New Convergence Estimates for Block Lanczos Methods for the Truncated SVD

Ming Gu, Christopher Melgaard, Christopher A. Wong

University of California, Berkeley

14 May 2014

Lanczos Bidiagonalization

Algorithm

Let $A \in \mathbb{R}^{m \times n}$ and pick initial matrix $Y \in \mathbb{R}^{n \times b}$. Set $U_0 = R_1 = 0$. Then an approximate rank k truncated SVD of A, given by $\tilde{A}_k$, is computed by

$[V_1, \sim] \leftarrow \mathrm{qr}(Y, 0)$

for $j = 1, \ldots, s$ **do**

$\quad [U_j, C_j] \leftarrow \mathrm{qr}(AV_j - U_{j-1}{R_j}^*, 0)$

$\quad [V_{j+1}, R_{j+1}] \leftarrow \mathrm{qr}(A^* U_j - V_j {C_j}^*, 0)$

end for

$$\tilde{B} \leftarrow \begin{bmatrix} C_1 & {R_2}^* & \\ & C_2 & \ddots \\ & & \ddots \end{bmatrix} \in \mathbb{R}^{sb \times sb}, \; [\tilde{U}, \tilde{\Sigma}, \tilde{V}] \leftarrow \mathrm{svd}(\tilde{B})$$

$$\tilde{A}_k \leftarrow \left(\begin{bmatrix} U_1 & \ldots & U_s \end{bmatrix} \tilde{U} \right) \tilde{\Sigma}_k \left(\begin{bmatrix} V_1 & \ldots & V_s \end{bmatrix} \tilde{V} \right)^*$$

Ideally $sb \ll \min(m, n)$ and hence we avoid doing the exact SVD on a very large matrix.

Some important properties

Let $U^{(s)} = \begin{bmatrix} U_1 & \dots & U_s \end{bmatrix}, V^{(s)} = \begin{bmatrix} V_1 & \dots & V_s \end{bmatrix}$, so $\tilde{A} = U^{(s)} \tilde{B} V^{(s)*}$.

- $AV^{(s)} = U^{(s)} \tilde{B}$
- Krylov subspace property: The columns of $V^{(s)}$ are an orthonormal basis for

$$K_s(A^*A, Y) = \mathrm{Col} \begin{bmatrix} Y & A^*AY & \dots & (A^*A)^{s-1}Y \end{bmatrix}.$$

- If $U\Sigma V^*$ is the true SVD of A, then

$$\|A - \tilde{A}\|_F = \|\Sigma - P^{(s)}\Sigma\|,$$

where $P^{(s)}$ is the orthogonal projection onto $K_s(\Sigma^*\Sigma, V^*Y)$.

Desired error estimates

The main question: How close is $\tilde{A}_k$ to A_k?

- Frobenius norm: Estimate $\delta = \delta(s, b, k, Y)$, where

$$\|A - \tilde{A}_k\|_F^2 = \|A - A_k\|_F^2 + \delta^2.$$

- Singular values: Estimate $\epsilon_j = \epsilon_j(s, b, k, Y)$, where

$$\sigma_j(\tilde{A}_k)^2 = \sigma_j(A)^2(1 - \epsilon_j), \quad 1 \le j \le k.$$

- Singular vectors: Measure the angles between K_s and the true singular vectors of A

Main strategy

Set $\Lambda = \Sigma^*\Sigma, \Omega = V^*Y$.
We need to analyze the Krylov subspace $K_s(\Lambda, \Omega)$. Write down the matrix

$$K = \begin{bmatrix} \Omega & \Lambda\Omega & \dots & \Lambda^{s-1}\Omega \end{bmatrix}.$$

Find a new basis for $K_s(\Lambda, \Omega)$ by choosing matrix R such that

$$KR = \begin{bmatrix} I_{sb} \\ E \end{bmatrix}.$$

This can be done by choosing $R = K^{-1}_{1:sb,1:sb}$, so

$$E = K_{sb+1:n,sb+1:n} K^{-1}_{1:sb,1:sb}.$$

Main strategy

New basis for $K_s(\Lambda, \Omega)$:

$$KR = \begin{bmatrix} I_{sb} \\ E \end{bmatrix}, \quad E = K_{sb+1:n,sb+1:n} K^{-1}_{1:sb,1:sb}.$$

The size of E is related to the tangents of the principal angles between the Krylov subspace and the first sb singular vectors:

$$|\tan \theta_j|^2 = \frac{(E^*(I + EE^*)^{-1}E)_{jj}}{((I + E^*E)^{-1})_{jj}},$$

where θ_j is the angle between K_s and the jth singular vector.

Norm estimates

We want $\delta = \delta(s, b, k, Y)$ where

$$\|A - \tilde{A}_k\|_F^2 = \|A - A_k\|_F^2 + \delta^2.$$

Estimate δ:

$$\begin{aligned} \|A - \tilde{A}_k\|_F^2 &= \|\Sigma - (P^{(s)}\Sigma)_k\|_F^2 \\ &\leq \|\Sigma - P^{(s)}\Sigma_k\|_F^2 \\ &= \|A - A_k\|_F^2 + \|\Sigma_k - P^{(s)}\Sigma_k\|_F^2. \end{aligned}$$

Therefore

$$\delta^2 \leq \|\Sigma_k - P^{(s)}\Sigma_k\|_F^2,$$

and we need to estimate this upper bound.

Norm estimates

We have

$$\delta^2 \leq \|\Sigma_k - P^{(s)}\Sigma_k\|_F^2,$$

Set $P^{(s)} = QQ^*$. Using our "E" representation,

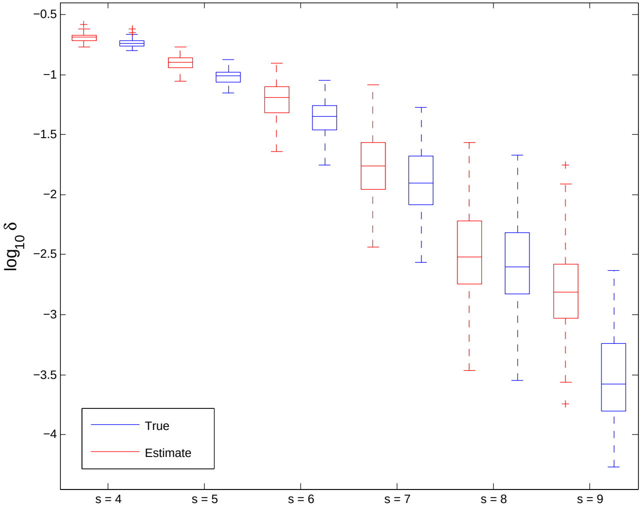
$$P^{(s)} = \begin{bmatrix} I \\ E \end{bmatrix} (I + E^*E)^{-1} \begin{bmatrix} I & E^* \end{bmatrix}.$$

Then, if $E = \begin{bmatrix} E_1 & E_2 \end{bmatrix}$, where E_1 is the first k columns of E, then

$$\|\Sigma_k - P^{(s)}\Sigma_k\|_F^2 = \|(I + E_1E_1^* + E_2E_2^*)^{-1/2}E_1\Sigma_k\|_F^2.$$

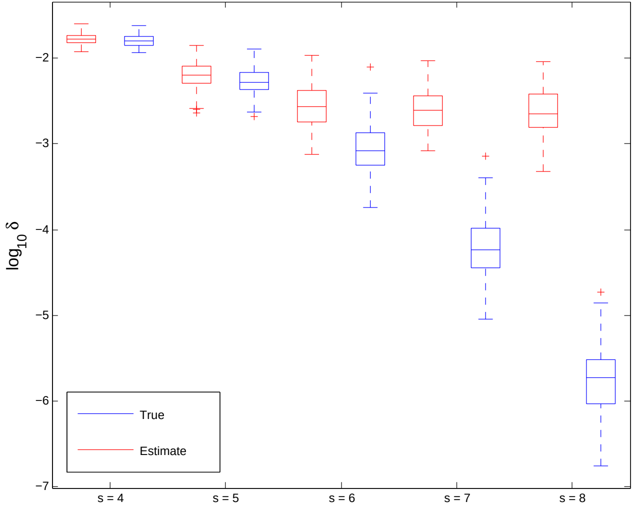
Norm estimates

We calculate the true and estimated δ with $A \in \mathbb{R}^{100\times 100}$ with $\sigma_j = 1/j$, fixing $k = 10$, $b = 3$, $s = 4, \ldots, 9$.



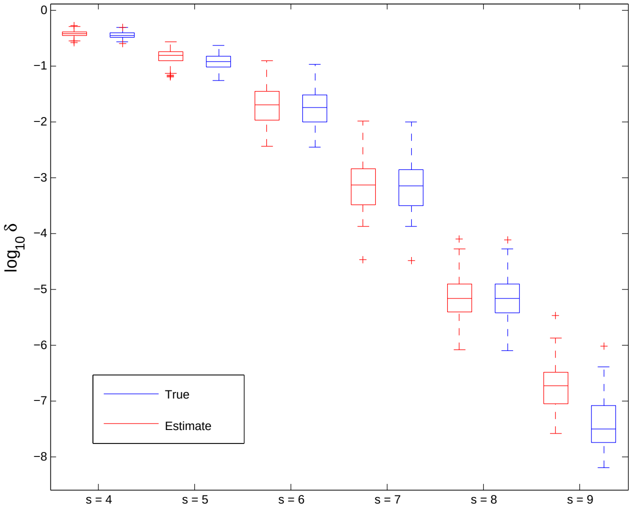
Norm estimates

We calculate the true and estimated δ with $A \in \mathbb{R}^{100\times 100}$ with $\sigma_j = 1/j^2$, fixing $k = 10$, $b = 3$, $s = 4, \ldots, 8$.



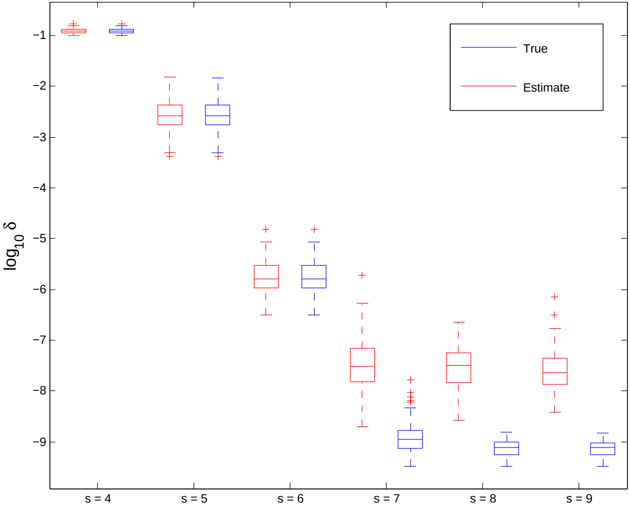
Norm estimates

We calculate the true and estimated δ with $A \in \mathbb{R}^{100\times 100}$ with $\sigma_j = (1.2)^{-j+1}$, fixing $k = 10$, $b = 3$, $s = 4, \ldots, 9$.



Norm estimates

We calculate the true and estimated δ with $A \in \mathbb{R}^{100\times 100}$ with $\sigma_j = 1/j$ for $j \leq 10$, and $\sigma_j = 1/j^2$ for $j > 10$, fixing $k = 10$, $b = 3$, $s = 4, \ldots, 9$.



Singular values

We want to find ϵ_j such that for $1 \leq j \leq k$,

$$\sigma_j(\tilde{A}_k)^2 = \sigma_j(A)^2(1 - \epsilon_j).$$

If $E_{:,1:j}$ indicates the first j columns of E, then

$$\sigma_j(\tilde{A}_k)^2 \geq \frac{\sigma_j(A)^2}{1 + \|E_{:,1:j}\|_2^2},$$

in which case

$$\epsilon_j \leq \frac{\|E_{:,1:j}\|_2^2}{1 + \|E_{:,1:j}\|_2^2}.$$

Therefore if $E_{:,1:j}$ is large, then the singular value estimate is inaccurate, but if it is small then ϵ_j is close to 0.

Singular values

Compare with existing estimate of Saad:
Let $\sigma_j = \sigma_j(A)$, and $\tilde{\sigma}_j = \sigma_j(\tilde{A}_k)$.

$$\epsilon_j \leq \left[\frac{K_j^{(s)} \tan\theta(Y, v_j)}{T_{s-j}(\hat{\gamma}_j)} \right]^2,$$

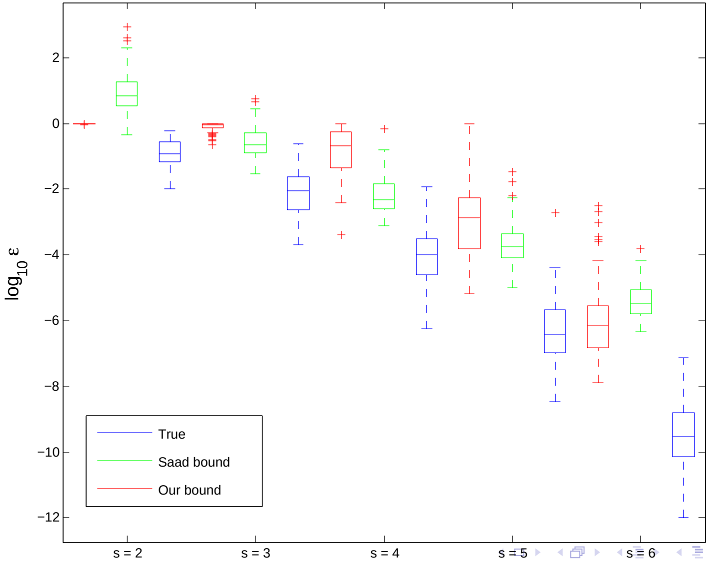
where

$$\hat{\gamma}_j = 1 + 2\frac{\sigma_j^2 - \sigma_{j+b}^2}{\sigma_{j+b}^2}, K_j = \begin{cases} \prod_{i \leq j-1} \frac{\tilde{\sigma}_i^2}{\tilde{\sigma}_i^2 - \sigma_j^2} & j > 1 \\ 1 & j = 1 \end{cases}$$

and $\theta(Y, v_j)$ is the angle between the subspace spanned by the initial block Y and the j-th right singular vector v_j.

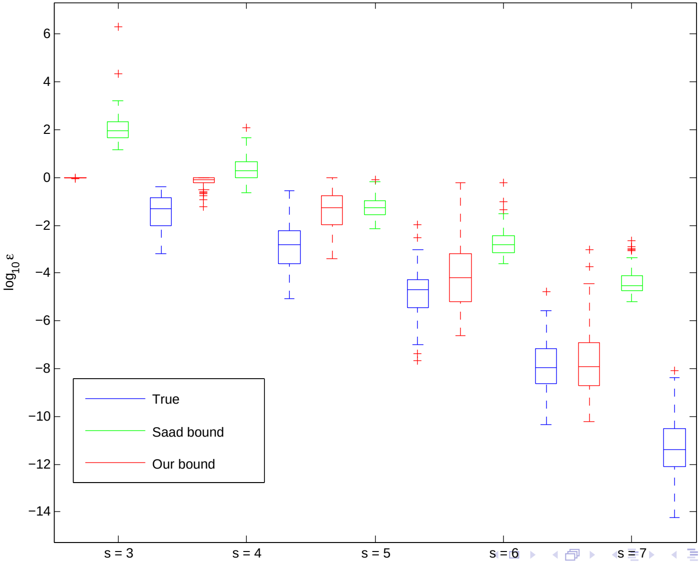
Singular values

Using $A \in \mathbb{R}^{100\times 100}$ with $\sigma_j = (1.2)^{-j+1}$, with $k = 1$, $b = 2$, $s = 2, 3, 4, 5, 6$, we compute estimates for ϵ_1 using our approach and using Saad's estimate.



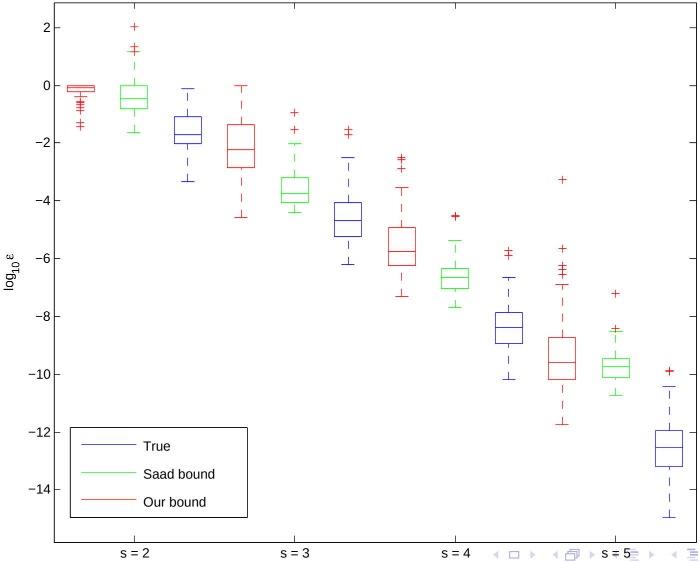
Singular values

Same matrix, with $k = 2$, $b = 2$, $s = 3, 4, 5, 6, 7$, we compute estimates for ϵ_2 using our approach and using Saad's estimate.



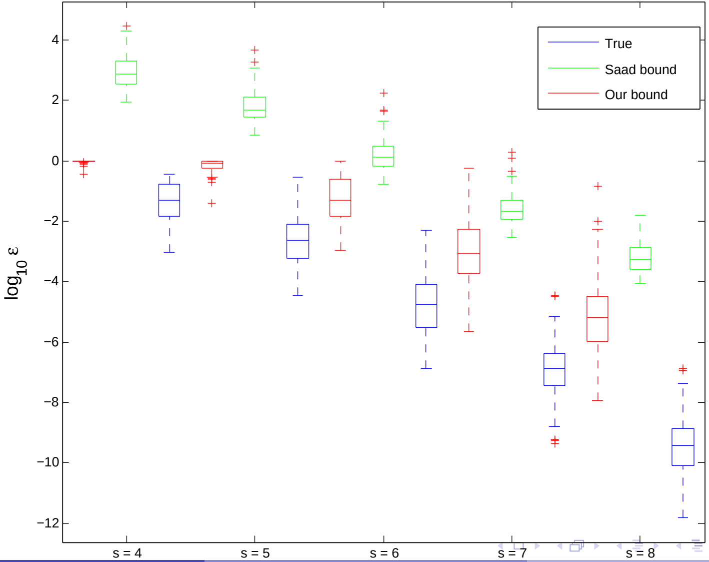
Singular values

Using $A \in \mathbb{R}^{100\times 100}$ with $\sigma_j = 1/j$, with $k = 1$, $b = 2$, $s = 2, 3, 4, 5$, we compute estimates for ϵ_1 using our approach and using Saad's estimate.



Singular values

Same A, $k = 4$, $s = 4, 5, 6, 7, 8$, we compute estimates for ϵ_4 using our approach and using Saad's estimate.



Singular vectors

Recall the estimate

$$\|A - \tilde{A}_k\|_F^2 \leq \|A - A_k\|_F^2 + \|\Sigma_k - P^{(s)}\Sigma_k\|_F^2.$$

We can rephrase this as

$$\|A - \tilde{A}_k\|_F^2 \leq \sum_{j=1}^{k} \sigma_j^2 |\sin\theta_j|^2 + \sum_{j \geq k+1} \sigma_j^2,$$

where θ_j is the angle between $K_s(A^*A, V^*Y)$ and the j-th singular vector v_j. If Q is a matrix of orthonormal basis vectors for the Krylov subspace, then

$$\cos\theta_j = \|Q^* v_j\|_2.$$

Using our matrix representation of Q from before, we obtain

$$|\sin\theta_j|^2 = (E^*(I + EE^*)^{-1}E)_{jj}.$$

Conclusions

- Norm estimates are provided where none existed before.
- Singular value estimates are competitive with previous results.
- Further work needed: Estimating E itself may not be easy; $E = A_1(A_2)^{-1}$,

$$A_i = \begin{bmatrix} \Omega_i & \Lambda_i \Omega_i & \dots & \Lambda_i^{s-1} \Omega_i \end{bmatrix}.$$

and A_1, A_2 are independent random variables.

References

- Gene H. Golub, Franklin T. Luk, and Michael L. Overton, *A block lanczos method for computing the singular values and corresponding singular vectors of a matrix*, ACM Trans. Math. Softw. **7** (1981), no. 2, 149–169.
- Ming Gu, *Subspace iteration randomization and singular value problems.*
- Y. Saad, *On the rates of convergence of the Lanczos and the block-Lanczos methods*, SIAM J. Numer. Anal. **17** (1980), no. 5, 687–706.
- Horst D. Simon and Hongyuan Zha, *Low-rank matrix approximation using the Lanczos bidiagonalization process with applications*, SIAM J. Sci. Comput. **21** (2000), no. 6, 2257–2274 (electronic).