Applications (This lecture and next 2-3)

- Numerical linear algebra: rank, condition number, bases for fundamental subspaces, pseudoinverse

- Compression: Principle Component Analysis.
- Recommender systems
- Ranking sports teams.

Topics

- Singular values and singular value decomposition.

Notes based on ALA 8.7 and LAA 7.4 .

Warmup

The diagonalization theorems we've seen for complete and symmetric matrices have played a role in many interesting applications. Unfortunately, not all matrices can be factored as $A = PDP^{-1}$ for a diagonal matrix D; for example, such a factorization makes no sense if A is not square! Fortunately, a factorization $A = PDQ^{-1}$ is possible for any $m \times n$ matrix A! A special factorization of this type, called the singular value decomposition, is one of the most useful and widely applicable matrix factorizations in linear algebra.

The singular value decomposition is based on the following key property of matrix diagonalization which we'll show can be captured in general rectangular matrices:

Key observation: The absolute values of the eigenvalues of a symmetric matrix A measure the amounts that A stretches or shrinks certain vectors (the eigenvectors). If $A\underline{x} = \lambda \underline{x}$ and $\|\underline{x}\| = 1$, then

$$\|A\underline{x}\| = \lambda\|\underline{x}\| = |\lambda|\|\underline{x}\| = |\lambda|.$$

If λ_1 is the eigenvalue with the greatest magnitude, i.e., if $|\lambda_1| \geq |\lambda_i|$ for $i = 1, \ldots, n$, then a corresponding unit eigenvector $\underline{v}_1$ identifies the direction in which stretching is greatest. That is, the length of $A\underline{x}$ is maximized when $\underline{x} = \underline{v}_1$, and $\|A\underline{v}_1\| = |\lambda_1|$.

This description is reminiscent of the optimization principle we saw for characterizing eigenvalues of symmetric matrices, albeit with a focus on maximizing length $\|A\underline{x}\|$ rather than the quadratic form $\underline{x}^T A \underline{x}$. What we'll see next is that this description of $\underline{v}_1$ and $|\lambda_1|$ has an analogue for rectangular matrices that will lead to the singular value decomposition.

Example: The matrix $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$ defines a linear map $\underline{x} \mapsto A\underline{x}$ from $\mathbb{R}^3$ to $\mathbb{R}^2$. If we consider the effects of this map on the unit sphere $\{\underline{x} \in \mathbb{R}^3 \mid \|\underline{x}\| = 1\}$, we observe that multiplication by A transforms this sphere in $\mathbb{R}^3$ into a ellipse in $\mathbb{R}^2$:

(LAA Ch 7.4)

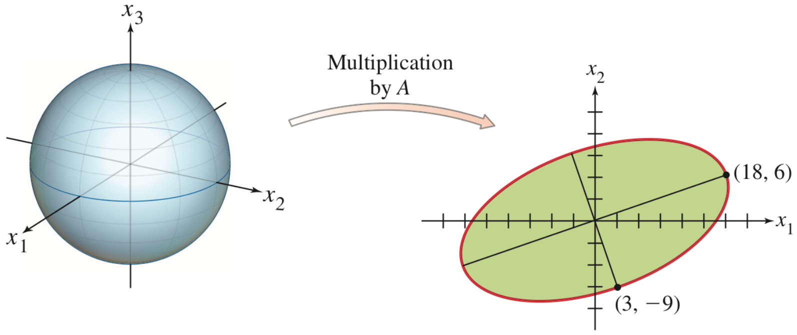

FIGURE 1 A transformation from $\mathbb{R}^3$ to $\mathbb{R}^2$.

Our task is to find a unit vector $\underline{x}$ at which the length $\|A\underline{x}\|$ is maximized, and compute this maximum length. That is, we want to solve the optimization problem

$$\text{maximize } \|A\underline{x}\|$$

over choices of $\underline{x}$ satisfying $\|\underline{x}\|=1$. Our first observation is that the quantity $\|A\underline{x}\|^2$ is maximized by the same $\underline{x}$ that maximizes $\|A\underline{x}\|$, but that $\|A\underline{x}\|^2$ is easier to work with. Specifically, note that

$$\|A\underline{x}\|^2 = \langle A\underline{x}, A\underline{x}\rangle = (A\underline{x})^T(A\underline{x}) = \underline{x}^TA^TA\underline{x} = \underline{x}^T(A^TA)\underline{x}.$$

So our task is to now find a unit vector $\|\underline{x}\|=1$ that maximizes the quadratic form $\underline{x}^T(A^TA)\underline{x}$ defined by the symmetric (positive semidefinite) matrix A^TA: we know how to do this. By our theorem characterizing eigenvalues of symmetric matrices from an optimization perspective, we know the maximum value is the largest eigenvalue λ_1 of the matrix A^TA, and is attained at the unit eigenvector $\underline{v}_1$ of A^TA corresponding to λ_1.

For the matrix in this example:

$$A^TA = \begin{bmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{bmatrix}\begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} = \begin{bmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{bmatrix}$$

and the eigenvalue/vector pairs are:

$$\lambda_1 = 360,\ \underline{v}_1 = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, \quad \lambda_2 = 90,\ \underline{v}_2 = \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix}, \quad \lambda_3 = 0,\ \underline{v}_3 = \begin{bmatrix} 2/3 \\ -2/3 \\ 1/3 \end{bmatrix}.$$

The maximum value of $\underline{x}^T(A^TA)\underline{x} = \|A\underline{x}\|^2$ is thus $\lambda_1 = 360$, and attained when $\underline{x} = \underline{v}_1$. The vector $A\underline{v}_1$ is a point on the ellipse in Fig 1 above farthest from the origin, namely

$$A\underline{v}_1 = \begin{bmatrix} 4 & 4 & 14 \\ 8 & 7 & -2 \end{bmatrix}\begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 18 \\ 6 \end{bmatrix}.$$

For $\|\underline{x}\|=1$, the maximum value of $\|A\underline{x}\|$ is $\|A\underline{v}_1\| = \sqrt{360} = 6\sqrt{10}$.

This example suggests that the effect of a matrix A on the unit sphere in $\mathbb{R}^3$ is related to the quadratic form $\underline{x}^TA^TA\underline{x}$. What we'll see next is that the entire geometric behavior of the map $\underline{x} \mapsto A\underline{x}$ is captured by this quadratic form.

The Singular Values of an $m \times n$ Matrix

Consider an $m \times n$ real matrix $A \in \mathbb{R}^{m \times n}$. Then $A^T A$ is an $n \times n$ symmetric matrix, and can be orthogonally diagonalized. Let $V = [v_1 \cdots v_n]$ be an orthogonal matrix composed of orthonormal eigenvectors of $A^T A$, and let $\lambda_1, \ldots, \lambda_n$ be the associated eigenvalues of $A^T A$. Then for $i = 1, \ldots, n$,

$$\begin{aligned}
\|A \underline{v}_i\|^2 &= (A \underline{v}_i)^T (A \underline{v}_i) = \underline{v}_i^T (A^T A \underline{v}_i) \\
&= \underline{v}_i^T (\lambda_i \underline{v}_i) \\
&= \lambda_i \underline{v}_i^T \underline{v}_i = \lambda_i \|\underline{v}_i\|^2 \\
&= \lambda_i .
\end{aligned}$$

This tells us that all of the eigenvalues $\lambda_i = \|A \underline{v}_i\|^2 \geq 0$, since norms can only take on nonnegative values, i.e., $A^T A$ is a positive semidefinite matrix. Let's assume that we've ordered our eigenvalues in decreasing order:

$$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n \geq 0.$$

The singular values of A are the positive square roots of the nonzero eigenvalues $\lambda_i > 0$ of $A^T A$, denoted σ_i. That is, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$, and $\lambda_{r+1} = \lambda_{r+2} = \cdots = \lambda_n = 0$ be a partition of the eigenvalues such that $\lambda_i > 0$ for $i = 1, \ldots, r$, and $\lambda_i = 0$ for $i = r+1, \ldots, n$. Then A has r singular values, defined as

$$\sigma_i = \sqrt{\lambda_i}, \quad i = 1, \ldots, r.$$

WARNING: Some texts include the zero eigenvalues $\lambda_{r+1}, \ldots, \lambda_n$ of $A^T A$ a singular values of A. This is simply a different convention, and is mathematically equivalent. However, we find our definition to be more natural for our purposes.

Example: Using the same $A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$ as the previous example, we have $\sigma_1 = \sqrt{360} = 6\sqrt{10}$, $\sigma_2 = \sqrt{90} = 3\sqrt{10}$. In this case, A only has two singular values as $\lambda_3 = 0$. For this example, $r = 2$, and $\lambda_1 = 360 > \lambda_2 = 90 > \lambda_3 = 0$. From the previous example, the first singular value of A is the maximum of $\|A\underline{x}\|$ over all $\|\underline{x}\| = 1$, attained at v_1. Our optimization based characterization of eigenvalues of symmetric matrices tells us that the second singular value of A is the maximum of $\|A\underline{x}\|$ over all unit vectors orthogonal to $\underline{v}_1$: this is attained by the second eigenvector u_2 of $A^T A$. For u_2 from the previous example,

$$A u_2 = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} \begin{bmatrix} -2/3 \\ -1/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 3 \\ -9 \end{bmatrix}.$$

This point is on the minor axis of the ellipse in Fig. 1 above, just as $A u_1$ is on the major axis (see Fig 2 below). The two singular values of A are the lengths of the major and minor semiaxes of the ellipse.

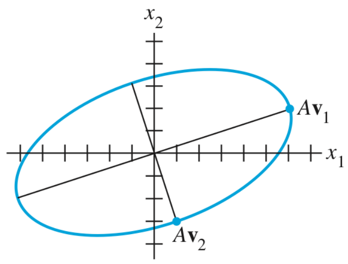

FIGURE 2

That $A\underline{v}_1$ and $A\underline{v}_2$ are orthogonal is no accident, as the next theorem shows.

Theorem: Suppose that $\underline{u}_1, \ldots, \underline{u}_n$ is an orthonormal basis for $\mathbb{R}^n$ composed of the eigenvectors of A^TA, ordered so that the corresponding eigenvalues of A^TA satisfy

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0,$$

where r denotes the number of nonzero eigenvalues of A^TA, i.e., the number of singular values $\sigma_i = \sqrt{\lambda_i} > 0$, $i = 1, \ldots, r$, of A. Then, $A\underline{u}_1, \ldots, A\underline{u}_r$ is an orthogonal basis for $\mathrm{Col}(A)$, and $\mathrm{rank}(A) = r$.

Proof: Because $\underline{v}_i$ and $\lambda_j \underline{v}_j$ are orthogonal for $i \neq j$,

$$(A\underline{v}_i)^T(A\underline{v}_j) = \underline{v}_i^T A^T A \underline{v}_j = \underline{v}_i^T \lambda_j \underline{v}_j = 0.$$

Thus, $A\underline{v}_1, \ldots, A\underline{v}_r$ are mutually orthogonal, and hence linearly independent. They are also clearly contained in $\mathrm{Col}(A)$. Now, for any $\underline{y} \in \mathrm{Col}(A)$, there must be an $\underline{x} \in \mathbb{R}^n$ such that $\underline{y} = A\underline{x}$. Expanding $\underline{x}$ in the basis $\underline{v}_1, \ldots, \underline{v}_n$, as $\underline{x} = c_1\underline{v}_1 + \cdots + c_n\underline{v}_n$ for some $c_1, \ldots, c_n \in \mathbb{R}$, we have:

$$\underline{y} = A\underline{x} = A(c_1\underline{v}_1 + \cdots + c_n\underline{v}_n) = c_1 A\underline{v}_1 + \cdots + c_r A\underline{v}_r + c_{r+1} A\underline{v}_{r+1} + \cdots + c_n A\underline{v}_n$$

$$= c_1 A\underline{v}_1 + \cdots + c_r A\underline{v}_r.$$

We used that $\|A\underline{v}_i\|^2 = \lambda_i = 0$ for $i = r+1, \ldots, n \iff A\underline{v}_i = 0$ for $i = r+1, \ldots, n$ in the last equality.

Therefore, we have that $\underline{y} \in \mathrm{span}\{A\underline{v}_1, \ldots, A\underline{v}_r\}$. Thus $A\underline{v}_1, \ldots, A\underline{v}_r$ is both linearly independent and a spanning set for $\mathrm{Col}(A)$; meaning it is an orthogonal basis for $\mathrm{Col}(A)$. Hence, by the Fundamental Theorem of Linear Algebra:

$$\mathrm{rank}(A) = \dim \mathrm{Col}(A) = r.$$

Numerical Note: In certain cases, the rank of A may be very sensitive to small changes in the entries of A. The obvious approach of counting the # of pivot columns in A does not work well if A is row reduced by a computer, as roundoff errors often create a row echelon form with full rank. In practice, the most reliable way of computing the rank of a large matrix A is to count the number of singular values larger than a small threshold ε (typically on the order of 10^{-12}, but can vary depend on applications). In this case, singular values smaller than ε are treated as zeros for all practical purposes, and the effective rank of A is computed by counting the remaining nonzero singular values.

The Singular Value Decomposition

The decomposition of A involves an $r \times r$ diagonal matrix Σ of the form

$$\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_r).$$

We note that because $r = \dim \operatorname{Col}(A) = \dim \operatorname{Row}(A)$ by the FTLA, we must have that $r \leq \min\{m, n\}$ if $A \in \mathbb{R}^{m \times n}$.

Theorem: Let $A \in \mathbb{R}^{m \times n}$ be an $m \times n$ matrix of rank $r > 0$. Then A can be factored as

$$A = U \Sigma V^T,$$

where $U \in \mathbb{R}^{m \times r}$ has orthonormal columns, so $U^T U = I_r$, $\Sigma = \operatorname{diag}(\sigma_1, \dots, \sigma_r)$ is a diagonal matrix with the singular values of A σ_i along the diagonal, and $V \in \mathbb{R}^{n \times r}$ has orthonormal columns, so $V^T V = I_r$.

Such a factorization of A is called its singular value decomposition, and the columns of U are called the left singular vectors of A, while the columns of V are called the right singular vectors of A.

Proof: Let λ_i and $\underline{v}_i$ be the eigenvalues/vectors of $A^T A$ as described previously, so that $A\underline{v}_1, \dots, A\underline{v}_r$ is an orthogonal basis for $\operatorname{col}(A)$. Normalize each $A\underline{v}_i$ to form an orthonormal basis for $\operatorname{col}(A)$:

$$\underline{u}_i = \frac{1}{\|A\underline{v}_i\|} A\underline{v}_i = \frac{1}{\sigma_i} A\underline{v}_i$$

and hence $A\underline{v}_i = \sigma_i \underline{u}_i$ for $i = 1, \dots, r$. Define the matrices

$$U = [\underline{u}_1 \cdots \underline{u}_r] \in \mathbb{R}^{m \times r} \quad \text{and} \quad V = [\underline{v}_1 \cdots \underline{v}_r] \in \mathbb{R}^{n \times r}$$

By construction, the columns of U are orthonormal: $U^T U = I_r$, and similarly for the columns of V: $V^T V = I_r$.

Let's define the following "full" matrices:

$$\hat{U} = [U \;\; U^\perp] \in \mathbb{R}^{m\times m} \quad \text{and} \quad \hat{V} = [V \;\; V^\perp] \in \mathbb{R}^{n\times n}.$$

Here, $V^\perp = [\underline{v}_{r+1} \cdots \underline{v}_n]$ has orthonormal columns spanning the orthogonal complement of $\operatorname{span}\{\underline{v}_1, \dots, \underline{v}_r\}$, so that the columns of $\hat{V}$ form an orthonormal basis of $\mathbb{R}^n$.

Similarly, let $U^\perp$ have orthonormal columns spanning the orthogonal complement of $\operatorname{span}\{\underline{u}_1, \dots, \underline{u}_r\}$, so the columns of $\hat{U}$ form an orthonormal basis for $\mathbb{R}^m$.

Finally, define $\hat{\Sigma} = \begin{array}{cc} \overbrace{}^{r} & \overbrace{}^{n-r} \\ \end{array}$ $\begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \begin{matrix} \}r \\ \}m-r \end{matrix}$. We first show that

$A = \hat{U}\hat{\Sigma}\hat{V}^T$, or equivalently (since $\hat{V}$ is orthogonal), $A\hat{V} = \hat{U}\hat{\Sigma}$. First,

$$A\hat{V} = [A\underline{v}_1 \cdots A\underline{v}_r \;\; A\underline{v}_{r+1} \cdots A\underline{v}_n] = [\sigma_1\underline{u}_1 \cdots \sigma_r\underline{u}_r \;\; \underline{0} \cdots \underline{0}],$$

Then, notice:

$$\hat{U}\hat{\Sigma} = [\underline{u}_1 \cdots \underline{u}_r \;\; \underline{u}_{r+1} \cdots \underline{u}_m] \left[\begin{array}{cccc|ccc} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & & \vdots & & \vdots \\ \vdots & & \ddots & 0 & & & \\ 0 & 0 & & \sigma_r & 0 & \cdots & 0 \\ \hline 0 & & \cdots & 0 & & & \\ \vdots & & & & & & \\ 0 & & \cdots & 0 & & & \end{array}\right] \begin{matrix} \left.\vphantom{\begin{matrix}a\\a\\a\\a\end{matrix}}\right\} r \\ \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} m-r \end{matrix}$$

$$= [\sigma\underline{u}_1 \cdots \sigma\underline{u}_r \;\; \underline{0} \cdots \underline{0}].$$

So that $A\hat{V} = \hat{U}\hat{\Sigma}$, or equivalently, $A = \hat{U}\hat{\Sigma}\hat{V}^T$. But, now, notice:

$$A = \hat{U}\hat{\Sigma}\hat{V}^T = [U \;\; U^\perp]\begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} V^T \\ (V^\perp)^T \end{bmatrix} = U\Sigma V^T,$$ proving our result.

NOTE: Some textbooks define the singular value decomposition of A as $A = \hat{U}\hat{\Sigma}\hat{V}^T$ — this is necessary when allowing for singular values equal to zero. When only considering nonzero singular values, as we do, $A = U\Sigma V^T$ is the appropriate definition of the SVD. This is sometimes called the compact SVD of A, but we will just call it the SVD.

Example: Let's use the results of the previous examples to construct the SVD of
$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}.$$

Step 1: Find an orthogonal diagonalization of A^TA. In general, for A with many columns, this is done numerically; here we use the data from before

$$A^TA = \hat{V}\Lambda\hat{V}^T = [\underline{v}_1\ \underline{v}_2\ \underline{v}_3]\begin{bmatrix}\lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3\end{bmatrix}\begin{bmatrix}\underline{v}_1^T \\ \underline{v}_2^T \\ \underline{v}_3^T\end{bmatrix}$$

with $(\lambda_i, \underline{v}_i)$ as specified above. with $\lambda_1 = 360$, $\lambda_2 = 90$, $\lambda_3 = 0$.

Step 2: Set up V and Σ: Arrange the nonzero eigenvalues of A^TA in decreasing order and compute the singular values. For this example:

$$\sigma_1 = 6\sqrt{10} \quad \text{and} \quad \sigma_2 = 3\sqrt{10},$$

and

$$\Sigma = \operatorname{diag}(\sigma_1, \sigma_2) = \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10}\end{bmatrix}.$$ Hence $\operatorname{rank} A = 2$, and $V \in \mathbb{R}^{3\times 2}$

The corresponding eigenvectors define the columns of V:

$$V = [\underline{v}_1\ \underline{v}_2] = \begin{bmatrix} 1/3 & -2/3 \\ 2/3 & -1/3 \\ 2/3 & 2/3\end{bmatrix}.$$

Step 3: Construct U: Since $\operatorname{rank} A = 2$, $U \in \mathbb{R}^{2\times 2}$. The columns of U are given by the normalized vectors obtained from $A\underline{v}_1$ and $A\underline{v}_2$. Recall that we showed above that $\|A\underline{v}_1\| = \sigma_1$ and $\|A\underline{v}_2\| = \sigma_2$, so $U = [\underline{u}_1, \underline{u}_2]$ with

$$\underline{u}_1 = \frac{A\underline{v}_1}{\sigma_1} = \frac{1}{6\sqrt{10}}\begin{bmatrix}18 \\ 6\end{bmatrix} = \begin{bmatrix}3/\sqrt{10} \\ 1/\sqrt{10}\end{bmatrix} \quad \text{and}$$

$$\underline{u}_2 = \frac{A\underline{v}_2}{\sigma_2} = \frac{1}{3\sqrt{10}}\begin{bmatrix}3 \\ -9\end{bmatrix} = \begin{bmatrix}1/\sqrt{10} \\ -3/\sqrt{10}\end{bmatrix}.$$

Finally, the SVD of A is:

$$A = \underbrace{\begin{bmatrix}3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10}\end{bmatrix}}_{U}\underbrace{\begin{bmatrix}6\sqrt{10} & 0 \\ 0 & 3\sqrt{10}\end{bmatrix}}_{\Sigma}\underbrace{\begin{bmatrix}1/3 & 2/3 & 2/3 \\ -2/3 & -1/3 & 2/3\end{bmatrix}}_{V^T}.$$

You can check that indeed $A = U\Sigma V^T$ here, and that $U^TU = V^TV = I_2$.

ONLINE NOTES: Please add example 4 from LAA 7.4, suitably modified to use the compact SVD form.

Linear Algebra Applications of the SVD

The next few classes will focus on engineering and AI applications of the SVD. For now, we highlight some more technical linear algebraic applications: these are all immensely important from a practical perspective, and form subroutines for most real-world applications of linear algebra.

<u>The Condition Number</u> Most numerical calculators that require solving a linear equation $A\underline{x}=\underline{b}$ are as reliable as possible when the SVD of A is used. Since the matrices U and V have orthonormal columns, they do not affect lengths or vectors between vectors. For example, for $U \in \mathbb{R}^{m\times r}$, we have:

$$\langle U\underline{x}, U\underline{y}\rangle = \underline{x}^T U^T U \underline{y} = \underline{x}^T\underline{y} = \langle \underline{x}, \underline{y}\rangle$$

for any $\underline{x}, \underline{y} \in \mathbb{R}^r$. Therefore, any numerical issues that arise will be due to the diagonal entries of Σ, i.e., due to the singular values of A.

In particular, if some of singular values are much larger than others, this means certain directions are stretched out much more than others, which can lead to roundoff errors. A natural way to quantify this notion is using the singular values of A. If $A \in \mathbb{R}^{n\times n}$ is an $n\times n$ invertible matrix, so that $r = \text{rank}(A) = n$, we define the condition number of A to be the ratio $\kappa(A) = \frac{\sigma_1}{\sigma_n}$ of the largest to smallest singular values of A. If A is not invertible, it is convention to set $\kappa(A)=\infty$, although the ratio σ_1/σ_r is a useful measure of the numerical stability of computing with a rectangular matrix $A\in\mathbb{R}^{m\times n}$.

ONLINE NOTES: Using numpy, show that ill-conditioned A can lead to bad solutions to $A\underline{x}=\underline{b}$ even if A is invertible. Case 1: assume we use $\hat{\underline{b}} = \underline{b} + \underline{n}$ for $\underline{n}$ some small measurement noise. Case 2: just make A super ill conditioned and show that $\underline{x} = A^{-1}\underline{b}$ computed using numpy doesn't actually satisfy $A\underline{x}=\underline{b}$.

Computing Bases of Fundamental Subspaces

Given an SVD for an $m\times n$ matrix $A\in\mathbb{R}^{m\times n}$ with $\text{rank}(A)=r$, let $\underline{u}_1,\dots,\underline{u}_r$ be the left singular vectors, $\underline{v}_1,\dots,\underline{v}_r$ the right singular vectors, and $\sigma_1,\dots,\sigma_r$ the singular values.

Recall that we showed that $\underline{u}_1,\dots,\underline{u}_r$ forms a basis for Col(A). Let $\underline{u}_{r+1},\dots,\underline{u}_m$ be an orthonormal basis for $\text{Col}(A)^\perp$ so that $\underline{u}_1,\dots,\underline{u}_m$ form a basis for $\mathbb{R}^m$, computed for example using the Gram-Schmidt Process. Then, by the FTLA, we have that $\text{Col}(A)^\perp = \text{Null}(A^T) = \text{span}\{\underline{u}_{r+1},\dots,\underline{u}_m\}$, i.e., these vectors form an orthonormal basis for $\text{Null}(A^T) = \text{LNull}(A)$.

Next, recall that $\underline{v}_1,\dots,\underline{v}_r, \underline{v}_{r+1},\dots,\underline{v}_n$, the eigenvectors of A^TA, form an orthonormal

basis of $\mathbb{R}^n$. Since $A\underline{v}_i = 0$ for $i = r+1, \dots, n$, the vectors $\underline{v}_{r+1}, \dots, \underline{v}_n$ span a subspace of $\text{Null}(A)$ of dimension $n-r$. But, by the FTLA, $\dim \text{Null}(A) = n - \text{rank}(A)$
$= n - r$.

Therefore, $\underline{v}_{r+1}, \dots, \underline{v}_n$ are an orthonormal basis for $\text{Null}(A)$.

Finally, $\text{Null}(A)^\perp = \text{Col}(A^T) = \text{Row}(A)$. But $\text{Null}(A)^\perp = \text{span}\{\underline{v}_1, \dots, \underline{v}_r\}$ since the $\underline{v}_i$ are an orthonormal basis for $\mathbb{R}^n$, and thus $\underline{v}_1, \dots, \underline{v}_r$ are an orthonormal basis for $\text{Row}(A)$.

Summarizing, we have:

- $\text{Col}(A) = \text{span}\{\underline{u}_1, \dots, \underline{u}_r\}$
- $\text{Col}(A)^\perp = \text{Null}(A^T) = \text{LNull}(A) = \text{span}\{\underline{u}_{r+1}, \dots, \underline{u}_m\}$
- $\text{Col}(A^T) = \text{Row}(A) = \text{span}\{\underline{v}_1, \dots, \underline{v}_r\}$
- $\text{Col}(A^T)^\perp = \text{Null}(A) = \text{span}\{\underline{v}_{r+1}, \dots, \underline{v}_n\}$

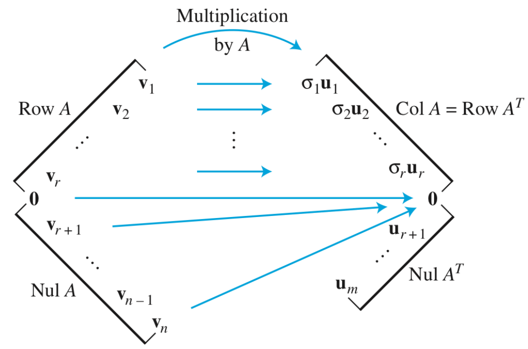

FIGURE 4 The four fundamental subspaces and the action of A.

Specializing these observations to square matrices, we have the following theorem characterizing invertible matrices:

<u>Theorem</u>: The following statements are equivalent for a square $n \times n$ matrix $A \in \mathbb{R}^{n\times n}$:

i) $\text{Col}(A)^\perp = \text{Null}(A^T) = \{0\}$
ii) $\text{Null}(A)^\perp = \text{Col}(A^T) = \text{Row}(A) = \mathbb{R}^n$
iii) $\text{Col}(A^T)^\perp = \text{Null}(A) = \{0\}$
iv) $\text{Null}(A^T)^\perp = \text{Col}(A) = \mathbb{R}^n$
v) A has rank $= n$
vi) A has n (nonzero) singular values

<u>The Pseudoinverse of A</u>

Recall the least squares problem of finding a vector $\underline{x}$ that minimizes the objective $\|A\underline{x} - b\|^2$. We saw that the least squares solution is given by the solution to the normal equations

$$A^TA\hat{\underline{x}} = A^T\underline{b}. \quad (NE)$$

Let's rewrite (NE) using the SVD $A = U\Sigma V^T$, $A^T = V\Sigma^T U^T = V\Sigma U^T$ $(\Sigma = \Sigma^T)$

$$A^TA\hat{\underline{x}} = V\Sigma \underbrace{U^TU}_{I}\Sigma V^T\hat{\underline{x}} = \underbrace{V\Sigma^2V^T\hat{\underline{x}}}_{(a)} = \underbrace{V\Sigma U^T\underline{b}}_{(b)}$$

Let's start by left multiplying (a) and (b) by V^T to take advantage of $V^TV = I_r$.

$$V^T(V\Sigma^2V^T\underline{f}) = V^T(V\Sigma U^T\underline{b}) \Rightarrow \Sigma^2 V^T\underline{f} = \Sigma U^T\underline{b}.$$

Now, let's isolate $V^T\underline{f}$ by multiplying both sides by Σ^{-2}:

$$V^T\underline{f} = \Sigma^{-1}U^T\underline{b}. \quad (*)$$

Finally, note that $\underline{f}$ satisfies $(*)$ iff $\hat{\underline{f}} = V\Sigma^{-1}U^T\underline{b} + \underline{n}$ (again since $V^TV = I$) for any $\underline{n} \in \mathrm{Null}(V^T) = \mathrm{Col}(V)^\perp$. The special solution $\underline{x}^* = V\Sigma^{-1}U^T\underline{b}$ can be shown to be the minimum norm least squares solution when several $\underline{f}$ exist such that $A\underline{f} = \underline{b}$. The matrix

$$A^+ = V\Sigma^{-1}U^T$$

is called the pseudoinverse of A, and is also known as the Moore-Penrose inverse of A.

If we look at $A\underline{x}^* = AA^+\underline{b}$, we observe that:

$$A\underline{x}^* = U\Sigma \underbrace{V^TV}_{I}\Sigma^{-1}U^T\underline{b} = UU^T\underline{b},$$

i.e., $A\underline{x}^*$ is the orthogonal projection $\hat{\underline{b}}$ of $\underline{b}$ onto $\mathrm{Col}(A)$.

ONLINE NOTES: Please add the Practice Problems on p.471 of LAA at the end of Ch. 7.4 + Solutions