Matrix Computations
Chapter 6: Singular value, Singular Vectors, and Singular Value Decomposition

Section 6.3 SVD for Linear Systems

Jie Lu
ShanghaiTech University

Linear Systems: Sensitivity Analysis

Given nonsingular $\mathbf{A} \in \mathbb{R}^{n\times n}$ and $\mathbf{y} \in \mathbb{R}^n$, let $\mathbf{x}$ be the solution to

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

Consider a perturbed version of the above system:

$$\hat{\mathbf{A}} = \mathbf{A} + \Delta\mathbf{A}, \quad \hat{\mathbf{y}} = \mathbf{y} + \Delta\mathbf{y}$$

where $\Delta\mathbf{A}$ and $\Delta\mathbf{y}$ are errors (e.g., floating point errors, measurement errors, etc.)

Let $\hat{\mathbf{x}}$ be a solution to the perturbed system

$$\hat{\mathbf{y}} = \hat{\mathbf{A}}\hat{\mathbf{x}}$$

Problem: Analyze how the solution error $\|\hat{\mathbf{x}} - \mathbf{x}\|_2$ scales with $\Delta\mathbf{A}$ and $\Delta\mathbf{y}$

Remark: We have already studied sensitivity analysis of linear systems in Section 1.3. Here, we focus on its relation with SVD

Condition Number

The condition number of matrix $\mathbf{A}$ is defined as

$$\kappa(\mathbf{A}) = \|\mathbf{A}\| \cdot \|\mathbf{A}^{-1}\|$$

Let the above norm be 2-norm. Then, $\|\mathbf{A}\|_2 = \sigma_{\max}(\mathbf{A})$, $\|\mathbf{A}^{-1}\|_2 = 1/\sigma_{\min}(\mathbf{A})$, and

$$\kappa(\mathbf{A}) = \frac{\sigma_{\max}(\mathbf{A})}{\sigma_{\min}(\mathbf{A})}$$

For nonsingular $\mathbf{A}$, $\sigma_{\max}(\mathbf{A}) \geq \sigma_{\min}(\mathbf{A}) > 0$

Thus, $\kappa(\mathbf{A}) \geq 1$, and $\kappa(\mathbf{A}) = 1$ if $\mathbf{A}$ is orthogonal

- $\mathbf{A}$ is said to be ill-conditioned if $\kappa(\mathbf{A})$ is very large, referring to the cases where $\mathbf{A}$ is close to singular

Sensitivity Analysis

Theorem

Let $\varepsilon > 0$ *be s.t.*

$$\frac{\|\Delta \mathbf{A}\|_2}{\|\mathbf{A}\|_2} \leq \varepsilon, \qquad \frac{\|\Delta \mathbf{y}\|_2}{\|\mathbf{y}\|_2} \leq \varepsilon.$$

If ε *is sufficiently small s.t.* $\varepsilon\kappa(\mathbf{A}) < 1$, *then*

$$\frac{\|\hat{\mathbf{x}} - \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq \frac{2\varepsilon\kappa(\mathbf{A})}{1 - \varepsilon\kappa(\mathbf{A})}$$

Implications:

- For small errors and in the worst-case sense, the relative error $\|\hat{\mathbf{x}} - \mathbf{x}\|_2 / \|\mathbf{x}\|_2$ tends to increase with the condition number
- In particular, for $\varepsilon\kappa(\mathbf{A}) \leq \frac{1}{2}$, the error bound is simplified to

$$\frac{\|\hat{\mathbf{x}} - \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \leq 4\varepsilon\kappa(\mathbf{A})$$

Proof

Proof (cont'd)

Interpretation of Linear Systems under SVD

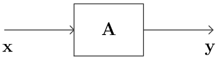
Consider the linear system

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

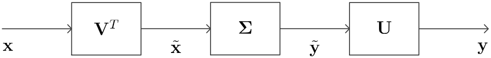
where $\mathbf{A} \in \mathbb{R}^{m\times n}$ is the system matrix, $\mathbf{x} \in \mathbb{R}^n$ is the system input, and $\mathbf{y} \in \mathbb{R}^m$ is the system output

Using SVD $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$, we can write

$$\mathbf{y} = \mathbf{U}\tilde{\mathbf{y}}, \qquad \tilde{\mathbf{y}} = \Sigma\tilde{\mathbf{x}}, \qquad \tilde{\mathbf{x}} = \mathbf{V}^T\mathbf{x}$$



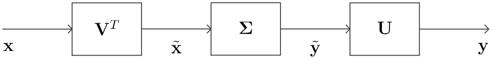
(a) linear system



(b) equivalent system

Interpretation of Linear Systems under SVD (cont'd)

(a) linear system



(b) equivalent system

Implication: *All* linear systems work by performing three processes in cascade

- $\tilde{\mathbf{x}} = \mathbf{V}^T\mathbf{x}$: Let $\mathbf{x}$ resolve into $\mathbf{v}_1, \ldots, \mathbf{v}_n$ (rotate by $\mathbf{V}^T$)
- $\tilde{\mathbf{y}} = \boldsymbol{\Sigma}\tilde{\mathbf{x}}$: Element-wise scale the first $p = \min\{m, n\}$ elements of $\tilde{\mathbf{x}}$ by $\sigma_i \geq 0$, $i = 1, \ldots, p$, and then either truncate or zero-pad to obtain the m-dimensional $\tilde{\mathbf{y}}$
- $\mathbf{y} = \mathbf{U}\tilde{\mathbf{y}}$: Reconstitute with basis $\mathbf{u}_1, \ldots, \mathbf{u}_m$ (rotate by $\mathbf{U}$)

Solution of Linear Systems via SVD

Problem: Given general $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{y} \in \mathbb{R}^m$, determine

- whether $\mathbf{y} = \mathbf{A}\mathbf{x}$ has a solution (more precisely, whether there exists an $\mathbf{x}$ such that $\mathbf{y} = \mathbf{A}\mathbf{x}$)
- what is the solution

It can be shown via SVD that

$$
\begin{aligned}
\mathbf{y} = \mathbf{A}\mathbf{x} \quad &\Longleftrightarrow \quad \mathbf{y} = \mathbf{U}_1 \tilde{\Sigma} \mathbf{V}_1^T \mathbf{x} \\
&\Longleftrightarrow \quad \mathbf{U}_1^T \mathbf{y} = \tilde{\Sigma} \mathbf{V}_1^T \mathbf{x}, \ \mathbf{U}_2^T \mathbf{y} = \mathbf{0} \\
&\Longleftrightarrow \quad \mathbf{V}_1^T \mathbf{x} = \tilde{\Sigma}^{-1} \mathbf{U}_1^T \mathbf{y}, \ \mathbf{U}_2^T \mathbf{y} = \mathbf{0} \\
&\Longleftrightarrow \quad \begin{array}{l} \mathbf{x} = \mathbf{V}_1 \tilde{\Sigma}^{-1} \mathbf{U}_1^T \mathbf{y} + \boldsymbol{\eta}, \\ \text{for any } \boldsymbol{\eta} \in \mathcal{R}(\mathbf{V}_2) = \mathcal{N}(\mathbf{A}), \\ \mathbf{U}_2^T \mathbf{y} = \mathbf{0} \end{array}
\end{aligned}
$$

Solution of Linear Systems via SVD (cont'd)

$$\mathbf{y} = \mathbf{A}\mathbf{x} \quad \Longleftrightarrow \quad \begin{array}{l} \mathbf{x} = \mathbf{V}_1 \tilde{\Sigma}^{-1} \mathbf{U}_1^T \mathbf{y} + \boldsymbol{\eta}, \\ \text{for any } \boldsymbol{\eta} \in \mathcal{R}(\mathbf{V}_2) = \mathcal{N}(\mathbf{A}), \\ \mathbf{U}_2^T \mathbf{y} = \mathbf{0} \end{array}$$

Case (a): Full-column rank $\mathbf{A}$, i.e., $r = n \leq m$

- There is no $\mathbf{V}_2$, and $\mathbf{U}_2^T \mathbf{y} = \mathbf{0}$ is equivalent to $\mathbf{y} \in \mathcal{R}(\mathbf{U}_1) = \mathcal{R}(\mathbf{A})$
- **Result**: The linear system has a solution if and only if $\mathbf{y} \in \mathcal{R}(\mathbf{A})$, and the solution, if exists, is uniquely given by $\mathbf{x} = \mathbf{V}_1 \tilde{\Sigma}^{-1} \mathbf{U}_1^T \mathbf{y}$

Case (b): Full-row rank $\mathbf{A}$, i.e., $r = m \leq n$

- There is no $\mathbf{U}_2$
- **Result**: The linear system always has a solution, and the solution is given by $\mathbf{x} = \mathbf{V}_1 \tilde{\Sigma}^{-1} \mathbf{U}^T \mathbf{y} + \boldsymbol{\eta}$ for any $\boldsymbol{\eta} \in \mathcal{N}(\mathbf{A})$

Least Squares via SVD

Consider the LS problem: Given general $\mathbf{A} \in \mathbb{R}^{m \times n}$,

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$$

For any $\mathbf{x} \in \mathbb{R}^n$,

$$\begin{aligned}
\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 = \|\mathbf{y} - \mathbf{U}\Sigma \underbrace{\mathbf{V}^T\mathbf{x}}_{=\tilde{\mathbf{x}}}\|_2^2 &= \| \underbrace{\mathbf{U}^T\mathbf{y}}_{=\tilde{\mathbf{y}}} - \Sigma\tilde{\mathbf{x}}\|_2^2 \\
&= \textstyle\sum_{i=1}^{r} |\tilde{y}_i - \sigma_i \tilde{x}_i|^2 + \sum_{i=r+1}^{p} |\tilde{y}_i|^2 \\
&\geq \textstyle\sum_{i=r+1}^{p} |\tilde{y}_i|^2
\end{aligned}$$

where the equality can be attained if $\tilde{\mathbf{x}}$ satisfies $\tilde{y}_i = \sigma_i \tilde{x}_i$ for $i = 1, \ldots, r$

Least Squares via SVD (cont'd)

It can be shown that such a $\tilde{\mathbf{x}}$ corresponds to

$$\mathbf{x} = \mathbf{V}_1 \tilde{\Sigma}^{-1} \mathbf{U}_1^T \mathbf{y} + \mathbf{V}_2 \mathbf{x}_2' \text{ for any } \mathbf{x}_2' \in \mathbb{R}^{n-r}$$

which is the desired LS solution

Pseudo-Inverse

Given $\mathbf{A} \in \mathbb{R}^{m \times n}$, a pseudo-inverse of $\mathbf{A}$ is defined as a matrix $\mathbf{A}^{\dagger} \in \mathbb{R}^{n \times m}$ satisfying the Moore-Penrose conditions:

(i) $\mathbf{A}\mathbf{A}^{\dagger}\mathbf{A} = \mathbf{A}$; (ii) $\mathbf{A}^{\dagger}\mathbf{A}\mathbf{A}^{\dagger} = \mathbf{A}^{\dagger}$; (iii) $\mathbf{A}\mathbf{A}^{\dagger}$ is symmetric (iv) $\mathbf{A}^{\dagger}\mathbf{A}$ is symmetric

Given the thin SVD $\mathbf{A} = \mathbf{U}_1 \tilde{\Sigma} \mathbf{V}_1^T$,

$$\mathbf{A}^{\dagger} = \mathbf{V}_1 \tilde{\Sigma}^{-1} \mathbf{U}_1^T$$

- $\mathbf{x}_{\text{LS}} = \mathbf{A}^{\dagger}\mathbf{y} + \boldsymbol{\eta}$ for any $\boldsymbol{\eta} \in \mathcal{R}(\mathbf{V}_2)$
- The same applies to the linear system $\mathbf{y} = \mathbf{A}\mathbf{x}$ that has a solution
- When $\mathbf{A}$ has full column rank
  - $\mathbf{A}^{\dagger} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$
  - $\mathbf{A}^{\dagger}\mathbf{A} = \mathbf{I}$
- when $\mathbf{A}$ has full row rank
  - $\mathbf{A}^{\dagger} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$
  - $\mathbf{A}\mathbf{A}^{\dagger} = \mathbf{I}$

Orthogonal Projections

- With SVD, the orthogonal projections of $\mathbf{y}$ onto $\mathcal{R}(\mathbf{A})$ and $\mathcal{R}(\mathbf{A})^\perp$ are given by

$$\Pi_{\mathcal{R}(\mathbf{A})}(\mathbf{y}) = \mathbf{A}\mathbf{x}_{\mathrm{LS}} = \mathbf{A}\mathbf{A}^\dagger\mathbf{y} = \mathbf{U}_1\mathbf{U}_1^T\mathbf{y}$$
$$\Pi_{\mathcal{R}(\mathbf{A})^\perp}(\mathbf{y}) = \mathbf{y} - \mathbf{A}\mathbf{x}_{\mathrm{LS}} = (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{y} = \mathbf{U}_2\mathbf{U}_2^T\mathbf{y}$$

- The orthogonal projector and orthogonal complement projector of $\mathbf{A}$ are given by

$$\mathbf{P}_{\mathbf{A}} = \mathbf{U}_1\mathbf{U}_1^T, \qquad \mathbf{P}_{\mathbf{A}}^\perp = \mathbf{U}_2\mathbf{U}_2^T$$

- Properties:
  - $\mathbf{P}_{\mathbf{A}}$ is idempotent, i.e., $\mathbf{P}_{\mathbf{A}}\mathbf{P}_{\mathbf{A}} = \mathbf{P}_{\mathbf{A}}$
  - $\mathbf{P}_{\mathbf{A}}$ is symmetric
  - The eigenvalues of $\mathbf{P}_{\mathbf{A}}$ are either 0 or 1
  - $\mathcal{R}(\mathbf{P}_{\mathbf{A}}) = \mathcal{R}(\mathbf{A})$
  - The same properties above apply to $\mathbf{P}_{\mathbf{A}}^\perp$, and $\mathbf{I} = \mathbf{P}_{\mathbf{A}} + \mathbf{P}_{\mathbf{A}}^\perp$

Minimum 2-Norm Solution to Underdetermined Linear Systems

Consider solving the linear system $\mathbf{y} = \mathbf{A}\mathbf{x}$ with fat $\mathbf{A} \in \mathbb{R}^{m \times n}$, $m < n$

- This is an underdetermined linear system: more unknowns n than the number of equations m

Assume $\mathbf{A}$ has full row rank. We already know that $\mathbf{x} = \mathbf{A}^\dagger \mathbf{y} + \boldsymbol{\eta}$ for any $\boldsymbol{\eta} \in \mathcal{R}(\mathbf{V}_2)$ is a solution

Now discard $\boldsymbol{\eta}$ and take $\mathbf{x} = \mathbf{A}^\dagger \mathbf{y}$ as one particular solution. This is the *unique* minimum 2-norm solution to $\mathbf{y} = \mathbf{A}\mathbf{x}$, i.e., it uniquely solves

$$\min_{\mathbf{x} \in \mathbb{R}^n} \; \|\mathbf{x}\|_2^2 \qquad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x}$$