EE731 Lecture Notes: Matrix Computations for Signal Processing

James P. Reilly©
Department of Electrical and Computer Engineering
McMaster University

October 17, 2005

Lecture 3

3 The Singular Value Decomposition (SVD)

In this lecture we learn about one of the most fundamental and important matrix decompositions of linear algebra: the SVD. It bears some similarity with the eigendecomposition (ED), but is more general. Usually, the ED is of interest only on symmetric square matrices, but the SVD may be applied to *any* matrix. The SVD gives us important information about the rank, the column and row spaces of the matrix, and leads to very useful solutions and interpretations of least squares problems. We also discuss the concept of *matrix projectors*, and their relationship with the SVD.

3.1 The Singular Value Decomposition (SVD)

We have found so far that the eigendecomposition is a useful analytic tool. However, it is only applicable on *square symmetric* matrices. We now consider the SVD, which may be considered a generalization of the ED to arbitrary matrices. Thus, with the SVD, all the analytical uses of the ED which before were restricted to symmetric matrices may now be applied to any form of matrix, regardless of size, whether it is symmetric or nonsymmetric, rank deficient, etc.

Theorem 1 *Let* $\mathbf{A} \in \Re^{m\times n}$. *Then* $\mathbf{A}$ *can be decomposed according to the singular value decomposition as*

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T \tag{1}$$

where $\mathbf{U}$ *and* $\mathbf{V}$ *are orthonormal and*

$$\mathbf{U} \in \Re^{m\times m}, \quad \mathbf{V} \in \Re^{n\times n}.$$

Let $p \stackrel{\Delta}{=} \min(m,n)$. *Then*

$$\boldsymbol{\Sigma} = \begin{matrix} p \\ m-p \end{matrix} \underset{\begin{matrix} p & n-p \end{matrix}}{\begin{bmatrix} \tilde{\boldsymbol{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}}$$

where $\tilde{\boldsymbol{\Sigma}} = diag(\sigma_1, \sigma_2, \ldots, \sigma_p)$ *and*

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \ldots \geq \sigma_p \geq 0.$$

The matrix $\boldsymbol{\Sigma}$ must be of dimension $\Re^{m\times n}$ (i.e., the same size as $\mathbf{A}$), to maintain dimensional consistency of the product in (1). It is therefore padded with zeros either on the bottom or to the right of the diagonal block, depending on whether $m > n$ or $m < n$, respectively.

Since $\mathbf{U}$ and $\mathbf{V}$ are orthonormal, we may also write (1) in the form:

$$\underset{m\times m}{\mathbf{U}^T} \; \underset{m\times n}{\mathbf{A}} \; \underset{n\times n}{\mathbf{V}} = \underset{m\times n}{\boldsymbol{\Sigma}} \tag{2}$$

where $\boldsymbol{\Sigma}$ is a diagonal matrix. The values σ_i which are defined to be positive, are referred to as the *singular values* of $\mathbf{A}$. The columns $\mathbf{u}_i$ and $\mathbf{v}_i$ of $\mathbf{U}$ and $\mathbf{V}$ are respectively called the left and right *singular vectors* of $\mathbf{A}$.

The SVD corresponding to (1) may be shown diagramatically in the following way:

$$\mathbf{A} = \underset{\substack{m\times m \\ \mathbf{U}}}{\begin{bmatrix} \vert & \vert & \vert & \vert \\ \vert & \vert & \vert & \vert \\ \vert & \vert & \vert & \vert \end{bmatrix}} \underset{\substack{m\times n \\ \mathbf{\Sigma}}}{\begin{bmatrix} \sigma_1 & & & & & \\ & \ddots & & \mathbf{0} & & \\ & & \sigma_p & & & \\ & & & 0 & & \\ & \mathbf{0} & & & \ddots & \\ & & & & & 0 \end{bmatrix}} \underset{\substack{n\times n \\ \mathbf{V}^T}}{\begin{bmatrix} \text{———} \\ \text{———} \\ \text{———} \\ \text{———} \end{bmatrix}} \tag{3}$$

Each line above represents a column of either $\mathbf{U}$ or $\mathbf{V}$.

3.2 Existence Proof of the SVD

Consider two vectors $\mathbf{x}$ and $\mathbf{y}$ where $||\mathbf{x}||_2 = ||\mathbf{y}||_2 = 1$, s.t. $\mathbf{Ax} = \sigma\mathbf{y}$, where $\sigma = ||\mathbf{A}||_2$. The fact that such vectors $\mathbf{x}$ and $\mathbf{y}$ can exist follows from the definition of the matrix 2-norm. We define orthonormal matrices $\mathbf{U}$ and $\mathbf{V}$ so that $\mathbf{x}$ and $\mathbf{y}$ form their first columns, as follows:

$$\begin{aligned} \mathbf{U} &= [\mathbf{y}, \mathbf{U}_1] \\ \mathbf{V} &= [\mathbf{x}, \mathbf{V}_1] \end{aligned}$$

That is, $\mathbf{U}_1$ consists of a set of non–unique orthonormal columns which are mutually orthogonal to themselves and to $\mathbf{y}$; similarly for $\mathbf{V}_1$.

We then define a matrix $\mathbf{A}_1$ as

$$\begin{aligned} \mathbf{U}^T\mathbf{AV} &= \mathbf{A}_1 \\ &= \begin{bmatrix} \mathbf{y}^T \\ \mathbf{U}_1{}^T \end{bmatrix} \mathbf{A}[\mathbf{x}, \mathbf{V}_1] \end{aligned} \tag{4}$$

The matrix $\mathbf{A}_1$ has the following structure:

$$\underbrace{\begin{bmatrix} \mathbf{y}^T \\ \mathbf{U}_1^T \end{bmatrix}}_{\text{orthonormal}} \quad \mathbf{A} \quad \underbrace{\begin{bmatrix} \mathbf{x} & \mathbf{V}_1 \end{bmatrix}}_{\text{orthonormal}} = \begin{bmatrix} \mathbf{y}^T \\ \mathbf{U}_1{}^T \end{bmatrix} \begin{bmatrix} \sigma\mathbf{y} & \mathbf{AV}_1 \end{bmatrix}$$

$$= \begin{array}{c} \scriptstyle \sigma y^T y \quad y^T A V_1 \\ \downarrow \quad\quad \downarrow \\ \left[\begin{array}{cc} \sigma & \mathbf{w}^T \\ \mathbf{0} & \mathbf{B} \end{array} \right] \\ \scriptstyle 1 \quad n-1 \end{array} \begin{array}{c} \scriptstyle 1 \\ \scriptstyle m-1 \end{array} \triangleq \mathbf{A}_1. \tag{5}$$

where $\mathbf{B} \triangleq \mathbf{U}_1^T \mathbf{A} \mathbf{V}_1$. The $\mathbf{0}$ in the (2,1) block above follows from the fact that $\mathbf{U}_1 \perp \mathbf{y}$, because $\mathbf{U}$ is orthonormal.

Now, we post-multiply both sides of (5) by the vector $\left[\begin{array}{c} \sigma \\ \mathbf{w} \end{array} \right]$ and take 2-norms:

$$\left|\left| \mathbf{A}_1 \left[\begin{array}{c} \sigma \\ \mathbf{w} \end{array} \right] \right|\right|_2^2 = \left|\left| \left[\begin{array}{cc} \sigma & \mathbf{w}^T \\ \mathbf{0} & \mathbf{B} \end{array} \right] \left[\begin{array}{c} \sigma \\ \mathbf{w} \end{array} \right] \right|\right|_2^2 \geq (\sigma^2 + \mathbf{w}^T \mathbf{w})^2. \tag{6}$$

This follows because the term on the extreme right is only the first element of the vector product of the middle term. But, as we have seen, matrix p-norms obey the following property:

$$||\mathbf{A}\mathbf{x}||_2 \leq ||\mathbf{A}||_2 \, ||\mathbf{x}||_2 \, . \tag{7}$$

Therefore using (6) and (7), we have

$$||\mathbf{A}_1||_2^2 \left|\left| \left[\begin{array}{c} \sigma \\ \mathbf{w} \end{array} \right] \right|\right|_2^2 \geq \left|\left| \mathbf{A}_1 \left[\begin{array}{c} \sigma \\ \mathbf{w} \end{array} \right] \right|\right|_2^2 \geq (\sigma^2 + \mathbf{w}^T \mathbf{w})^2. \tag{8}$$

Note that $\left|\left| \left[\begin{array}{c} \sigma \\ \mathbf{w} \end{array} \right] \right|\right|_2^2 = \sigma^2 + \mathbf{w}^T \mathbf{w}$. Dividing (8) by this quantity, we obtain

$$||\mathbf{A}_1||_2^2 \geq \sigma^2 + \mathbf{w}^T \mathbf{w}. \tag{9}$$

But, we defined $\sigma = ||\mathbf{A}||_2$. Therefore, the following must hold:

$$\sigma = ||\mathbf{A}||_2 = \left|\left| \mathbf{U}^T \mathbf{A} \mathbf{V} \right|\right|_2 = ||\mathbf{A}_1||_2 \tag{10}$$

where the equality on the right follows because the matrix 2-norm is invariant to matrix pre- and post-multiplication by an orthonormal matrix. By comparing (9) and (10), we have the result $\mathbf{w} = \mathbf{0}$.

Substituting this result back into (5), we now have

$$\mathbf{A}_1 = \begin{bmatrix} \sigma & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix}. \tag{11}$$

The whole process repeats using only the component $\mathbf{B}$, until $\mathbf{A}_n$ becomes diagonal.

□

It is instructive to consider an alternative proof for the SVD. The following is useful because it is a *constructive* proof, which shows us how to form the components of the SVD.

Theorem 2 *Let $\boldsymbol{A} \in \Re^{m \times n}$ be a rank r matrix ($r \leq p = \min(m, n)$). Then there exist orthonormal matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ such that*

$$\boldsymbol{U}^T \boldsymbol{A} \boldsymbol{V} = \begin{bmatrix} \tilde{\boldsymbol{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{12}$$

where

$$\tilde{\boldsymbol{\Sigma}} = diag(\sigma_1, \ldots, \sigma_r), \qquad \sigma_i > 0, \;\; i = 1, \ldots, r. \tag{13}$$

Proof:

Consider the square symmetric positive semi–definite matrix $\boldsymbol{A}^T \boldsymbol{A}$[1]. Let the eigenvalues greater than zero be $\sigma_1^2, \sigma_2^2, \ldots, \sigma_r^2$. Then, from our knowledge of the eigendecomposition, there exists an orthonormal matrix $\boldsymbol{V} \in \Re^{n \times n}$ such that

$$\boldsymbol{V}^T \boldsymbol{A}^T \boldsymbol{A} \boldsymbol{V} = \begin{bmatrix} \tilde{\boldsymbol{\Sigma}}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{14}$$

where $\tilde{\boldsymbol{\Sigma}^2} = \mathrm{diag}[\sigma_1^2, \ldots, \sigma_r^2]$. We now partition $\boldsymbol{V}$ as $[\boldsymbol{V}_1 \quad \boldsymbol{V}_2]$, where $\boldsymbol{V}_1 \in \Re^{n \times r}$. Then (14) has the form

$$\underset{n}{\begin{bmatrix} \boldsymbol{V}_1^T \\ \boldsymbol{V_2}^T \end{bmatrix}} \boldsymbol{A}^T \boldsymbol{A} \begin{bmatrix} \underset{r}{\boldsymbol{V}_1} & \underset{n-r}{\boldsymbol{V}_2} \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{\Sigma}}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{15}$$

[1] The concept of *positive definiteness* is discussed next lecture. It means all the eigenvalues are greater than or equal to zero.

Then by equating corresponding blocks in (15) we have

$$\begin{aligned} \boldsymbol{V}_1^T \boldsymbol{A}^T \boldsymbol{A} \boldsymbol{V}_1 &= \tilde{\boldsymbol{\Sigma}}^2 \quad (r \times r) &(16)\\ \boldsymbol{V}_2^T \boldsymbol{A}^T \boldsymbol{A} \boldsymbol{V}_2 &= \boldsymbol{0}. \quad (n-r) \times (n-r) &(17) \end{aligned}$$

From (16), we can write

$$\tilde{\boldsymbol{\Sigma}}^{-1} \boldsymbol{V}_1^T \boldsymbol{A}^T \boldsymbol{A} \boldsymbol{V}_1 \tilde{\boldsymbol{\Sigma}}^{-1} = \boldsymbol{I}. \tag{18}$$

Then, we define the matrix $\boldsymbol{U}_1 \in \Re^{m \times r}$ from (18) as

$$\boldsymbol{U}_1 = \boldsymbol{A} \boldsymbol{V}_1 \tilde{\boldsymbol{\Sigma}}^{-1}. \tag{19}$$

Then from (18) we have $\boldsymbol{U}_1^T \boldsymbol{U}_1 = \mathbf{I}$ and it follows that

$$\boldsymbol{U}_1^T \boldsymbol{A} \boldsymbol{V}_1 = \tilde{\boldsymbol{\Sigma}}. \tag{20}$$

From (17) we also have

$$\boldsymbol{A} \boldsymbol{V}_2 = \boldsymbol{0}. \tag{21}$$

We now choose a matrix $\boldsymbol{U}_2$ so that $\boldsymbol{U} = [\boldsymbol{U}_1 \;\; \boldsymbol{U}_2]$, where $\boldsymbol{U}_2 \in \Re^{m \times (m-r)}$, is orthonormal. Then from (19) and because $\boldsymbol{U}_1 \perp \boldsymbol{U}_2$, we have

$$\boldsymbol{U}_2^T \boldsymbol{U}_1 = \boldsymbol{U}_2^T \boldsymbol{A} \boldsymbol{V}_1 \tilde{\boldsymbol{\Sigma}}^{-1} = \boldsymbol{0}. \tag{22}$$

Therefore

$$\boldsymbol{U}_2^T \boldsymbol{A} \boldsymbol{V}_1 = \boldsymbol{0}. \tag{23}$$

Combining (20), (21) and (23), we have

$$\boldsymbol{U}^T \boldsymbol{A} \boldsymbol{V} = \begin{bmatrix} \boldsymbol{U}_1^T \boldsymbol{A} \boldsymbol{V}_1 & \boldsymbol{U}_1^T \boldsymbol{A} \boldsymbol{V}_2 \\ \boldsymbol{U}_2^T \boldsymbol{A} \boldsymbol{V}_1 & \boldsymbol{U}_2^T \boldsymbol{A} \boldsymbol{V}_2 \end{bmatrix} = \begin{bmatrix} \tilde{\boldsymbol{\Sigma}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \tag{24}$$

□

The proof can be repeated using an eigendecomposition on the matrix $\boldsymbol{A}\boldsymbol{A}^T \in \Re^{m \times m}$ instead of on $\boldsymbol{A}^T \boldsymbol{A}$. In this case, the roles of the orthonormal matrices $\boldsymbol{V}$ and $\boldsymbol{U}$ are interchanged.

The above proof is useful for several reasons:

- It is short and elegant.
- We can also identify which part of the SVD is not unique. Here, we assume that $\boldsymbol{A}^T\boldsymbol{A}$ has no repeated non–zero eigenvalues. Because $\boldsymbol{V}_2$ are the eigenvectors corresponding to the zero eigenvalues of $\boldsymbol{A}^T\boldsymbol{A}$, $\boldsymbol{V}_2$ is not unique when there are repeated zero eigenvalues. This happens when $m < n+1$, (i.e., $\boldsymbol{A}$ is sufficiently short) or when the nullity of $\boldsymbol{A} \geq 2$, or a combination of these conditions.

  By its construction, the matrix $\boldsymbol{U}_2 \in \Re^{m \times m-r}$ is not unique whenever it consists of two or more columns. This happens when $m - 2 \geq r$.

  It is left as an exercise to show that similar conclusions on the uniqueness of $\boldsymbol{U}$ and $\boldsymbol{V}$ can be made when the proof is developed using the matrix $\boldsymbol{A}\boldsymbol{A}^T$.

3.3 Partitioning the SVD

Following the second proof, we assume that $\mathbf{A}$ has $r \leq p$ non-zero singular values (and $p - r$ zero singular values). Later, we see that $r = \text{rank}(\boldsymbol{A})$. For convenience of notation, we arrange the singular values as:

$$\underbrace{\underset{\text{max}}{\sigma_1} \geq \cdots \geq \underset{\substack{\text{min} \\ \text{non-zero} \\ \text{s.v.}}}{\sigma_r}}_{r \text{ non-zero s.v's}} > \underbrace{\sigma_{r+1} = \cdots = \sigma_p = 0}_{p-r \text{ zero s.v.'s}}$$

In the remainder of this lecture, we use the SVD partitioned in both $\boldsymbol{U}$ and $\boldsymbol{V}$. We can write the SVD of $\boldsymbol{A}$ in the form

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1^T \\ \boldsymbol{V}_2^T \end{bmatrix} \tag{25}$$

where where $\tilde{\boldsymbol{\Sigma}} \in \Re^{r \times r} = \text{diag}(\sigma_1, \ldots, \sigma_r)$, and $\boldsymbol{U}$ is partitioned as

$$\boldsymbol{U} = \begin{bmatrix} \underset{r}{\boldsymbol{U}_1} & \underset{m-r}{\boldsymbol{U}_2} \end{bmatrix} \; {\scriptstyle m} \tag{26}$$

The columns of $\boldsymbol{U}_1$ are the left singular vectors associated with the r nonzero singular values, and the columns of $\boldsymbol{U}_2$ are the left singular vectors associated with the zero singular values. $\mathbf{V}$ is partitioned in an analogous manner:

$$\boldsymbol{V} = \begin{bmatrix} \underset{r}{\boldsymbol{V}_1} & \underset{n-r}{\boldsymbol{V}_2} \end{bmatrix} \; n \tag{27}$$

3.4 Interesting Properties and Interpretations of the SVD

The above partition reveals many interesting properties of the SVD:

3.4.1 $\mathbf{rank}(\mathbf{A}) = r$

Using (25), we can write $\boldsymbol{A}$ as

$$\begin{aligned} \boldsymbol{A} &= \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\Sigma}} \boldsymbol{V}_1^T \\ \mathbf{0} \end{bmatrix} \\ &= \boldsymbol{U}_1 \tilde{\boldsymbol{\Sigma}} \boldsymbol{V}_1^T \\ &= \boldsymbol{U}_1 \boldsymbol{B} \end{aligned} \tag{28}$$

where $\boldsymbol{B} \in \Re^{r\times n} \triangleq \tilde{\boldsymbol{\Sigma}} \boldsymbol{V}_1^T$. From (28) it is clear that the $ith, i = 1, \ldots, r$ column of $\boldsymbol{A}$ is a linear combination of the columns of $\boldsymbol{U}_1$, whose coefficients are given by the ith column of $\boldsymbol{B}$. But since there are $r \leq n$ columns in $\boldsymbol{U}_1$, there can only be r linearly independent columns in $\boldsymbol{A}$. Thus, if $\boldsymbol{A}$ has r non-zero singular values, it follows from the definition of rank that $\text{rank}(\boldsymbol{A}) = r$. It is straightforward to show the converse: if $\boldsymbol{A}$ is rank r, then it has r nonzero singular values.

This point is analogous to the case previously considered in Lecture 2, where we saw rank is equal to the number of non-zero eigenvalues, when $\mathbf{A}$ is a square symmetric matrix. In this case however, the result applies to any matrix. This is another example of how the SVD is a generalization of the eigendecomposition.

Determination of rank when $\sigma_1, \ldots, \sigma_r$ are distinctly greater than zero, and when $\sigma_{r+1}, \ldots, \sigma_p$ are exactly zero is easy. But often in practice, due to finite precision arithmetic and fuzzy data, σ_r may be very small, and σ_{r+1} may be not quite zero. Hence, in practice, determination of rank is not so easy. A common method is to declare rank $\mathbf{A} = r$ if $\sigma_{r+1} \leq \epsilon$, where ϵ is a small number specific to the problem considered.

3.4.2 $N(\mathbf{A}) = \mathbf{R}(\boldsymbol{V}_2)$

Recall the nullspace $N(\mathbf{A}) = \{\boldsymbol{x} \neq \mathbf{0} \mid \boldsymbol{A}\boldsymbol{x} = \mathbf{0}\}$. So, we investigate the set $\{\boldsymbol{x}\}$ such that $\boldsymbol{A}\boldsymbol{x} = \mathbf{0}$. Let $\boldsymbol{x} \in \text{span}(\boldsymbol{V}_2)$; i.e., $x = \boldsymbol{V}_2\boldsymbol{c}$, where $\boldsymbol{c} \in \Re^{n-r}$. By substituting (25) for $\boldsymbol{A}$, by noting that $\boldsymbol{V}_1 \perp \boldsymbol{V}_2$ and that $\boldsymbol{V}_1^T\boldsymbol{V}_1 = \boldsymbol{I}$, we have

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} &= \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \boldsymbol{c} \end{bmatrix} \\ &= \mathbf{0}. \end{aligned} \tag{29}$$

Thus, span$(\boldsymbol{V}_2)$ is at least a subspace of $N(\boldsymbol{A})$. However, if $\boldsymbol{x}$ contains any components of $\boldsymbol{V}_1$, then (29) will not be zero. But since $\boldsymbol{V} = [\boldsymbol{V}_1\boldsymbol{V}_2]$ is a complete basis in $\Re^n$, we see that $\boldsymbol{V}_2$ alone is a basis for the nullspace of $\boldsymbol{A}$.

3.4.3 $R(\mathbf{A}) = R(\mathbf{U}_1)$

Recall that the definition of range $R(\mathbf{A})$ is $\{\mathbf{y} \mid \mathbf{y} = \mathbf{A}\mathbf{x},\ \mathbf{x} \in \Re^n\}$. From (25),

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} &= \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1^T \\ \boldsymbol{V}_2^T \end{bmatrix} \boldsymbol{x} \\ &= \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{d}_1 \\ \boldsymbol{d}_2 \end{bmatrix} \end{aligned} \tag{30}$$

where

$$\begin{matrix} r \\ n-r \end{matrix} \quad \begin{bmatrix} \boldsymbol{d}_1 \\ \boldsymbol{d}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{V}_1^T \\ \boldsymbol{V}_2^T \end{bmatrix} \boldsymbol{x}. \tag{31}$$

From the above we have

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x} &= \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{bmatrix} \begin{bmatrix} \tilde{\boldsymbol{\Sigma}}\boldsymbol{d}_1 \\ \mathbf{0} \end{bmatrix} \\ &= \boldsymbol{U}_1 \left(\tilde{\boldsymbol{\Sigma}}\boldsymbol{d}_1 \right) \end{aligned} \tag{32}$$

We see that as $\mathbf{x}$ moves throughout $\Re^n$, the quantity $\tilde{\mathbf{\Sigma}} \boldsymbol{d}_1$ moves throughout $\Re^r$. Thus, the quantity $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ in this context consists of all linear combinations of the columns of $\boldsymbol{U}_1$. Thus, an orthonormal basis for $R(\boldsymbol{A})$ is $\boldsymbol{U}_1$.

3.4.4 $R(\mathbf{A}^T) = R(\mathbf{V}_1)$

Recall that $R(\mathbf{A}^T)$ is the set of all linear combinations of rows of $\mathbf{A}$. Our property can be seen using a transposed version of the argument in Section 3.4.3 above. Thus, $\mathbf{V}_1$ is an orthonormal basis for the rows of $\mathbf{A}$.

3.4.5 $R(\boldsymbol{A})_\perp = R(\boldsymbol{U}_2)$

From Sect. 3.4.3, we see that $R(\mathbf{A}) = R(\mathbf{U}_1)$. Since from (25), $\boldsymbol{U}_1 \perp \boldsymbol{U}_2$, then $\boldsymbol{U}_2$ is a basis for the orthogonal complement of $R(\boldsymbol{A})$. Hence the result.

3.4.6 $||\mathbf{A}||_2 = \sigma_1 = \sigma_{\max}$

This is easy to see from the definition of the 2-norm and the ellipsoid example of section 3.6.

3.4.7 Inverse of A

If the svd of a square matrix $\boldsymbol{A}$ is given, it is easy to find the inverse. Of course, we must assume $\boldsymbol{A}$ is full rank, (which means $\sigma_i > 0$) for the inverse to exist. The inverse of $\boldsymbol{A}$ is given from the svd, using the familiar rules, as

$$\boldsymbol{A}^{-1} = \boldsymbol{V}\mathbf{\Sigma}^{-1}\boldsymbol{U}^T. \tag{33}$$

The evaluation of $\mathbf{\Sigma}^{-1}$ is easy because $\mathbf{\Sigma}$ is square and diagonal. Note that this treatment indicates that the singular values of $\boldsymbol{A}^{-1}$ are $[\sigma_n^{-1}, \sigma_{n-1}^{-1}, \ldots, \sigma_1^{-1}]$.

3.4.8 The SVD diagonalizes any system of equations

Consider the system of equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, for an arbitrary matrix $\boldsymbol{A}$. Using the SVD of $\mathbf{A}$, we have

$$\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x} = \boldsymbol{b}. \tag{34}$$

Let us now represent $\boldsymbol{b}$ in the basis $\mathbf{U}$, and $\mathbf{x}$ in the basis $\mathbf{V}$, in the same way as in Sect. 3.6. We therefore have

$$\boldsymbol{c} = \begin{matrix} r \\ m-r \end{matrix} \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{U}_1^T \\ \boldsymbol{U}_2^T \end{bmatrix} \boldsymbol{b} \tag{35}$$

and

$$\boldsymbol{d} = \begin{matrix} r \\ n-r \end{matrix} \begin{bmatrix} \boldsymbol{d}_1 \\ \boldsymbol{d}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{V}_1^T \\ \boldsymbol{V}_2^T \end{bmatrix} \boldsymbol{x} \tag{36}$$

Substituting the above into (34), the system of equations becomes

$$\boldsymbol{\Sigma}\boldsymbol{d} = \boldsymbol{c}. \tag{37}$$

This shows that as long as we choose the correct bases, *any* system of equations can become diagonal. This property represents the power of the SVD; it allows us to transform arbitrary algebraic structures into their simplest forms.

Eq. (37) can be expanded as

$$\begin{bmatrix} \tilde{\boldsymbol{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{d}_1 \\ \boldsymbol{d}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_1 \\ \boldsymbol{c}_2 \end{bmatrix} \tag{38}$$

The above equation reveals several interesting facts about the solution of the system of equations. First, if $m > n$ ($\boldsymbol{A}$ is tall) and $\boldsymbol{A}$ is full rank, then the right blocks of zeros in $\boldsymbol{\Sigma}$ are empty. In this case, the system of equations can be satisfied only if $\boldsymbol{c}_2 = \mathbf{0}$. This implies that $\boldsymbol{U}_2^T\boldsymbol{b} = \mathbf{0}$, or that $\boldsymbol{b} \in R(\boldsymbol{U}_1) = R(\boldsymbol{A})$ for a solution to exist.

If $m < n$ ($\boldsymbol{A}$ is short) and full rank, then the bottom blocks of zeros in $\boldsymbol{\Sigma}$ are empty. This implies that a solution to the system of equations exists for any

$\boldsymbol{c}$ or $\boldsymbol{b}$. We note however in this case that $\boldsymbol{d}_1 = \tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{c}$ and $\boldsymbol{d}_2$ is arbitrary. The solution $\boldsymbol{x}$ is not unique and is given by $\boldsymbol{x} = \boldsymbol{V}_1\tilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{c} + \boldsymbol{V}_2\boldsymbol{d}_2$, where $\boldsymbol{d}_2$ is any $n - r$ vector.

If $\boldsymbol{A}$ is not full rank, then none of the zero blocks in (38) are empty. This implies that the two paragraphs above both apply in this case.

3.4.9 The "rotation" interpretation of the SVD

From the SVD relation $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$, we have

$$\boldsymbol{A}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{\Sigma}. \tag{39}$$

Note that since $\boldsymbol{\Sigma}$ is diagonal, the matrix $\boldsymbol{U}\boldsymbol{\Sigma}$ on the right has orthogonal columns, whose 2–norm's are equal to the corresponding singular value. We can therefore interpret the matrix $\boldsymbol{V}$ as an orthonormal matrix which rotates the rows of $\boldsymbol{A}$ so that the result is a matrix with orthogonal columns. Likewise, we have

$$\boldsymbol{U}^T\boldsymbol{A} = \boldsymbol{\Sigma}\boldsymbol{V}^T. \tag{40}$$

The matrix $\boldsymbol{\Sigma}\boldsymbol{V}^T$ on the right has orthogonal rows with 2–norm equal to the corresponding singular value. Thus, the orthonormal matrix $\boldsymbol{U}^T$ operates (rotates) the columns of $\boldsymbol{A}$ to produce a matrix with orthogonal rows.

In the case where $m > n$, ($\boldsymbol{A}$ is tall), then the matrix $\boldsymbol{\Sigma}$ is also tall, with zeros in the bottom $m - n$ rows. Then, only the first n columns of $\boldsymbol{U}$ are relevant in (39), and only the first n rows of $\boldsymbol{U}^T$ are relevant in (40). When $m < n$, a corresponding transposed statement replacing $\boldsymbol{U}$ with $\boldsymbol{V}$ can be made.

3.5 Relationship between SVD and ED

It is clear that the eigendecomposition and the singular value decomposition share many properties in common. The price we pay for being able to perform a diagonal decomposition on an *arbitray* matrix is that we need two orthonormal matrices instead of just one, as is the case for square symmetric matrices. In this section, we explore further relationships between the ED and the SVD.

Using (25), we can write

$$\begin{aligned} \boldsymbol{A}^T\boldsymbol{A} &= \boldsymbol{V}\begin{bmatrix} \tilde{\boldsymbol{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\boldsymbol{U}^T\boldsymbol{U}\begin{bmatrix} \tilde{\boldsymbol{\Sigma}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\boldsymbol{V}^T \\ &= \boldsymbol{V}\begin{bmatrix} \tilde{\boldsymbol{\Sigma}}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\boldsymbol{V}^T. \end{aligned} \tag{41}$$

Thus it is apparent, that the eigenvectors $\mathbf{V}$ of the matrix $\boldsymbol{A}^T\boldsymbol{A}$ are the right singular vectors of $\mathbf{A}$, and that the singular values of $\boldsymbol{A}$ squared are the corresponding nonzero eigenvalues. Note that if $\boldsymbol{A}$ is short $(m < n)$ and full rank, the matrix $\boldsymbol{A}^T\boldsymbol{A}$ will contain $n - m$ additional zero eigenvalues that are not included as singular values of $\mathbf{A}$. This follows because the rank of the matrix $\boldsymbol{A}^T\boldsymbol{A}$ is m when $\boldsymbol{A}$ is full rank, yet the size of $\boldsymbol{A}^T\boldsymbol{A}$ is $n \times n$.

As discussed in *Golub and van Loan*, the SVD is numerically more stable to compute than the ED. However, in the case where $n >> m$, the matrix $\mathbf{V}$ of the SVD of $\mathbf{A}$ becomes large, which means the SVD on $\mathbf{A}$ becomes more costly to compute, relative to the eigendecomposition of $\boldsymbol{A}^T\boldsymbol{A}$.

Further, we can also say, using the form $\boldsymbol{A}\boldsymbol{A}^T$, that

$$\begin{aligned} \boldsymbol{A}\boldsymbol{A}^T &= \boldsymbol{U}\begin{bmatrix} \tilde{\boldsymbol{\Sigma}}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\boldsymbol{V}^T\boldsymbol{V}\begin{bmatrix} \tilde{\boldsymbol{\Sigma}}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\boldsymbol{U}^T \\ &= \boldsymbol{U}\begin{bmatrix} \tilde{\boldsymbol{\Sigma}}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\boldsymbol{U}^T \end{aligned} \tag{42}$$

which indicates that the eigenvectors of $\boldsymbol{A}\boldsymbol{A}^T$ are the left singular vectors $\boldsymbol{U}$ of $\boldsymbol{A}$, and the singular values of $\boldsymbol{A}$ squared are the nonzero eigenvalues of $\boldsymbol{A}\boldsymbol{A}^T$. Notice that in this case, if $\boldsymbol{A}$ is tall and full rank, the matrix $\boldsymbol{A}\boldsymbol{A}^T$ will contain $m - n$ additional zero eigenvalues that are not included as singular values of $\mathbf{A}$.

We now compare the fundamental defining relationships for the ED and the SVD:

For the ED, if $\boldsymbol{A}$ is symmetric, we have:

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T \rightarrow \boldsymbol{A}\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{\Lambda},$$

where $\boldsymbol{Q}$ is the matrix of eigenvectors, and $\boldsymbol{\Lambda}$ is the diagonal matrix of eigenvalues. Writing this relation column-by-column, we have the familiar eigenvector/eigenvalue relationship:

$$\boldsymbol{A}\boldsymbol{q}_i = \lambda_i \boldsymbol{q}_i \quad i = 1, \ldots, n. \qquad * \tag{43}$$

For the SVD, we have

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \rightarrow \mathbf{A}\mathbf{V} = \mathbf{U}\mathbf{\Sigma}$$

or

$$\boldsymbol{A}\boldsymbol{v}_i = \sigma_i \boldsymbol{u}_i \quad i = 1, \ldots, p, \qquad * \tag{44}$$

where $p = \min(m, n)$. Also, since $\boldsymbol{A}^T = \boldsymbol{V}\boldsymbol{\Sigma}\boldsymbol{U}^T \rightarrow \boldsymbol{A}^T\boldsymbol{U} = \boldsymbol{V}\boldsymbol{\Sigma}$, we have

$$\boldsymbol{A}^T\boldsymbol{u}_i = \sigma_i \boldsymbol{v}_i \quad i = 1, \ldots, p. \qquad * \tag{45}$$

Thus, by comparing (43), (44), and (45), we see the singular vectors and singular values obey a relation which is similar to that which defines the eigenvectors and eigenvalues. However, we note that in the SVD case, the fundamental relationship expresses left singular values in terms of right singular values, and vice-versa, whereas the eigenvectors are expressed in terms of themselves.

Exercise: compare the ED and the SVD on a square symmetric matrix, when i) $\boldsymbol{A}$ is positive definite, and ii) when $\boldsymbol{A}$ has some positive and some negative eigenvalues.

3.6 Ellipsoidal Interpretation of the SVD

The singular values of $\mathbf{A}$, where $\mathbf{A} \in \Re^{m \times n}$ are the lengths of the semi-axes of the hyperellipsoid E given by:

$$E = \{\mathbf{y} \mid \mathbf{y} = \mathbf{A}\mathbf{x}, ||\mathbf{x}||_2 = 1\}.$$

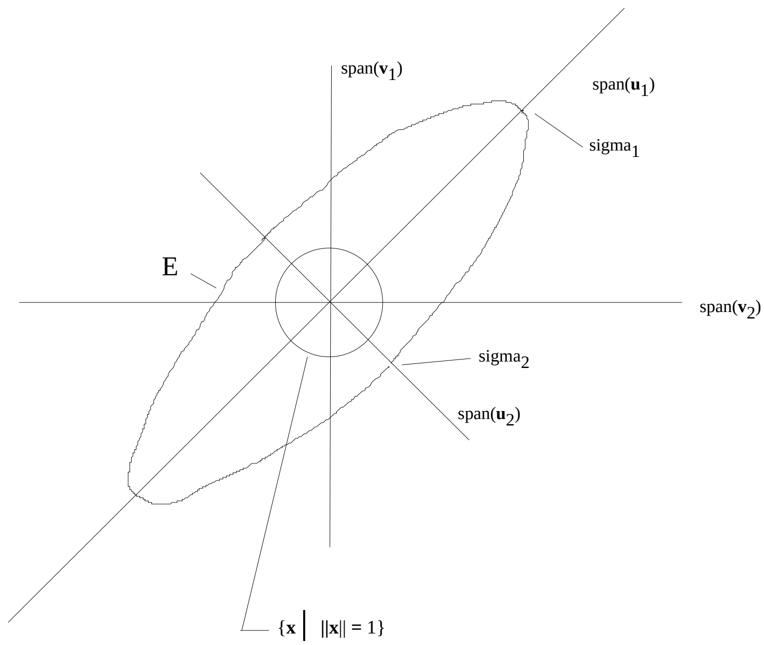

Figure 1: The ellipsoidal interpretation of the SVD. The locus of points $E = \{\boldsymbol{y} \mid \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}, ||\boldsymbol{x}||_2 = 1\}$ defines an ellipse. The principal axes of the ellipse are aligned along the left singular vectors $\boldsymbol{u}_i$, with lengths equal to the corresponding singular value.

That is, E is the set of points mapped out as $\mathbf{x}$ takes on all possible values such that $||\mathbf{x}||_2 = 1$, as shown in Fig. 1. To appreciate this point, let us look at the set of $\mathbf{y}$ corresponding to $\{\boldsymbol{x} \mid ||\boldsymbol{x}||_2 = 1\}$. We take

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{A}\boldsymbol{x} \\ &= \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T\boldsymbol{x}. \end{aligned} \tag{46}$$

Let us change bases for both $\mathbf{x}$ and $\mathbf{y}$. Define

$$\begin{aligned} \boldsymbol{c} &= \boldsymbol{U}^T\boldsymbol{y} \\ \boldsymbol{d} &= \boldsymbol{V}^T\boldsymbol{x}. \end{aligned} \tag{47}$$

Then (46) becomes

$$\boldsymbol{c} = \boldsymbol{\Sigma}\boldsymbol{d}. \tag{48}$$

We note that $||\boldsymbol{d}||_2 = 1$ if $||\boldsymbol{x}||_2 = 1$. Thus, our problem is transformed into observing the set $\{\boldsymbol{c}\}$ corresponding to the set $\{\boldsymbol{d} \mid ||\boldsymbol{d}||_2 = 1\}$. The set $\{\boldsymbol{c}\}$ can be determined by evaluating 2-norms on each side of (48):

$$\sum_{i=1}^{p} \left(\frac{c_i}{\sigma_i}\right)^2 = \sum_{i=1}^{p} (d_i)^2 = 1. \tag{49}$$

We see that the set $\{\boldsymbol{c}\}$ defined by (49) is indeed the canonical form of an ellipse in the basis $\mathbf{U}$. Thus, the principal axes of the ellipse are aligned along the columns $\boldsymbol{u}_i$ of $\mathbf{U}$, with lengths equal to the corresponding singular value σ_i. This interpretation of the SVD is useful later in our study of *condition numbers.*

3.7 An Interesting Theorem

First, we realize that the SVD of $\mathbf{A}$ provides a “sum of outer-products” representation:

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T = \sum_{i=1}^{p} \sigma_i \mathbf{u}_i {\mathbf{v}_i}^T, \quad p = \min(m, n). \tag{50}$$

Given $\mathbf{A} \in \Re^{m \times n}$ with rank r, then what is the matrix $\mathbf{B} \in \Re^{m \times n}$ with rank $k < r$ closest to $\mathbf{A}$ in 2-norm? What is this 2-norm distance? This question is answered in the following theorem:

Theorem 3 *Define*

$$\mathbf{A}_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i{}^T, \qquad k \leq r, \tag{51}$$

then

$$\min_{rank(B)=k} ||\mathbf{A} - \mathbf{B}||_2 = ||\mathbf{A} - \mathbf{A}_k||_2 = \sigma_{k+1}.$$

In words, this says the closest rank $k < r$ matrix $\mathbf{B}$ matrix to $\mathbf{A}$ in the 2–norm sense is given by $\mathbf{A}_k$. $\mathbf{A}_k$ is formed from $\mathbf{A}$ by excluding contributions in (50) associated with the smallest singular values.

Proof:

Since $\mathbf{U}^T \mathbf{A}_k \mathbf{V} = \text{diag}(\sigma_1 \ldots \sigma_k, 0 \ldots 0)$ it follows that $\text{rank}(\mathbf{A}_k) = k$, and that

$$\begin{aligned} ||\mathbf{A} - \mathbf{A}_k||_2 &= \left|\left|\mathbf{U}^T(\mathbf{A} - \mathbf{A}_k)\mathbf{V}\right|\right|_2 \\ &= ||\ \text{diag}(0 \ldots 0, \sigma_{k+1} \ldots \sigma_r, 0 \ldots 0)||_2 \\ &= \sigma_{k+1}. \end{aligned} \tag{52}$$

where the first line follows from the fact the the 2-norm of a matrix is invariant to pre– and post–multiplication by an orthonormal matrix (properties of matrix p-norms, Lecture 2). Further, it may be shown that, for any matrix $\mathbf{B} \in \Re^{m \times n}$ of rank $k < r$, [2]

$$||\mathbf{A} - \mathbf{B}||_2 \geq \sigma_{k+1} \tag{53}$$

Comparing (52)and (53), we see the closest rank k matrix to $\mathbf{A}$ is $\mathbf{A}_k$ given by (51).

□

[2] Golub and van Loan pg. 73.

This result is very useful when we wish to approximate a matrix by another of lower rank. For example, let us look at the Karhunen-Loeve expansion as discussed in Lecture 1. For a sample $\mathbf{x}_n$ of a random process $\mathbf{x} \in \Re^m$, we express $\mathbf{x}$ as

$$\mathbf{x}_i = \mathbf{V}\boldsymbol{\theta}_i \tag{54}$$

where the columns of $\mathbf{V}$ are the eigenvectors of the covariance matrix $\mathbf{R}$. We saw in Lecture 2 that we may represent $\mathbf{x}_i$ with relatively few coefficients by setting the elements of $\boldsymbol{\theta}$ associated with the smallest eigenvalues of $\mathbf{R}$ to zero. The idea was that the resulting distortion in $\mathbf{x}$ would have minimum energy.

This fact may now be seen in a different light with the aid of this theorem. Suppose we retain the $j = r$ elements of a given $\boldsymbol{\theta}$ associated with the largest r eigenvalues. Let $\tilde{\boldsymbol{\theta}} \stackrel{\Delta}{=} [\theta_1, \theta_2, \ldots, \theta_r, 0, \ldots, 0]^T$ and $\tilde{\mathbf{x}} = \mathbf{V}\tilde{\boldsymbol{\theta}}$. Then

$$\begin{aligned}
\tilde{\mathbf{R}} &= E(\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T) \\
&= E(\mathbf{V}\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}^T\mathbf{V}) \\
&= \mathbf{V}\begin{bmatrix} E\mid\theta_1\mid^2 & & & & & \\ & \ddots & & & & \\ & & E\mid\theta_r\mid^2 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}\mathbf{V}^T \\
&= \boldsymbol{V}\tilde{\boldsymbol{\Lambda}}\boldsymbol{V}^T,
\end{aligned} \tag{55}$$

where $\tilde{\boldsymbol{\Lambda}} = \mathrm{diag}\,[\lambda_1 \ldots, \lambda_r, 0 \ldots, 0]$. Since $\tilde{\mathbf{R}}$ is positive definite, square and symmetric, its eigendecomposition and singular value decomposition are identical; hence, $\lambda_i = \sigma_i, i = 1, \ldots, r$. Thus from this theorem, and (55), we know that the covariance matrix $\tilde{\mathbf{R}}$ formed from truncating the K-L coefficients is the closest rank–r matrix to the true covariance matrix $\mathbf{R}$ in the 2–norm sense.

4 Orthogonal Projections

4.1 Sufficient Conditions for a Projector

Suppose we have a subspace $S = R(\boldsymbol{X})$, where $\boldsymbol{X} = [\mathbf{x}_1 \ldots \mathbf{x}_n] \in \Re^{m \times n}$ is full rank, $m > n$, and an arbitrary vector $\mathbf{y} \in \Re^m$. How do we find a matrix $\mathbf{P} \in \Re^{m \times m}$ so that the product $\boldsymbol{P}\boldsymbol{y} \in S$?

The matrix $\mathbf{P}$ is referred to as a *projector*. That is, we can project an arbitrary vector $\mathbf{y}$ onto the subspace S, by premultiplying $\mathbf{y}$ by $\mathbf{P}$. Note that this projection has non-trivial meaning only when $m > n$. Otherwise, $\mathbf{y} \in S$ already for arbitrary $\mathbf{y}$.

A matrix $\mathbf{P}$ is a projection matrix onto S if:

1. $R(\mathbf{P}) = S$
2. $\mathbf{P}^2 = \mathbf{P}$
3. $\mathbf{P}^T = \mathbf{P}$

A matrix satisfying condition (2) is called an *idempotent* matrix. This is the fundamental property of a projector.

We now show that these three conditions are *sufficient* for $\mathbf{P}$ to be a projector. An arbitrary vector $\mathbf{y}$ can be expressed as

$$\mathbf{y} = \mathbf{y}_s + \mathbf{y}_c \tag{56}$$

where $\mathbf{y}_s \in S$ and $\mathbf{y}_c \in S_\perp$ (the orthogonal complement subspace of S). We see that $\mathbf{y}_s$ is the desired projection of $\mathbf{y}$ onto S. Thus, in mathematical terms, our objective is to show that

$$\mathbf{P}\mathbf{y} = \mathbf{y}_s. \tag{57}$$

Because of condition 2, $\mathbf{P}^2 = \mathbf{P}$, hence

$$\mathbf{P}\mathbf{p}_i = \mathbf{p}_i \quad i = 1, \ldots, m \tag{58}$$

where $\mathbf{p}_i$ is a column of $\mathbf{P}$. Because $\mathbf{y}_s \in S$, and also $(\mathbf{p}_1 \dots \mathbf{p}_m) \in S$ (condition1), then $\mathbf{y}_s$ can be expressed as a linear combination of the $\mathbf{p}_i$'s:

$$\mathbf{y}_s = \sum_{i=1}^{m} c_i \mathbf{p}_i, \quad c_i \in \Re. \tag{59}$$

Combining (58) and (59), we have

$$\mathbf{P}\mathbf{y}_s = \sum_{i=1}^{m} c_i \mathbf{P}\mathbf{p}_i = \sum_{i=1}^{m} c_i \mathbf{p}_i = \mathbf{y}_s. \tag{60}$$

If $R(\mathbf{P}) = S$ (condition 1), then $\mathbf{P}\mathbf{y}_c = 0$. Hence,

$$\mathbf{P}\mathbf{y} = \mathbf{P}(\mathbf{y}_s + \mathbf{y}_c) = \mathbf{P}\mathbf{y}_s = \mathbf{y}_s. \tag{61}$$

i.e., $\mathbf{P}$ projects $\mathbf{y}$ onto S, if $\mathbf{P}$ obeys conditions 1 and 2. Furthermore, by repeating the above proof, and using condition 3, we have

$$\mathbf{y}^T \mathbf{P} \in S$$

i.e., $\mathbf{P}$ projects both column- and row–vectors onto S, by pre- and post-multiplying, respectively. Because this property is a direct consequence of the three conditions above, then these conditions are *sufficient* for $\mathbf{P}$ to be a projector.

□

4.2 A Definition for P

Let $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_n], \quad \mathbf{x}_i \in \Re^m, n < m$ be full rank. Then the matrix $\mathbf{P}$ where

$$\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \tag{62}$$

is a projector onto $S = R(\mathbf{X})$. Other definitions of $\mathbf{P}$ equivalent to (62) will follow later after we discuss *pseudo inverses*.

Note that when $\boldsymbol{X}$ has orthonormal columns, then the projector becomes $\boldsymbol{X}\boldsymbol{X}^T \in \Re^{m \times m}$, which according to our previous discussion on orthonormal matrices in Chapter 2, is *not* the $m \times m$ identity.

Exercises:

- prove (62).
- How is $\boldsymbol{P}$ in (62) formed if $r = \text{rank}(\boldsymbol{X}) < n$?

Theorem 4 *The projector onto S defined by (62) is unique.*

Proof:

Let $\boldsymbol{Y}$ be any other $m \times n$ full rank matrix such that $R(\boldsymbol{Y}) = S$. Since $\boldsymbol{X}$ and $\boldsymbol{Y}$ are both in S, each column of $\boldsymbol{Y}$ must be a linear combination of the columns of $\boldsymbol{X}$. Therefore, there exists a full-rank matrix $\boldsymbol{C} \in \Re^{n \times n}$ so that

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{C}. \tag{63}$$

The projector $\boldsymbol{P}_1$ formed from $\boldsymbol{Y}$ is therefore

$$\begin{aligned}
\boldsymbol{P}_1 &= \boldsymbol{Y}(\boldsymbol{Y}^T\boldsymbol{Y})^{-1}\boldsymbol{Y}^T \\
&= \boldsymbol{X}\boldsymbol{C}(\boldsymbol{C}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{C})^{-1}\boldsymbol{C}^T\boldsymbol{X}^T \\
&= \boldsymbol{X}\boldsymbol{C}\boldsymbol{C}^{-1}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{C}^{-T}\boldsymbol{C}^T\boldsymbol{X}^T \\
&= \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T \\
&= \boldsymbol{P}.
\end{aligned} \tag{64}$$

Thus, the projector formed from (62) onto S is unique, regardless of the set of vectors used to form $\mathbf{X}$, provided the corresponding matrix $\mathbf{X}$ is full rank and that $R(\mathbf{X}) = S$.

□

In Section 4.1 we discussed *sufficient* conditions for a projector. This means that while these conditions are enough to specify a projector, there may be other conditions which also specify a projector. But since we have now proved the projector is unique, the conditions in Section 4.1 are also *necessary.*

4.3 The Orthogonal Complement Projector

Consider the vector $\mathbf{y}$, and let $\mathbf{y}_s$ be the projection of $\mathbf{y}$ onto our subspace S, and $\mathbf{y}_c$ be the projection onto the orthogonal complement subspace $S_\perp$.

Thus,

$$\begin{aligned} \mathbf{y} &= \mathbf{y}_s + \mathbf{y}_c \\ &= \mathbf{P}\mathbf{y} + \mathbf{y}_c. \end{aligned} \tag{65}$$

Therefore we have

$$\begin{aligned} \mathbf{y} - \mathbf{P}\mathbf{y} &= \mathbf{y}_c \\ (\mathbf{I} - \mathbf{P})\,\mathbf{y} &= \mathbf{y}_c. \end{aligned} \tag{66}$$

It follows that if $\mathbf{P}$ is a projector onto S, then the matrix $(\mathbf{I} - \mathbf{P})$ is a projector onto $S_\perp$. It is easily verified that this matrix satisfies the all required properties for this projector.

4.4 Orthogonal Projections and the SVD

Suppose we have a matrix $\mathbf{A} \in \Re^{m \times n}$ of rank r. Then, using the partitions of (25), we have these useful relations:

1. $\mathbf{V}_1\mathbf{V}_1{}^T$ is the orthogonal projector onto $[N(\mathbf{A})]^\perp = R(\mathbf{A}^T)$.
2. $\mathbf{V}_2\mathbf{V}_2{}^{\perp T}$ is the orthogonal projector onto $N(\mathbf{A})$
3. $\mathbf{U}_1\mathbf{U}_1{}^T$ is the orthogonal projector onto $R(\mathbf{A})$
4. $\mathbf{U}_2\mathbf{U}_2{}^T$ is the orthogonal projector onto $[R(\mathbf{A})]^\perp = N(\mathbf{A}^T)$

To justify these results, we show each projector listed above satisfies the three conditions for a projector:

1. First, we must show that each projector above is in the range of the corresponding subspace (condition 1). In Sects. 3.6.2 and 3.6.3, we have already verified that $\mathbf{V}_2$ is a basis for $N(\mathbf{A})$, and that $\mathbf{U}_1$ is a basis for $R(\mathbf{A})$, as required. It is easy to verify that the remaining two projectors above (no.'s 1 and 4 respectively) also have the appropriate ranges.
2. From the orthonormality property of each of the matrix partitions above, it is easy to see condition 2 (idempotency) holds in each case.

3. Finally, each matrix above is symmetric (condition 3). Therefore, each matrix above is a projector onto the corresponding subspace.