Improving LSQR with oversampling: application for inverse problems

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

\textsuperscript{1}: School of Mathematical and Statistical Sciences, Arizona State University, renaunt@asu.edu,

\textsuperscript{2}: Institute of Geophysics, University of Tehran, Hubei Subsurface Multi-scale Imaging Key Laboratory, Institute of Geophysics and Geomatics, China University of Geosciences, Wuhan, China

Modern Challenges in Imaging: In the Footsteps of Allan MacLeod Cormack
Motivation and Background
  Inversion Undersampled Magnetic / Gravity Data
  Basic technique: the singular value decomposition (SVD)
  Focusing Inversion: Reweighted Regularization
  [LK83, WR07]

Numerical Methods for Large Scale: Approximating the SVD
  Krylov: Golub Kahan Bidiagonalization - LSQR [PS82]
  Randomized SVD [HMT11]
  Enhancing LSQR by Oversampling: SVDS [Lar98, BR05]

Properties and Simulations
  Contrast Hybrid SVDS
  Angles between singular vectors
  Angles between subspaces
  Image Restoration with Focusing Inversion
  Undersampled Focusing Inversion of Geophysical Data

Conclusions and Future Work
Observation point \( \mathbf{r} = (x, y, z) \)

Vertical magnetic anomaly \( m(\mathbf{r}) \) is given using Biot-Savart Law

\[
m(\mathbf{r}) \propto \int_{d\Omega} K(\mathbf{r}, \mathbf{r}') \kappa(\mathbf{r}') d\delta
\]

Susceptibility \( \kappa(\mathbf{r}') \) at \( \mathbf{r}' = (x', y', z') \)

Linear Relation \( \mathbf{m} = G\kappa \) (or \( \mathbf{b} = A\mathbf{x} \))

**Aim:** Given surface observations \( m_{ij} \) find susceptibility \( \kappa_{ijk} \)

Underdetermined, measurements 5000, unknowns 75000

Practical Approaches for Large Scale Ill-Posed Problems needed
The Model, True Data and Noisy Data

True Measured Data Using Kernel

Measured Data Using Kernel

Measured Data SNR: 18.7907

Target Structure
Example Results: Magnetic for subspace size $k$

**Figure:** LSQR: $k = 15, 1\text{s}$

**Figure:** LSQROS: $10\% \; k = 15, 1\text{s}$

**Figure:** RSVD: $k = 1000, 10\% \; 1073\text{s}$

**Figure:** RSVD: power iteration $k = 1000, 10\% \; 2211\text{s}$
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:

\[ x = \sum_{i=1}^{r} \frac{s_i}{\sigma_i} v_i, \quad s_i = u_i^T b \]
Filtered and Truncated solution

\[ x = \sum_{i=1}^{k} \gamma_i(\alpha) \frac{s_i}{\sigma_i} v_i \], \quad \gamma_i(\alpha) = \frac{\sigma_i^2}{\sigma_i^2 + \alpha^2}, \quad i = 1 \ldots k, \]

Solves Standard Form

\[ x(\alpha) = \arg\min_x \{ \| b - Ax \|^2 + \alpha^2 \| x \|^2 \} \]

\[ x_k(\alpha) \approx \arg\min_x \{ \| b - A_k x \|^2 + \alpha_k^2 \| x \|^2 \} \]

Generalized Tikhonov - \( L \) invertible (transfer to standard form)

\[ x(\alpha) = \arg\min_x \{ \| b - Ax \|^2 + \alpha^2 \| Lx \|^2 \} \]

\[ x_k(\alpha) \approx L^{-1} \left( \arg\min_y \{ \| b - A_k L^{-1} y \|^2 + \alpha_k^2 \| y \|^2 \} \right) \]
Iterative Reweighted Regularization: Focusing Inversion [LK83, WR07] with iteration count $t$:

$$\|Ax - b\|^2 + \alpha^2\|L(t)(x^{(t)} - x^{(t-1)})\|^2, \quad t > 0$$

Regularization operator $L(t)$. $\epsilon$ ensures $L(t)$ invertible

$$(L(t))_{ii} = ((x_i^{(t-1)} - x_i^{(t-2)})^2 + \epsilon)^{-1/4} \quad \epsilon > 0$$

Invertibility use $(L(t))^{-1}$ as right preconditioner for $A$

$$(L(t))^{-1}_{ii} = ((x_i^{(t-1)} - x_i^{(t-2)})^2 + \epsilon)^{1/4} \quad \epsilon > 0$$

Regularization parameter $\alpha_k$ automatic update each $t$.

Cost of $L(t)$ is minimal: it is diagonal

Generalized Tikhonov regularization: System $A_k(L(t))^{-1}$
Define $\beta_1 := \|b\|_2$, $e_1^{(k+1)}$ first column of $I_{k+1}$ and $\beta_1 H_{k+1} e_1^{(k+1)} = b$

Factorize $AG_k = H_{k+1} B_k$, Lanczos vectors $G_k$ and $H_{k+1}$

Lanczos vectors span $\mathcal{K}_{k+1}\{AA^T, b\}$ and $\mathcal{K}_k\{A^T A, A^T b\}$. and are column orthogonal. $B_k \in \mathcal{R}^{(k+1) \times k}$ is lower bidiagonal.

Projected Problem

$$B_k w_k \approx \beta_1 e_1^{(k+1)}, \quad x_k = G_k w_k$$

Hybrid projected problem

$$x_k = G_k \left( \arg\min \{ \| B_k w_k - \beta_1 e_1^{(k+1)} \|^2 + \alpha^2 \| w_k \|^2 \} \right)$$

Solution defined by SVD of $B_k = \tilde{U} \tilde{\Sigma} \tilde{V}^T$

Ritz vectors, columns of $G_k \tilde{V}$ and $H_{k+1} \tilde{U}$, give $\tilde{A}_k$.

**Approximate SVD:**

$$\tilde{A}_k = (H_{k+1} \tilde{U}) \tilde{\Sigma} (G_k \tilde{V})^T$$
A \in \mathcal{R}^{m \times n}, \text{ target rank } k, \text{ oversampling parameter } p,
\begin{align*}
k + p &\ll m, \ m \geq n. \text{ Power factor } q. \text{ Compute } \\
A &\approx \boxed{A_k} = \bar{U}_k \bar{\Sigma}_k \bar{V}_k^T.
\end{align*}

1: \text{ Generate a Gaussian random matrix } \Omega \in \mathcal{R}^{n \times (k+p)}.
2: \text{ Compute } Y = A\Omega \in \mathcal{R}^{m \times (k+p)}. \ Y = \text{qr}(Y)
3: \text{ If } q > 0 \text{ repeat } q \text{ times } \{ [Y, \sim] = \text{qr}(A^T \text{qr}(A Y)) \} \text{ Power }
4: \text{ Form } B = Y^T A \in \mathcal{R}^{(k+p) \times n}. \ (Q = Y)
5: \text{ Find 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: \bar{U}_k = Q U_B(:, 1 : k), \ \bar{V}_k = V_B(:, 1 : k), \ \bar{\Sigma}_k = \Sigma_B(1 : k, 1 : k)

\begin{itemize}
    \item Hybrid projected problem
    \begin{align*}
x_k &= \text{argmin}\{ \| \bar{A}_k x_k - b \|^2 + \alpha^2 \| x_k \|^2 \}
\end{align*}
\end{itemize}

\begin{itemize}
    \item Solution defined by approximation \( \bar{A}_k = \bar{U}_k \bar{\Sigma}_k \bar{V}_k^T \)
\end{itemize}
Extending RSVD for $A \in \mathbb{R}^{m \times n}$, $m << n$ : Undersampled [VRA18]

Compute $\overline{U}_k \in \mathbb{R}^{m \times k}$, $\overline{\Sigma}_k \in \mathbb{R}^{k \times k}$, $\overline{V}_k \in \mathbb{R}^{n \times k}$.

1. Generate a Gaussian random matrix $\Omega \in \mathbb{R}^{(k+p) \times m}$.
2. Compute matrix $Y = \Omega A \in \mathbb{R}^{(k+p) \times n}$.
3. Compute $Q \in \mathbb{R}^{n \times (k+p)}$ via QR factorization $Y^T = QR$.
4. If $q > 0$ repeat $q$ times \{ $[Q, \sim] = qr(A^T qr(AQ))$ \} Power
5. Form $B = AQ \in \mathbb{R}^{m \times (k+p)}$ using factored form of $Q$.
6. Compute the matrix $B^T B \in \mathbb{R}^{(k+p) \times (k+p)}$.
7. Compute the eigen-decomposition of $B^T B$;
   \[ [\tilde{V}_{k+p}, D_{k+p}] = \text{eig}(B^T B). \]
8. Compute $\overline{V}_k = Q\tilde{V}_l(:, 1 : k)$; $\overline{\Sigma}_k = \sqrt{D_l}(1 : k, 1 : k)$; and
   $\overline{U}_k = B\tilde{V}_k(:, 1 : k)\overline{\Sigma}_k^{-1}$.

Yields $\overline{A}_k = \overline{U}_k \overline{\Sigma}_k \overline{V}_k^T$.
Semi-convergence of LSQR, TSVD, RSVD and power RSVD $q = 1$

A $k$ inherits the ill-conditioning of $A_k$

AIM: optimal stable $k$
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.
Theoretical Properties: Contrasting spectrum of Surrogates

Figure: LSQR: (with reorthogonalization) Good Approximation of fewer dominant singular values for a problem of size $4096 \times 4096$ using the LSQR algorithm with Krylov subspace of size $k$, as compared to the exact singular values of the problem.
The LSQR / RSVD spectrum: Key Properties

**LSQR**
- Good estimates of extremal singular values
- Interior eigenvalue approximations *improve* for increasing $k$
- Dominant spectrum *stabilizes*, increasing $k$
- $\tilde{A}_k$ is not an approximation to $A_k$
- Ill-conditioning is captured.

**RSVD**
- Approximates dominant singular values with sufficient oversampling
- With power iteration improved $\overline{A}_k \approx A_k$
- Does not capture ill-conditioning.
LSQROS: Apply LSQR for Krylov space size $k + p$

Use LSQR to space size $k + p$: \( B_{k+p} \in \mathbb{R}^{(k+p+1) \times (k+p)} \).

\[
AG_{k+p} = H_{k+p+1}B_{k+p}
\]

Find SVD (Enlarged Krylov space)

\[
[\tilde{U}, \tilde{\Sigma}, \tilde{V}] = \text{svd}(B_{k+p})
\]

Truncate

\[
\tilde{U} = \tilde{U}(:, 1:k), \quad \tilde{V} = \tilde{V}(:, 1:k), \quad \tilde{\Sigma} = \tilde{\Sigma}(1:k, 1:k)
\]
Extending LSQR: Oversampling

Figure: LSQR add oversampling: Good Approximation of fewer dominant singular values for a problem of size $4096 \times 4096$ using the LSQR algorithm with extended Krylov subspace of size $k$, as compared to the exact singular values of the problem. Oversampled 100%

No small values
Alternative view of oversampling: svds

- Apply svds find dominant singular space of size $k$:

\[
[\tilde{U}, \tilde{\Sigma}, \tilde{V}] = \text{svds}(B_{k+p}, k)
\]

\[\tilde{U} \in \mathcal{R}^{(k+p+1) \times k}, \tilde{V} \in \mathcal{R}^{(k+p) \times k}, \tilde{\Sigma} \text{ is size } k \times k.\]

- Uses tolerances to determine required size of SVD from $B_{k+p}$ that is needed.

- When $p$ relatively large compared to $k$, the SVD may not be of size $k + p$.

- Use existing software: eg Propack: lansvd [Lar98, BR05].
Why not just abandon LSQR and apply svds directly to $A$?

For $A \in \mathbb{R}^{m \times n}$, target rank $k$, using space size $k + p$

- Directly use svds to find dominant space of size $k$ using extended Krylov space $k + p$.

\[
[U_k, \Sigma_k, V_k] = \text{svds}(A, k, 'SubspaceDimension', k + p)
\]

- Approximate SVD is immediate if tolerance is met.
- Can adjust $p$ if tolerance is not met. e.g. iterate on $k + p$ to force acceptable tolerance for size $k$.

Why LSQROS and not svds($A$)?
Comparison of relative error : SVDSB converges to SVDS
Comparison of time SVDS more expensive than SVDSB
Gravity: Problem Size 5000 by 75000 (One Step)

Comparison of Time: SVDSB expensive increasing $k$

But LSQR also expensive as compared to RSVD
Hybrid RSVD: RSVD with regularization and oversampling

Relative Errors larger than TSVD for small $k$
Relative Errors less than TSVD for small $k$

SVDSB reduces semiconvergence issue of LSQR
Preliminary Summary

SVDS / SVDSB

1. SVDS can be used in place of LSQR
2. SVDS can be applied directly to projected problem
3. SVDSB cheaper than SVDS.

RSVD / LSQR options

1. OS for LSQR is effective for small $k$
2. RSVD is not effective for small $k$

Spectrum approximation is not a sufficient guide for accuracy
RSVD and LSQR provide approximate TSVD (see references)

<table>
<thead>
<tr>
<th></th>
<th>TSVD</th>
<th>LSQR</th>
<th>RSVD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>$k$</td>
<td>$k$</td>
<td>$k + p$</td>
</tr>
<tr>
<td>Model SVD</td>
<td>$A_k$</td>
<td>$\tilde{A}_k$</td>
<td>$\overline{A}_k$</td>
</tr>
<tr>
<td>Terms $s_i$</td>
<td>$U_k \Sigma_k V_k^T$</td>
<td>$(H_{k+1} \tilde{U}) \Sigma (G_k \tilde{V})^T$</td>
<td>$(H_{k+1} \overline{U}<em>k)</em>{i} \Sigma (G_k \tilde{V})^T$</td>
</tr>
<tr>
<td>Basis</td>
<td>columns of $V_k$</td>
<td>columns of $G_k \tilde{V}_k$</td>
<td>columns of $G_k \tilde{V}_k$</td>
</tr>
<tr>
<td></td>
<td>$\sigma_{k+1}$</td>
<td>$\sigma_{k+1}$</td>
<td>$\sigma_{k+1}$</td>
</tr>
<tr>
<td>$\sin(\langle V_k, \cdot \rangle)$</td>
<td>Delft et al [DDLT91]</td>
<td>Delft et al [DDLT91]</td>
<td>Delft et al [DDLT91]</td>
</tr>
<tr>
<td></td>
<td>Theorem $\tilde{A}_k$</td>
<td>Theorem $\tilde{A}_k$</td>
<td>Theorem $\tilde{A}_k$</td>
</tr>
</tbody>
</table>

**Accuracy is well-studied**

Other Properties: rank $k$ approximation of $A$
Theorems on error of near rank $k$ best approximation $\|A - \overline{A}_k\|$

Theorem (RSVD Proto: with power iteration $q$ [HMT11])

$$E(\|A - \overline{A}_k\|) \leq \left(1 + \left[1 + 4\sqrt{\frac{2\min\{m, n\}}{k - 1}}\right]^{1/(2q+1)}\right)\sigma_{k+1}$$

Theorem (LSQR [Jia17]: Fast decay of singular values

$\sigma_i = \zeta \rho^{-i}$, $\rho > 2$ and noise level contaminates at coefficient $\ell$)

$\tilde{A}_k = H_{k+1}B_kG_k^T$ is a near best rank $k$ approximation to $A$ for $k = 1, 2, \ldots, \ell - 1$. 


Theorems on approximation of the spectral space: Angles between subspaces (vectors) formed by TSVD and approximate TSVD:

Theorem ([DDLT91]: For LSQR ($\sigma_i \neq \sigma_j$) and $\| A - \tilde{A}_k \| \leq \nu_k \| A \| = \nu_k \sigma_1$ if $2\nu_k < \min_{i \neq j} |\sigma_i - \sigma_j|$, then )

$$\max(\sin \theta(u_i, \tilde{u}_i), \sin \theta(v_i, \tilde{v}_i)) \leq \frac{\nu_k}{\min_{i \neq j} |\sigma_i - \sigma_j| - \nu_k} \leq 1.$$  

Theorem (Convergence of Lanczos Vectors [Saa11, Theorem 6.3]. $L_i(\sigma_n^2) = \prod_{j=1}^{i-1} \frac{(\sigma_j)^2 - \sigma_n^2}{(\sigma_j)^2 - \sigma_i^2} \cdot \rho_i = \frac{\sigma_i^2 - \sigma_{i+1}^2}{\sigma_{i+1}^2 - \sigma_n^2}$, Chebyshev polynomial $C_{k-i}$)

$$\tan(\Theta(v_i, G_k)) \leq \frac{L_i(\sigma_n^2)}{C_{k-i}(1 + 2\rho_i)} \tan(\Theta(v_i, G_1)).$$

Theorem ([Sai19, Theorem 4] RSVD canonical subspace angles $i = 1 \to k$. $\gamma_i = \sigma_{k+1}/\sigma_i$)

$$\max(\sin \theta(u_i, \bar{u}_i), \sin \theta(v_i, \bar{v}_i)) \leq \gamma_i \frac{\gamma_k^{2q}}{1 - \gamma_k} \| (V_{\perp}^T \Omega)(V_k^T \Omega)^\dagger \|_2$$
Contrasting sines of angles between singular vectors

Figure: Fixed $k = 200$ and increasing $p$
Contrasting sines of angles between singular vectors

Figure: Fixed $k = 500$ and increasing $p$
Contrasting Sines of Subspace Canonical Angles:

Figure: RSVDP ($p = 10$): $\sin \Theta(U_\ell, \overline{U}_\ell)$, $\sin \Theta(V_\ell, \overline{V}_\ell)$ Increasing $k$: $\ell = f k$, $f$ is percent

Sine of subspace angle RSVDP $p = 10$
Contrasting Sines of Subspace Canonical Angles:

**Figure:** RSVDQ ($p = 10$): $\sin \Theta(U_\ell, \overline{U}_\ell)$, $\sin \Theta(V_\ell, \overline{V}_\ell)$ Increasing $k$: $\ell = f k$, $f$ is percent.
Figure: LSQROS \((p = 10)\): \(\sin \Theta(U_\ell, \tilde{U}_\ell)\), \(\sin \Theta(V_\ell, \tilde{V}_\ell)\) Increasing \(k\): \(\ell = f k\), \(f\) is percent.

Sine of subspace angle LSQRO \(p = 10\)
Contrasting Sines of Subspace Canonical Angles:

Figure: LSQROS ($p = 50$); $\sin \Theta(U_\ell, \tilde{U}_\ell)$, $\sin \Theta(V_\ell, \tilde{V}_\ell)$ Increasing $k$:

$\ell = f \frac{k}{f}, f$ is percent
Observations: LSQR and RSVD

1. Canonical angles between the singular vectors are far smaller for LSQR than RSVD and RSVDQ, particularly with oversampling.

2. Canonical angles between dominant subspaces are far smaller for LSQR than RSVD for equivalent small \( k \).

3. RSVD does not capture the subspace of rank \( k \) from a \( k + p \) estimate as well as LSQROS - canonical angles are larger.

4. Subspace alignment stabilizes for LSQROS.

5. Conclude: LSQROS better mimics TSVD.
Restoration Grain size $256 \times 256$: SNR 20: Dominant space of size 500

Figure: True Data

Figure: Blurred Noisy Data

Figure: LSQR

Figure: RSVDQ

Figure: LSQROS

Stabilization with LSQROS and 10% oversampling
Gravity Results: Volume Rendering

**LSQR** $k = 1000$, **LSQROS** $k = 150$, 10% oversampling

**RSVD** $k = 1000$ and **RSVDQ** $k = 1000$, 10% oversampling
Gravity Results: UPRE for parameter estimation

**LSQR with** $k = 1000$, time $1016s$, 0% oversampling (14 iterations)

**LSQROS with** $k = 150$, time $43s$, 10% oversampling (14 iterations)

**RSVD with** $k = 1000$, time $339s$, 10% oversampling (15 iterations)

**RSVDQ with** $k = 1000$, time $537s$ 10% oversampling (13 iterations)
Gravity Results: Relative Error

LSQR and LSQROS, 10% oversampling

RSVD and RSVDQ, 10% oversampling
Conclusions: LSQR with Extended Krylov

Canonical Angles Accurate dominate subspace is critical.

Extension of Krylov Space Improves dominant space accuracy.

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

Cost While LSQROS more expensive than LSQR, provides the dominant subspace more accurately for $p$ small.

Hybrid Implementations stabilize the solution errors.

Heuristics verified on a practical application.

Future
  ▶ Apply for Generalized Regularizers
  ▶ Stabilize RSVD oversampling choice using $\text{svds}$?
Some key references

James Baglama and Lothar Reichel.
Augmented implicitly restarted lanczos bidiagonalization methods.

Percy Deift, James Demmel, Luen-Chau Li, and Carlos Tomei.
The bidiagonal singular value decomposition and Hamiltonian mechanics.

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.

Rasmus Larsen.
Lanczos bidiagonalization with partial reorthogonalization.

B. J. Last and K. Kubik.
Compact gravity inversion.

Christopher C. Paige and Michael A. Saunders.
LSQR: an algorithm for sparse linear equations and sparse least squares.

Y. Saad.
A. Saibaba.
Randomized subspace iteration: Analysis of canonical angles and unitarily invariant norms. 

Saeed Vatankhah, Rosemary A. Renaut, and Vahid E. Ardestani.
A fast algorithm for regularized focused 3-D inversion of gravity data using the randomized SVD. 

Brendt Wohlberg and Paul Rodríguez.
An iteratively reweighted norm algorithm for minimization of total variation functionals. 