Contrasting properties of RSVD and LSQR algorithms for solutions of ill-posed problems: Approximating the SVD

Rosemary Renaut\textsuperscript{1}  Anthony Helmstetter \textsuperscript{1}  Saeed Vatankhah\textsuperscript{2}

1: School of Mathematical and Statistical Sciences, Arizona State University, renaut@asu.edu, anthony.helmstetter@asu.edu

2: Institute of Geophysics, University of Tehran, svatan@ut.ac.ir

International Conference on Mathematics of Data Science
November 2018
Outline

Background: TSVD surrogate for the small scale
Standard Approaches to Estimate Regularization Problem
Convergence of the regularization parameter for UPRE
Algorithm Verification

Methods for the Large Scale: Approximating the SVD
Krylov: Golub Kahan Bidiagonalization - LSQR
Randomized SVD
Simulations: Hybrid RSVD and Hybrid LSQR

Conclusions: RSVD - LSQR
Main Results
Relevance to Data Science
Simple Ill-Posed Problem: Image Restoration

Mildly ill-posed problem: Slow decay of singular values. SNR 13
Notation: Spectral Decomposition of the Solution: The SVD

Consider general discrete problem

\[ Ax = b, \quad A \in \mathbb{R}^{m \times n}, \quad b \in \mathbb{R}^m, \quad x \in \mathbb{R}^n. \]

Singular value decomposition (SVD) of \( A \) rank \( r \leq \min(m, n) \)

\[ A = U \Sigma V^T = \sum_{i=1}^{r} u_i \sigma_i v_i^T, \quad \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_r). \]

Singular values \( \sigma_i \), singular vectors \( u_i, v_i \), rank \( r \).

Expansion for the solution:

\[ x = \sum_{i=1}^{r} \frac{s_i}{\sigma_i} v_i, \quad s_i = u_i^T b \]
Truncated SVD of size $k$ gives best rank-$k$ approximation to $A$. Surrogate model is given by $A_k \approx U_k \Sigma_k V_k^T$.

Filtered and Truncated solution

$$x = \sum_{i=1}^{k} \gamma_i(\alpha) \frac{s_i}{\sigma_i} v_i$$

Filter Factor $\gamma_i(\alpha)$ ($\gamma_i = 0$ when $i > k$)

Regularization parameters:

- truncation $k$ - find the size for the surrogate model.
- regularization parameter $\alpha$ for the hybrid surrogate.
Regularization Parameter Estimation: Find $\alpha^{\text{opt}}$ to minimize $F(\alpha)$

Filter function $\gamma_i(\alpha)$ and complement $\phi_i(\alpha)$.

$\phi(\alpha) = \frac{\alpha^2}{\alpha^2 + \sigma_i^2} = 1 - \gamma_i(\alpha)$, $i = 1 : r$, $\phi_i = 1$, $i > k$.

**Unbiased Predictive Risk**: Minimize functional, noise level $\eta^2$

$$U_k(\alpha) = \sum_{i=1}^{k} \phi_i^2(\alpha)s_i^2 - 2\eta^2 \sum_{i=1}^{k} \phi_i(\alpha)$$

**GCV**: Minimize rational function, $m^* = \min\{m, n\}$

$$G(\alpha) = \frac{\left(\sum_{i=1}^{m^*} \phi_i^2(\alpha)s_i^2\right)}{\left(\sum_{i=1}^{m^*} \phi_i(\alpha)\right)^2}$$

How does $\alpha^{\text{opt}} = \arg\min F(\alpha)$ depend on $k$?
Convergence $\alpha_k$ with $k$ for GCV and UPRE: Examples Restore Tools

Different noise levels: GCV and UPRE

**Grain**
- Mildly Ill-Posed
  \[ \sigma_i = \zeta i^{-\tau}, \frac{1}{2} \leq \tau \leq 1 \]

**Satellite**
- Moderately Ill-Posed
  \[ \sigma_i = \zeta i^{-\tau}, \tau > 1 \]

$\alpha_k$ converges with $k$ and depends on noise level.

Supports use of truncated SVD as surrogate
Assumptions (Normalization)
The system is normalized so that we may assume \( \sigma_1 = 1 \).

Assumptions (Decay Rate)
The measured coefficients \( s_i \) decay according to
\[
|s_i|^2 = \sigma_i^{2(1+\nu)} > \sigma^2 \text{ for } 0 < \nu < 1, 1 \leq i \leq \ell, \text{ i.e. the dominant measured coefficients follow the decay rate of the exact coefficients.}
\]

Assumptions (Noise in Coefficients)
There exists \( \ell \) such that \( E(|s_i|^2) = \sigma^2 \) for all \( i > \ell \), i.e. that the coefficients \( s_i \) are noise dominated for \( i > \ell \).
Theorem
Suppose Assumptions 2 and 3, and that $U_k(\alpha_k)$ is a minimum for $U_k(\alpha)$. Then $\alpha_k > \alpha_\ell > \sigma_{\ell+1}/\sqrt{1 - \sigma_{\ell+1}^2} = \alpha_{\min}$ for $k \geq \ell$.

Theorem
Suppose the decay rate and noise assumptions, and that $\alpha^{\text{opt}}$, and each $\alpha_k$, $k > \ell$ are unique on $\sigma_{\ell+1}/\sqrt{1 - \sigma_{\ell+1}^2} < \alpha < 1$. Then

- $\{\alpha_k\}_{k>\ell}$ is on the average increasing with $\lim_{k \to r} E(\alpha_k) = E(\alpha^{\text{opt}})$.
- $\{U_k(\alpha_k)\}$ is increasing.

Theory can be used to estimate $k$ and $\alpha_k$. 
Comparing Automatic Parameter Estimates by TSVD and SVD

Figure: Box plots comparing parameter estimates $\alpha_k$ with $\alpha^{\text{opt}}$ for problem Satellite computed from 100 runs for noise levels 1\%, 5\%, and 10\%.

Robust algorithm verifies choice of $k$ and $\alpha_k$ with increasing $k$
Comparing Automatic Relative Errors TSVD and SVD

Figure: Box plots comparing relative errors using estimated $k$ and $\alpha_k$ for Full and Truncated SVD: for problem Satellite computed from 100 runs for noise levels 1%, 5%, and 10%.

Surrogate found automatically and error is less than full space
Remark (Observations for UPRE)

1. Find $\alpha_k$ for surrogate model $TSVD A_k = U_k \Sigma_k V_k^T$ with $k$ terms.
2. Determine optimal $k$ as $\alpha_k$ converges to $\alpha^{opt}$.
3. With UPRE for large enough $k$ the full problem is regularized: i.e. $\gamma_i(\alpha_k) \approx 0$ for $i > k$.

Remark (Extending to Large Scale)

- The TSVD for large problems is not feasible?
- Use iterative methods, randomized SVD to find the surrogate model of $A$. 

Large Scale - Hybrid LSQR: Given \( k \) defines range space

**LSQR** Let \( \beta_1 := \|b\|_2 \), and \( e_1^{(k+1)} \) first column of \( I_{k+1} \)
Generate, lower bidiagonal \( B_k \in \mathcal{R}^{(k+1) \times k} \), column orthonormal \( H_{k+1} \in \mathcal{R}^{m \times (k+1)} \), \( G_k \in \mathcal{R}^{n \times k} \)

\[
AG_k = H_{k+1}B_k, \quad \beta_1H_{k+1}e_1^{(k+1)} = b.
\]

**Projected Problem** on projected space: (standard Tikhonov)

\[
w_k(\zeta_k) = \arg\min_{w \in \mathcal{R}^k} \{\|B_k w - \beta_1e_1^{(k+1)}\|_2^2 + \zeta_k^2\|w\|_2^2\}.
\]

**Projected Solution** depends on \( \zeta_k^{\text{opt}} \): Let \( B_k = \tilde{U}\tilde{\Sigma}\tilde{V}^T \)

\[
x_k(\zeta_k^{\text{opt}}) = G_k w_k(\zeta_k^{\text{opt}}) = \beta_1G_k \sum_{i=1}^{k+1} \gamma_i(\zeta_k^{\text{opt}}) \frac{\tilde{u}_i^T e_1^{(k+1)}}{\tilde{\sigma}_i} \tilde{v}_i
\]

\[
= \sum_{i=1}^{k} \gamma_i(\zeta_k^{\text{opt}}) \frac{\tilde{u}_i^T (H_{k+1}^T b)}{\tilde{\sigma}_i} G_k \tilde{v}_i = \sum_{i=1}^{k} \gamma_i(\zeta_k^{\text{opt}}) \frac{\tilde{s}_i}{\tilde{\sigma}_i} G_k \tilde{v}_i
\]

**Approximate SVD**: \( \tilde{A}_k = (H_{k+1}\tilde{U})\tilde{\Sigma}(G_k\tilde{V})^T \)
Hybrid Randomized Singular Value Decomposition: Proto [HMT11]

\( A \in \mathcal{R}^{m \times n} \), target rank \( k \), oversampling parameter \( p \), \( k + p \ll m \). Power factor \( q \). Compute \( A \approx \overline{A}_k = \overline{U}_k \overline{\Sigma}_k \overline{V}_k^T \).

1: Generate a Gaussian random matrix \( \Omega \in \mathcal{R}^{n \times (k+p)} \).
2: Compute \( Y = A\Omega \in \mathcal{R}^{m \times (k+p)} \). \( Y = \text{orth}(Y) \).
3: If \( q > 0 \) repeat \( q \) times \{\( Y = A(A^TY), \ Y = \text{orth}(Y) \}\). Power
4: Form \( B = Y^T A \in \mathcal{R}^{(k+p) \times n} \). \( (Q = Y) \)
5: Economy SVD \( B = U_B \Sigma_B V_B^T, U_B \in \mathcal{R}^{(k+p) \times (k+p)}, V_B \in \mathcal{R}^{k \times k} \)
6: \( \overline{U}_k = QU_B(:, 1 : k), \overline{V}_k = V_B(:, 1 : k), \overline{\Sigma}_k = \Sigma_B(1 : k, 1 : k) \)

Projected RSVD Problem

\[ x_k(\mu_k) = \arg\min_{x \in \mathcal{R}^k} \{ \| \overline{A}_k x - b \|_2^2 + \mu_k^2 \| x \|_2^2 \} \]

\[ = \sum_{i=1}^{k} \gamma_i(\mu_k) \frac{\overline{u}_i^T b}{\overline{\sigma}_i} \overline{v}_i. \]

\[ = \sum_{i=1}^{k} \gamma_i(\mu_k) \frac{\overline{s}_i}{\overline{\sigma}_i} \overline{v}_i. \]

Approximate SVD \( \overline{A}_k = \overline{U}_k \overline{\Sigma}_k \overline{V}_k^T \).
RSVD and LSQR provide approximate TSVD (see references)

<table>
<thead>
<tr>
<th>Model</th>
<th>TSVD</th>
<th>LSQR</th>
<th>RSVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVD</td>
<td>$A_k$</td>
<td>$\tilde{A}_k$</td>
<td>$\overline{A}_k$</td>
</tr>
<tr>
<td></td>
<td>$U_k \Sigma_k V_k^T$</td>
<td>$(H_{k+1} \tilde{U}) \tilde{\Sigma} (G_k \tilde{V})^T$</td>
<td>$\overline{U}_k \Sigma_k \overline{V}_k^T$</td>
</tr>
<tr>
<td>Terms</td>
<td>$s_i u_i^T b$</td>
<td>$(H_{k+1} \tilde{U}_k)^T b$</td>
<td>$\overline{u}_i^T b$</td>
</tr>
<tr>
<td>Basis</td>
<td>$v_i$</td>
<td>$(G_k \tilde{V}_k)_i$</td>
<td>$\overline{v}_i$</td>
</tr>
<tr>
<td>Coeff</td>
<td>$\gamma_i(\alpha_k) \frac{s_i}{\sigma_i} v_i$</td>
<td>$\gamma_i(\zeta_k) \frac{\tilde{s}_i}{\tilde{\sigma}_i} (G_k \tilde{V})_i$</td>
<td>$\gamma_i(\mu_k) \frac{s_i}{\sigma_i} \overline{v}_i$</td>
</tr>
<tr>
<td>$|A - A_k|$</td>
<td>Theorem $\tilde{A}_k$</td>
<td>Theorem $\overline{A}_k$</td>
<td></td>
</tr>
<tr>
<td>sin($\langle V_k, \overline{V}_k \rangle$)</td>
<td>Golub [GvL96]</td>
<td>Jia [Jia17]</td>
<td>Saibaba [Sai]</td>
</tr>
</tbody>
</table>

Accuracy depends on the surrogate model?
Relative Errors using Approximate LSQR/RSVD with oversampling

OS LSQR conquers semi-convergence for small $k$. 
Hybrid LSQR: LSQR with regularization

Relative Errors for Regularized Solutions

Relative Errors less than TSVD for small $k$
Hybrid RSVD: RSVD with regularization

Relative Errors for Regularized Solutions $q = 2$

Relative Errors larger than TSVD for small $k$
Questions to address

1. Both algorithms show semi-convergence.
2. But what is happening with RSVD accuracy?
3. Why is OS for LSQR effective?
4. Relation of $\alpha_k$, $\zeta_k$, $\mu_k$.
5. Can automatic algorithm be applied

**Investigate the surrogate approximation for RSVD and LSQR**
Contrasting RSVD and LSQR spectrum: Mildly Ill-posed

**Figure:** RSVD: Good Approximation of Dominant Singular Values for a problem of size $4096 \times 4096$ using the RSVD algorithm using 100% oversampling, as compared to the exact singular values of the problem.
Figure: LSQR: Good Approximation of fewer dominant singular values for a problem of size $4096 \times 4096$ using the LSQR algorithm with a Krylov subspace of size $k$, as compared to the exact singular values of the problem.
**Figure:** LSQR: Good Approximation of fewer dominant singular values for a problem of size $4096 \times 4096$ using the LSQR algorithm with a Krylov subspace of size $k$, as compared to the exact singular values of the problem. Oversampled 100%
The Lanczos algorithm provides good estimates of extremal singular values

- LSQR exhibits **semi-convergence** as a result.
- LSQR interior eigenvalue approximations *improve* with increasing \( k \) - approximations *stabilize* with increasing \( k \).
- RSVD approximates dominant singular values, does not capture ill-conditioning.
**Figure:** Rank $k$ approximation error RSVD Power with $q = 2$

Power Iteration assists error reduction.
Figure: Rank $k$ approximation error RSVD $q = 2$ and OS LSQR

Oversampling LSQR improves rank $k$ estimate
Figure: RSVD: The canonical angles increase exponentially for subspace $j$ to subspace $k$ from $4096 \times 4096$ using the RSVD algorithm and decrease with OS: Example Size $k = 400$. 

![Graph showing RSVD angles for different percentages of data](image-url)
Figure: RSVD with power iteration 2: The canonical angles increase exponentially for subspace $j$ to subspace $k$ from $4096 \times 4096$ using the RSVD algorithm and decrease with OS: Example Size $k = 400$
**Figure:** LSQR: The canonical angles increase after some subspace size $j^*$ to subspace $k$ from $4096 \times 4096$ using the RSVD algorithm: Example Size $k = 400$
IMPACT: $V$ Basis Matrices (2D) - Lower basis vectors

**LSQR**

**RSVD**

$k = 100$ $p = 100\%$

**RSVD** $q = 2$

$k = 400$ $p = 100\%$
Observations: LSQR and RSVD

1. LSQR : semi-convergence
2. OS LSQR : overcomes semi-convergence
3. RSVD has smaller rank $k$ error than LS.
4. BUT RSVD does not capture the subspace of rank $k$ from a $k + p$ estimate as well as LSQR - canonical angles are larger.
5. Plots of the basis support the reduced accuracy of the RSVD subspaces

Restored solutions at optimal $k = 750, 50$ for RSVD, LSQR, resp.
Restored Regularized Solutions noise level with SNR $\approx 13$

Figure: LSQR $k = 50$

- $k=50$ $p=0\%$
- $k=50$ $p=10\%$
- $k=50$ $p=20\%$
- $k=50$ $p=80\%$
Restored Regularized Solutions noise level with SNR \( \approx 13 \)

Figure: RSVD \( k = 50 \)

- \( k=50 \) p=0\% q=2
- \( k=50 \) p=10\% q=2
- \( k=50 \) p=20\% q=2
- \( k=50 \) p=80\% q=2
Restored Regularized Solutions noise level with SNR $\approx 13$

Figure: LSQR $k = 750$

- $k=750$ $p=0$
- $k=750$ $p=10$
- $k=750$ $p=20$
- $k=750$ $p=80$
Restored Regularized Solutions noise level with SNR $\approx 13$

**Figure:** RSVD $k = 750$

- $k=750$ $p=0\%$ $q=2$
- $k=750$ $p=10\%$ $q=2$
- $k=750$ $p=20\%$ $q=2$
- $k=750$ $p=80\%$ $q=2$
Dominant Subspace Finding dominant singular space of model matrix is important: Oversampling

RSVD / LSQR Trade offs depend on speed by which singular values decrease (degree of ill-posedness)

Cost While LSQR costs more per iteration, provides the dominant subspace more accurately for $k$ small.

Hybrid Implementations stabilize the solution errors.

Future Investigate transfer of noise to the RSVD subspace - apparently inaccurate.
Remark (Messages of the Analysis)

- SVD plays a role in analysis of large datasets?
- Impact of approximating the spectrum by surrogates?
- Important to understand impact of noise on spectrum
- Important to analyze the methods
Some key references

Gene H. Golub and Charles F. van Loan.
*Matrix computations.*

N. Halko, P. G. Martinsson, and J. A. Tropp.
Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions.

Zhongxiao Jia.
The regularization theory of the Krylov iterative solvers LSQR and CGLS for linear discrete ill-posed problems, part I: the simple singular value case.

Convergence of regularization parameters for solutions using the filtered truncated singular value decomposition.
Submitted.

Arvind K. Saibaba.
Analysis of randomized subspace iteration: Canonical angles and unitarily invariant norms.