Singular Value Decomposition (SVD)
Outline

- Linear problem
- SVD (Singular Value Decomposition)
- Generalized inverse
- Controlling ill-conditioning
- Truncated SVD
- Understanding information from data alone
Our statement of the inverse problem

- Given observations: $d_j^{obs}, \ j = 1, \ldots, N$
  - Uncertainties: $\epsilon_j$
  - Ability for forward modelling: $\mathcal{F}[m] = d$

- Find the earth model that gave rise to the data.
Linear problem

\[ d_j = \int_v g_j(x)m(x)\,dx \quad \text{and} \quad g_j(x) = e^{jpx}\cos(2\pi jqx) \]

- \(d_j\): j-th datum
- \(g_j\): kernel function for j-th datum
- \(m\): model

Each datum in an inner product of the model with the kernel function. Evaluate one case

\[ d_j = g \cdot m = 4.89 \]
Basic challenges for the inverse problem:

Non-uniqueness

- Data are
  \[ d_i = \sum_{j=1}^{M} G_{ij} m_j \]
  \[ \mathbf{d} = \mathbf{Gm} \quad \mathbf{G}: (N \times M) \]

- \( M > N \) (underdetermined problem) \( \rightarrow \) infinitely many solutions
- Ill-conditioned
  - Small changes in the data yields large changes in the model

How do we understand and deal with these elements?
Solving \( \mathbf{Gm} = \mathbf{d} \)

Given

\[
\mathbf{Gm} = \mathbf{d}
\]

\( \mathbf{G} \in \mathbb{R}^{N \times M} \quad \mathbf{m} \in \mathbb{R}^{M} \quad \mathbf{d} \in \mathbb{R}^{N} \)

Want to write

\[
\mathbf{m} = \mathbf{G}^{-1}\mathbf{d}
\]

\( \mathbf{G}^{-1}: \text{does not exist} \)

\( \mathbf{G}: \text{not sparse} \)

Example

\[
m_1 + 2m_2 = 2
\]

Write as

\[
\mathbf{Gm} = \mathbf{d}
\]

\( \mathbf{G} = [1, 2] \quad \mathbf{d} = 2 \quad \mathbf{m} = (m_1, m_2)^{\top} \)

Think

\[
\mathbf{m} = \mathbf{G}^{-1}\mathbf{d}
\]

Use whatever language you use (e.g. matlab and python). What do you get
Singular Value Decomposition

\[ \mathbf{G} = \mathbf{U}_p \mathbf{\Lambda} \mathbf{V}_p^T \]

\[ d_j = \int_v g_j(x)m(x)dx \]

\[ \mathbf{G} = \begin{pmatrix} g_1 \\ \vdots \\ g_N \end{pmatrix} \]

\[ (M \times N) \]

\[ \mathbf{U}_p = \begin{pmatrix} \mathbf{u}_1 & \ldots & \mathbf{u}_p \end{pmatrix} \]

\[ (N \times P) \]

\[ \mathbf{\Lambda}_p = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_p \end{pmatrix} \]

\[ (P \times P) \]

\[ \mathbf{V}_p = \begin{pmatrix} \mathbf{v}_1 & \ldots & \mathbf{v}_p \end{pmatrix} \]

\[ (M \times P) \]

\[ \lambda_i: \text{singular values} \ (\lambda_1 \geq \lambda_2 \geq \ldots \lambda_p > 0) \]

\[ \mathbf{v}_i: \text{singular vectors (model space; } \mathbf{V}_p^T \mathbf{V}_p = \mathbf{I}_p) \]

\[ \mathbf{u}_i: \text{singular vectors (data space; } \mathbf{U}_p^T \mathbf{U}_p = \mathbf{I}_p) \]
Singular Value Decomposition

\[ Gv_i = \lambda_i u_i \]

\[ G^T u_i = \lambda_i v_i \]

\( m^\parallel \): activated portion of model space

\( m^\perp \): annihilator space
Solving with SVD

\[ \mathbf{Gm} = \mathbf{d} \quad \text{with} \quad \mathbf{G} = \mathbf{U}\Lambda\mathbf{V}^T \]

\[ \mathbf{U}\Lambda\mathbf{V}^T \mathbf{m} = \mathbf{d} \]

\[ \Lambda\mathbf{V}^T \mathbf{m} = \mathbf{U}^T \mathbf{d} \]

\[ \mathbf{V}^T \mathbf{m} = \Lambda^{-1} \mathbf{U}^T \mathbf{d} \]

\[ \mathbf{VV}^T \mathbf{m} = \mathbf{V}\Lambda^{-1} \mathbf{U}^T \mathbf{d} \]

Effects: Successively add vectors

Each vector \( \mathbf{v}_i \) scaled by \( \frac{\mathbf{u}_i^T \mathbf{d}}{\lambda_i} \to \infty \) if \( \lambda_i \to 0 \)

Define

\[ \mathbf{G}^\dagger = \mathbf{V}\Lambda^{-1} \mathbf{U}^T \]

\[ \mathbf{m}_c = \mathbf{G}^\dagger \mathbf{d} \]

\[ \mathbf{m}_\parallel = \mathbf{m}_c = \mathbf{VV}^T \mathbf{m} = \mathbf{V}\Lambda^{-1} \mathbf{U}^T \mathbf{d} \]

\[ \mathbf{m}_c = \mathbf{V}\Lambda^{-1} \mathbf{U}^T \mathbf{d} = \sum_{i=1}^{N} \left( \frac{\mathbf{u}_i^T \mathbf{d}}{\lambda_i} \right) \mathbf{v}_i \]

\[ \phi_d = \| \mathbf{Gm} - \mathbf{d} \|^2 \]

\[ \phi_m = \| \mathbf{m} \|^2 \]

\[ \mathbf{m}_c = \sum_{i=1}^{p} \left( \frac{\mathbf{u}_i^T \mathbf{d}}{\lambda_i} \right) \mathbf{v}_i \]
Ill-conditionedness

$$m_c = \sum_{i=1}^{p} \left( \frac{u_i^T d}{\lambda_i} \right) v_i$$

Small singular value  ↔  High oscillatory
Large amplitude of noise

<table>
<thead>
<tr>
<th>$\lambda_i$</th>
<th>$v_i$</th>
<th>$u_i$</th>
<th>$u_i^T d^{true}$</th>
<th>$u_i^T d^{true} / \lambda_i$</th>
<th>$u_i^T n$</th>
<th>$u_i^T n / \lambda_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda_1 = 0.43$</td>
<td>![Line graph]</td>
<td>![Line graph]</td>
<td>-0.29</td>
<td>-2.13</td>
<td>-0.16</td>
<td>-1.16</td>
</tr>
<tr>
<td>$\lambda_2 = 0.11$</td>
<td>![Line graph]</td>
<td>![Line graph]</td>
<td>0.41</td>
<td>3.8</td>
<td>0.02</td>
<td>0.22</td>
</tr>
<tr>
<td>$\lambda_3 = 0.09$</td>
<td>![Line graph]</td>
<td>![Line graph]</td>
<td>0.16</td>
<td>1.9</td>
<td>0.08</td>
<td>0.96</td>
</tr>
<tr>
<td>$\lambda_{11} = 10^{-6}$</td>
<td>![Line graph]</td>
<td>![Line graph]</td>
<td>$3.4 \times 10^{-7}$</td>
<td>0.32</td>
<td>-0.02</td>
<td>$-1.9 \times 10^4$</td>
</tr>
<tr>
<td>$\lambda_{10} = 2.3 \times 10^{-17}$</td>
<td>![Line graph]</td>
<td>![Line graph]</td>
<td>$-6.2 \times 10^{-17}$</td>
<td>-2.7</td>
<td>-0.0001</td>
<td>$4.6 \times 10^{13}$</td>
</tr>
</tbody>
</table>
Truncated SVD

• If data are inaccurate, noise is also amplified by \(1/\lambda_i\)

\[
m_c = \sum_{i=1}^{q} \left( \frac{u_i^T d}{\lambda_i} \right) v_i + \sum_{i=q+1}^{N} \left( \frac{u_i^T d}{\lambda_i} \right) v_i
\]

Cause more harm than good

• So

\[
m_c = \sum_{i=1}^{q} \left( \frac{u_i^T d}{\lambda_i} \right) v_i
\]

• Solution lies in a small sub-space

• Treats non-uniqueness and ill-conditioning
### Truncated SVD

<table>
<thead>
<tr>
<th>$\mathbf{v}_i$</th>
<th>$\lambda_i$</th>
<th>$\mathbf{m}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbf{v}_3$</td>
<td>$\lambda_3 = 0.09$</td>
<td>$\mathbf{m}<em>c = \sum</em>{i=1}^{q} \left( \frac{\mathbf{u}_i^T \mathbf{d}}{\lambda_i} \right) \mathbf{v}_i$ $q = 3$</td>
</tr>
<tr>
<td>$\mathbf{v}_5$</td>
<td>$\lambda_5 = 0.04$</td>
<td>$\mathbf{m}<em>c = \sum</em>{i=1}^{q} \left( \frac{\mathbf{u}_i^T \mathbf{d}}{\lambda_i} \right) \mathbf{v}_i$ $q = 5$</td>
</tr>
<tr>
<td>$\mathbf{v}_6$</td>
<td>$\lambda_6 = 0.02$</td>
<td>$\mathbf{m}<em>c = \sum</em>{i=1}^{q} \left( \frac{\mathbf{u}_i^T \mathbf{d}}{\lambda_i} \right) \mathbf{v}_i$ $q = 6$</td>
</tr>
<tr>
<td>$\mathbf{v}_{19}$</td>
<td>$\lambda_{19} = 2.3 \times 10^{-17}$</td>
<td>$\mathbf{m}<em>c = \sum</em>{i=1}^{q} \left( \frac{\mathbf{u}_i^T \mathbf{d}}{\lambda_i} \right) \mathbf{v}_i$ $q = 18$</td>
</tr>
</tbody>
</table>
Plot Tikhonov curve

$q = 3$
Plot Tikhonov curve

$q = 3$

$q = 4$
Plot Tikhonov curve

$q = 3$

$q = 4$

$q = 5$

$q = 6$
Plot Tikhonov curve

$q = 3$

$q = 4$

$q = 5$

$q = 6$
Consider Tikhonov solution

Data misfit:  \[ \phi_d = \| \mathbf{Gm} - \mathbf{d} \|^2 \]

Regularization:  \[ \phi_m = \| \mathbf{m} \|^2 \]

\[
\begin{align*}
\text{minimize} & \quad \phi = \phi_d + \beta \phi_m \\
\text{minimum} & \quad \nabla \mathbf{g} = \nabla_m \phi = 0
\end{align*}
\]

Differentiate w.r.t vector

\[
\frac{\partial \phi}{\partial \mathbf{m}} = 2 \mathbf{G}^T(\mathbf{Gm} - \mathbf{d}) + 2\beta \mathbf{m} = 0
\]

So

\[
(\mathbf{G}^T \mathbf{G} + \beta \mathbf{I}) \mathbf{m} = \mathbf{G}^T \mathbf{d}
\]

\[
\mathbf{G}^T \mathbf{G}: M \times M \text{ (rank deficient)} \\
\text{rank}(\mathbf{G}) \leq N, M > N
\]

but $\beta > 0$, so adding $\beta \mathbf{I}$ makes all eigen values $\geq \beta$

Solve

\[
\mathbf{A} \mathbf{m} = \mathbf{b}
\]

with letting  \[ \mathbf{A} = \mathbf{G}^T \mathbf{G} + \beta \mathbf{I} \]

\[ \mathbf{b} = \mathbf{G}^T \mathbf{d} \]
Tikhonov curve

\[(G^T G + \beta I)\mathbf{m} = G^T \mathbf{d}\]
Tikhonov solution vs TSVD

$$(G^T G + \beta I)m = G^T d$$

So solving by TSVD, or Tikhonov yields almost identical
Tikhonov, TSVD, and weighted TSVD

\[(G^T G + \beta I) m = G^T d\]

- In fact, exact correspondence is obtained by modifying to include weights

\[m_c = VTA^{-1} Ud\]

\[m_c = \sum_{i=1}^{p} t_i \left( \frac{u_i^T d}{\lambda_i} \right) v_i\]

\[T = \text{diag}(t_1, \ldots, t_p) \quad 0 \leq t_i \leq 1\]

\[t_i = \frac{\lambda_i^2}{\lambda_i^2 + \beta}\]
Summary

\[ \mathbf{G} \mathbf{m} = \mathbf{d} \]
\[ \mathbf{G} = \mathbf{U} \Lambda \mathbf{V}^T \]

\[ \mathbf{m}_c = \sum_{i=1}^{p} t_i \left( \frac{\mathbf{u}_i^T \mathbf{d}}{\lambda_i} \right) \mathbf{v}_i \]

- Small singular values:
  - Associated with high frequency vectors in model space
  - Amplify the noise

- Handling instabilities
  1. By truncating (keep only p)
  2. Adjust their influence

\[ t_i = \frac{\lambda_i}{\lambda_i^2 + \beta} \]

this exactly the same as solving the Tikhonov problem

\[
\text{minimize } \phi = \phi_d + \beta \phi_m \\
= \| \mathbf{G} \mathbf{m} - \mathbf{d} \|^2 + \beta \| \mathbf{m} \|^2
\]
What does the solution tell us

• Full data set can recover information about $m^{\parallel}$

• Regularized solution is in a reduced region of model space

Geophysical model lies outside of this region
To explore that we need to incorporate more information
Outline

• Linear problem
• SVD (Singular Value Decomposition)
• Generalized inverse
• Controlling ill-conditioning
• Truncated SVD
• Understanding information from data alone
Next up …