Lecture 5 - *More on the SVD*

OBJECTIVE:

We show how the SVD can provide low-rank approximations of matrices in terms of the 2-norm and Frobenius norm.

$\diamond$ CHANGING BASES

The SVD tells us that all matrices are diagonal if you know how to look at them.
i.e., with the proper bases for the domain and range spaces.

Let $\mathbf{b} \in \mathbb{R}^m$ be expanded in the basis of left singular vectors of $\mathbf{A}$ (columns of $\mathbf{U}$).

Let $\mathbf{x} \in \mathbb{R}^n$ be expanded in the basis of the right singular vectors of $\mathbf{A}$ (columns of $\mathbf{V}$).

i.e.,

$$\mathbf{b}' = \mathbf{U}^T\mathbf{b}, \qquad \mathbf{x}' = \mathbf{V}^T\mathbf{x}$$

So now,

$$\begin{aligned}\mathbf{b} = \mathbf{A}\mathbf{x} \quad &\Leftrightarrow \quad \mathbf{U}^T\mathbf{b} = \mathbf{U}^T\mathbf{A}\mathbf{x} = \mathbf{U}^T\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\mathbf{x} \\ &\Leftrightarrow \quad \mathbf{b}' = \boldsymbol{\Sigma}\mathbf{x}'\end{aligned}$$

So viewing $\mathbf{b} = \mathbf{A}\mathbf{x}$ in the right-hand bases gives $\mathbf{b}' = \boldsymbol{\Sigma}\mathbf{x}'$
i.e., $\mathbf{A}$ reduces to the diagonal matrix $\boldsymbol{\Sigma}$.

$\diamond$ SVD and Eigenvalue Decomposition

If $\mathbf{A} \in \mathbb{R}^{m\times m}$ is nondefective, then it has a complete set of m linearly independent eigenvectors.
If we put the eigenvalues of $\mathbf{A}$ into a diagonal matrix $\boldsymbol{\Lambda}$ and gather the eigenvectors into a matrix $\mathbf{X}$, then the eigenvalue decomposition of $\mathbf{A}$ is given by

$$\mathbf{A} = \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^{-1}.$$

In this case, if $\mathbf{b} = \mathbf{A}\mathbf{x}$ and we expand $\mathbf{b}$, $\mathbf{x} \in \mathbb{R}^m$ as $\mathbf{b}' = \mathbf{X}^{-1}\mathbf{b}$, $\mathbf{x}' = \mathbf{X}^{-1}\mathbf{x}$.
Then

$$\mathbf{b}' = \mathbf{\Lambda}\mathbf{x}'$$

However, what looks similar at first glance has fundamental differences:

1. SVD uses two different bases (which ones?) whereas eigenvalue decomposition uses only one.

2. SVD uses orthonormal bases; $\mathbf{X}$ is generally not orthogonal.

3. SVD exists for all matrices, no matter what m and n, no matter if matrix is defective. Eigenvalue decomposition exists only for square, nondefective matrices.

◇ MATRIX PROPERTIES VIA SVD

The power of SVD comes from all the information that can be gleaned from it:
For the following, assume $\mathbf{A} \in \mathbb{R}^{m\times n}$.
Let $p = \min(m, n)$, $r \leq p$ be the number of nonzero singular values of $\mathbf{A}$.

Theorem 1. rank$(\mathbf{A}) = r =$ *the number of nonzero singular values.*

Theorem 2.

$$\begin{aligned} \mathrm{range}(\mathbf{A}) &= \mathrm{span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r) \\ \mathrm{null}(\mathbf{A}) &= \mathrm{span}(\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \ldots, \mathbf{v}_n) \end{aligned}$$

Theorem 3.

$$\begin{aligned} \|\mathbf{A}\|_2 &= \sigma_1 \\ \|\mathbf{A}\|_F &= \sqrt{\sigma_1^2 + \sigma_2^2 + \ldots + \sigma_r^2} \end{aligned}$$

Theorem 4. *The nonzero singular values of* $\mathbf{A}$ *are the (positive) square roots of the nonzero eigenvalues of* $\mathbf{A}^T\mathbf{A}$ *or* $\mathbf{A}\mathbf{A}^T$ *(these matrices have the same nonzero eigenvalues).*

Theorem 5. *If* $\mathbf{A}$ *is symmetric, the singular values of* $\mathbf{A}$ *are the absolute value of the eigenvalues of* $\mathbf{A}$*.*

Theorem 6. *For* $\mathbf{A} \in \mathbb{R}^{m\times m}$, $|\det(\mathbf{A})| = \prod_{i=1}^{m} \sigma_i$.

◇ LOW-RANK APPROXIMATIONS

Another way to understand the SVD is to consider how a matrix may be represented by a sum of rank-one matrices.

Theorem 7.

$$\mathbf{A} = \sum_{j=1}^{r} \sigma_j \mathbf{u}_j \mathbf{v}_j^T$$

i.e., $\mathbf{A}$ *is the sum of* r *rank-one matrices.*

What is so useful about this expansion is that the ν^{th} *partial sum captures as much of the "energy" of* $\mathbf{A}$ *as possible by a matrix of at most rank-*ν.

In this case, "energy" is defined by the 2-norm or the Frobenius norm.

Theorem 8. *For any* ν *with* $0 \leq \nu \leq r$ *define*

$$\mathbf{A}_\nu = \sum_{j=1}^{\nu} \sigma_j \mathbf{u}_j \mathbf{v}_j^T$$

If $\nu = p = \min(m, n)$, *define* $\sigma_{\nu+1} = 0$.
Then,

$$\|\mathbf{A} - \mathbf{A}_\nu\|_2 = \min_{\substack{\mathbf{B} \in \mathbb{R}^{m \times n} \\ \mathrm{rank}(\mathbf{B}) \leq \nu}} \|\mathbf{A} - \mathbf{B}\|_2 = \sigma_{\nu+1}$$

Geometric interpretation:

What is the best approximation of a hyperellipsoid by a line segment?
$\longrightarrow$ the line segment corresponding to the longest axis.

What is the best approximation by a 2-D ellipse?
$\longrightarrow$ the ellipse made up of the two longest axes.

etc.
$\longrightarrow$ we build up better approximations by taking more and more axes of less and less importance.

Theorem 9. *For any ν such that $0 \leq \nu \leq r$, $\mathbf{A}_\nu$ also satisfies*

$$\begin{aligned} \|\mathbf{A} - \mathbf{A}_\nu\|_F &= \min_{\substack{\mathbf{B} \in \mathbb{R}^{m \times n} \\ \operatorname{rank}(\mathbf{B}) \leq \nu}} \|\mathbf{A} - \mathbf{B}\|_F \\ &= \sqrt{\sigma_{\nu+1}^2 + \ldots + \sigma_r^2} \end{aligned}$$

The idea of successive approximation has application to image compression and information retrieval.

$\diamond$ COMPUTATION OF THE SVD

Alas, this topic will not be covered here, but we will cover the basic algorithms required (Householder bidiagonalization and shifted QR iteration).
So, with a bit of help (see Golub and Van Loan or Sarah's thesis) you'll even be able to program this nontrivial algorithm.

The SVD is the best (most accurate) way to determine the following quantities:

1. $\mathrm{rank}(\mathbf{A})$ (see Theorem 5.1)

2. orthonormal basis, $\mathrm{range}(\mathbf{A})$, $\mathrm{null}(\mathbf{A})$ (see Theorem 5.2)

3. $\|\mathbf{A}\|_2$ (see Theorem 5.3)

4. optimal low-rank approximations to $\mathbf{A}$ in $\|\cdot\|_2$ or $\|\cdot\|_F$ (see Theorems 5.8, 5.9)

Also, least-squares problems, intersection of subspaces, regularization, etc.