**

With $\hat{K}_p$ and $K_q$ as previously defined,

$$\text{span} \{K_q\} \supseteq \text{span} \{\hat{K}_p\} \quad (16)$$
Convergence of RBL

Let $B_k$ be the approximation returned by the RBL algorithm. With notation as previously defined, if the random starting $\Omega$ is initialized such that the block Vandermonde formed by the sub-blocks of $V_{q-p}$ is invertible, then for all $1 \leq j \leq k$, and all choices\(^a\) of $s, r,
\[
\sigma_j \geq \sigma_j(B_k) \geq \frac{\sigma_{j+s}}{\sqrt{1 + C^2 T_{2p+1}^{-2} \left(1 + 2 \cdot \frac{\sigma_j - \sigma_{j+s+r+1}}{\sigma_{j+s+r+1}}\right)}}
\]

where $p = q - \frac{k+r}{b}$, and $C$ 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.
Rewriting bounds in comparable forms:

<table>
<thead>
<tr>
<th>citation</th>
<th>bound</th>
<th>req. on b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saad [Saa80]</td>
<td>$\lambda_j^{(q)} \geq \frac{\lambda_j}{1+L_j^{(q)}^2 \tan^2 \Theta(U,V) T_{q-j}^{-2} \left(1+2 \frac{\lambda_j - \lambda_{j+b}}{\lambda_{j+b}}\right)}$</td>
<td>$b \geq k$</td>
</tr>
<tr>
<td>Musco [MM15] spec. indep.</td>
<td>$\sigma_j^{(q)} \geq \frac{\sigma_j}{\sqrt{1+C_1^2 \log^2(n) q^{-2} \frac{\sigma_{k+1}}{\sigma_j^2}}}$</td>
<td>$b \geq k$</td>
</tr>
<tr>
<td>Musco [MM15] spec. dep.</td>
<td>$\sigma_j^{(q)} \geq \frac{\sigma_j}{\sqrt{1+C_2 ne^{-q\sqrt{\min(1,\sigma_k/\sigma_{b+1}-1)}} \frac{\sigma_{k+1}}{\sigma_j^2}}}$</td>
<td>$b \geq k$</td>
</tr>
<tr>
<td>Current Work</td>
<td>$\sigma_j^{(q)} \geq \frac{\sigma_j}{\sqrt{1+C_3^2 T_{2q+1-2(k+r)}/b \left(1+2 \frac{\sigma_{j+r+1}}{\sigma_{j+r+1}}\right)}}$</td>
<td>$b \geq 1$ $bq \geq k + r$</td>
</tr>
</tbody>
</table>
Recall: \( \{a_q\} \) convergence *superlinearly* to \( a \) if

\[
\lim_{q \to \infty} \frac{|a_{q+1} - a|}{|a_q - a|} = 0
\]  

(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 \( B_k \) such that \( \sigma_j(B_k) \to \sigma_j \) superlinearly.
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

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

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 $T_{2q+1}^{-1} - 2((k+r)/b) \left( 1 + 2 \frac{\sigma_j - \sigma_{j+r+1}}{\sigma_{j+r+1}} \right)$ as $r$ varies, for Daily Activities and Sports Dataset, $k = j = 100$, $b = 10$, $q = 20$. 
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 - $\mathbf{A} \in \mathbb{R}^{9120 \times 5625}$. 

![Graph showing the relationship between the number of MATVECs and relative error for different block sizes and RSI and RBL variants.]
Figure: Eigenfaces Dataset - \( \mathbf{A} \in \mathbb{R}^{10304 \times 400} \).
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.


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