

## DATA PARTITIONING

### 1. Mappings as data

#### 1.1. Vector spaces of mappings and matrix representations

A vector space  $\mathcal{L}$  can be formed from all linear mappings from the vector space  $\mathcal{U} = (U, S, +, \cdot)$  to another vector space  $\mathcal{V} = (V, S, +, \cdot)$

$$\mathcal{L} = \{L, S, +, \cdot\}, L = \{f|f: U \rightarrow V, f(au + bv) = af(u) + bf(v)\},$$

with addition and scaling of linear mappings defined by  $(f + g)(u) = f(u) + g(u)$  and  $(af)(u) = af(u)$ . Let  $B = \{u_1, u_2, \dots\}$  denote a basis for the domain  $U$  of linear mappings within  $\mathcal{L}$ , such that the linear mapping  $f \in \mathcal{L}$  is represented by the matrix

$$A = [f(u_1) \ f(u_2) \ \dots].$$

When the domain and codomain are the real vector spaces  $U = \mathbb{R}^n$ ,  $V = \mathbb{R}^m$ , the above is a standard matrix of real numbers,  $A \in \mathbb{R}^{m \times n}$ . For linear mappings between infinite dimensional vector spaces the matrix is understood in a generalized sense to contain an infinite number of columns that are elements of the codomain  $V$ . For example, the indefinite integral is a linear mapping between the vector space of functions that allow differentiation to any order,

$$\int: \mathcal{C}^\infty \rightarrow \mathcal{C}^\infty \quad v(x) = \int u(x) dx$$

and for the monomial basis  $B = \{1, x, x^2, \dots\}$ , is represented by the generalized matrix

$$A = \left[ x \quad \frac{1}{2}x^2 \quad \frac{1}{3}x^3 \quad \dots \right].$$

Truncation of the basis expansion  $u(x) = \sum_{j=1}^{\infty} u_j x^j$  where  $u_j \in \mathbb{R}$  to  $n$  terms, and sampling of  $u \in \mathcal{C}^\infty$  at points  $x_1, \dots, x_m$ , forms a standard matrix of real numbers

$$A = \left[ x \quad \frac{1}{2}x^2 \quad \frac{1}{3}x^3 \quad \dots \right] \in \mathbb{R}^{m \times n}, \quad \mathbf{x}^j = \begin{bmatrix} x_1^j \\ \vdots \\ x_m^j \end{bmatrix}.$$

As to be expected, matrices can also be organized as vector space  $\mathcal{M}$ , which is essentially the representation of the associated vector space of linear mappings,

$$\mathcal{M} = (M, S, +, \cdot) \quad M = \{A|A = [f(u_1) \ f(u_2) \ \dots]\}.$$

The addition  $C = A + B$  and scaling  $S = aR$  of matrices is given in terms of the matrix components by

$$c_{ij} = a_{ij} + b_{ij}, s_{ij} = ar_{ij}.$$

#### 1.2. Measurement of mappings

From the above it is apparent that linear mappings and matrices can also be considered as data, and a first step in analysis of such data is definition of functionals that would attach a single scalar label to each linear mapping of matrix. Of particular interest is the definition of a norm functional that characterizes in an appropriate sense the size of a linear mapping.

Consider first the case of finite matrices with real components  $A \in \mathbb{R}^{m \times n}$  that represent linear mappings between real vector spaces  $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$ . The columns  $\mathbf{a}_1, \dots, \mathbf{a}_n$  of  $A \in \mathbb{R}^{m \times n}$  could be placed into a single column vector  $\mathbf{c}$  with  $mn$  components

$$\mathbf{c} = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{bmatrix}.$$

Subsequently the norm of the matrix  $A$  could be defined as the norm of the vector  $\mathbf{c}$ . An example of this approach is the Frobenius norm

$$\|A\|_F = \|\mathbf{c}\|_2 = \left( \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2}.$$

A drawback of the above approach is that the structure of the matrix and its close relationship to a linear mapping is lost. A more useful characterization of the size of a mapping is to consider the amplification behavior of linear mapping. The motivation is readily understood starting from linear mappings between the reals  $f: \mathbb{R} \rightarrow \mathbb{R}$ , that are of the form  $f(x) = ax$ . When given an argument of unit magnitude  $|x| = 1$ , the mapping returns a real number with magnitude  $|a|$ . For mappings  $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  within the plane, arguments that satisfy  $\|\mathbf{x}\|_2 = 1$  are on the unit circle with components  $\mathbf{x} = [\cos \theta \ \sin \theta]$  have images through  $f$  given analytically by

$$f(\mathbf{x}) = A\mathbf{x} = [\mathbf{a}_1 \ \mathbf{a}_2] \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} = \cos \theta \mathbf{a}_1 + \sin \theta \mathbf{a}_2,$$

and correspond to ellipses.

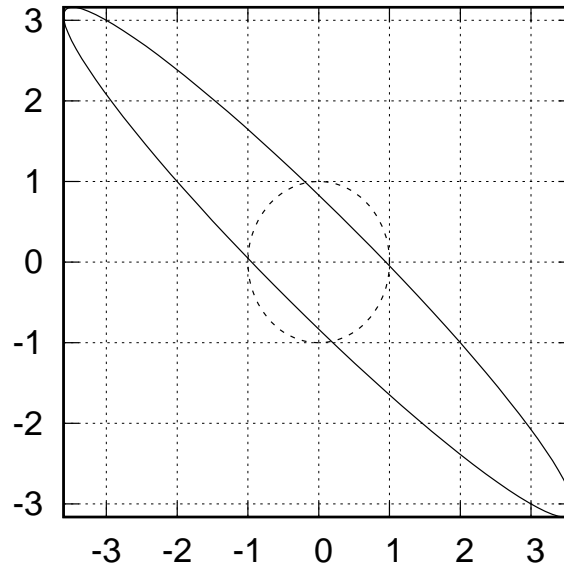


Figure 1. Mapping of unit circle by  $f(\mathbf{x}) = A\mathbf{x}$ ,  $A = \begin{bmatrix} 2 & 3 \\ -1 & -3 \end{bmatrix}$ .

From the above the mapping associated  $A$  amplifies some directions more than others. This suggests a definition of the size of a matrix or a mapping by the maximal amplification unit norm vectors within the domain.

DEFINITION. For vector spaces  $U, V$  with norms  $\|\cdot\|_U: U \rightarrow \mathbb{R}_+$ ,  $\|\cdot\|_V: V \rightarrow \mathbb{R}_+$ , the *induced norm* of  $f: U \rightarrow V$  is

$$\|f\| = \sup_{\|\mathbf{x}\|_U=1} \|f(\mathbf{x})\|_V.$$

DEFINITION. For vector spaces  $\mathbb{R}^n, \mathbb{R}^m$  with norms  $\|\cdot\|^{(n)}: \mathbb{R}^n \rightarrow \mathbb{R}_+$ ,  $\|\cdot\|^{(m)}: \mathbb{R}^m \rightarrow \mathbb{R}_+$ , the *induced norm* of matrix  $A \in \mathbb{R}^{m \times n}$  is

$$\|A\| = \sup_{\|\mathbf{x}\|^{(n)}=1} \|A\mathbf{x}\|^{(m)}.$$

In the above, any vector norm can be used within the domain and codomain.

## 2. The Singular Value Decomposition (SVD)

The fundamental theorem of linear algebra partitions the domain and codomain of a linear mapping  $f: U \rightarrow V$ . For real vectors spaces  $U = \mathbb{R}^n$ ,  $V = \mathbb{R}^m$  the partition properties are stated in terms of spaces of the associated matrix  $A$  as

$$C(A) \oplus N(A^T) = \mathbb{R}^m \quad C(A) \perp N(A^T) \quad C(A^T) \oplus N(A) = \mathbb{R}^n \quad C(A^T) \perp N(A).$$

The dimension of the column and row spaces  $r = \dim C(A) = \dim C(A^T)$  is the rank of the matrix,  $n - r$  is the nullity of  $A$ , and  $m - r$  is the nullity of  $A^T$ . A infinite number of bases could be defined for the domain and codomain. It is of great theoretical and practical interest bases with properties that facilitate insight or computation.

### 2.1. Orthogonal matrices

The above partitions of the domain and codomain are orthogonal, and suggest searching for orthogonal bases within these subspaces. Introduce a matrix representation for the bases

$$U = [u_1 \ u_2 \ \dots \ u_m] \in \mathbb{R}^{m \times m}, V = [v_1 \ v_2 \ \dots \ v_n] \in \mathbb{R}^{n \times n},$$

with  $C(U) = \mathbb{R}^m$  and  $C(V) = \mathbb{R}^n$ . Orthogonality between columns  $u_i, u_j$  for  $i \neq j$  is expressed as  $u_i^T u_j = 0$ . For  $i = j$ , the inner product is positive  $u_i^T u_i > 0$ , and since scaling of the columns of  $U$  preserves the spanning property  $C(U) = \mathbb{R}^m$ , it is convenient to impose  $u_i^T u_i = 1$ . Such behavior is concisely expressed as a matrix product

$$U^T U = I_m,$$

with  $I_m$  the identity matrix in  $\mathbb{R}^m$ . Expanded in terms of the column vectors of  $U$  the first equality is

$$[u_1 \ u_2 \ \dots \ u_m]^T [u_1 \ u_2 \ \dots \ u_m] = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_m^T \end{bmatrix} [u_1 \ u_2 \ \dots \ u_m] = \begin{bmatrix} u_1^T u_1 & u_1^T u_2 & \dots & u_1^T u_m \\ u_2^T u_1 & u_2^T u_2 & \dots & u_2^T u_m \\ \vdots & \vdots & \ddots & \vdots \\ u_m^T u_1 & u_m^T u_2 & \dots & u_m^T u_m \end{bmatrix} = I_m.$$

It is useful to determine if a matrix  $X$  exists such that  $UX = I_m$ , or

$$UX = U [x_1 \ x_2 \ \dots \ x_m] = [e_1 \ e_2 \ \dots \ e_m].$$

The columns of  $X$  are the coordinates of the column vectors of  $I_m$  in the basis  $U$ , and can readily be determined

$$Ux_j = e_j \Rightarrow U^T Ux_j = U^T e_j \Rightarrow I_m x_j = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_m^T \end{bmatrix} e_j \Rightarrow x_j = (U^T)_j,$$

where  $(U^T)_j$  is the  $j^{\text{th}}$  column of  $U^T$ , hence  $X = U^T$ , leading to

$$U^T U = I = U U^T.$$

Note that the second equality

$$[u_1 \ u_2 \ \dots \ u_m] [u_1 \ u_2 \ \dots \ u_m]^T = [u_1 \ u_2 \ \dots \ u_m] \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_m^T \end{bmatrix} = u_1 u_1^T + u_2 u_2^T + \dots + u_m u_m^T = I$$

acts as normalization condition on the matrices  $U_j = u_j u_j^T$ .

DEFINITION. A square matrix  $U$  is said to be orthogonal if  $U^T U = U U^T = I$ .

## 2.2. Intrinsic basis of a linear mapping

Given a linear mapping  $f: U \rightarrow V$ , expressed as  $y = f(x) = Ax$ , the simplest description of the action of  $A$  would be a simple scaling, as exemplified by  $g(x) = ax$  that has as its associated matrix  $aI$ . Recall that specification of a vector is typically done in terms of the identity matrix  $b = Ib$ , but may be more insightfully given in some other basis  $Ax = Ib$ . This suggests that especially useful bases for the domain and codomain would reduce the action of a linear mapping to scaling along orthogonal directions, and evaluate  $y = Ax$  by first re-expressing  $y$  in another basis  $U$ ,  $Us = Iy$  and re-expressing  $x$  in another basis  $V$ ,  $Vr = Ix$ . The condition that the linear operator reduces to simple scaling in these new bases is expressed as  $s_i = \sigma_i r_i$  for  $i = 1, \dots, \min(m, n)$ , with  $\sigma_i$  the scaling coefficients along each direction which can be expressed as a matrix vector product  $s = \Sigma r$ , where  $\Sigma \in \mathbb{R}^{m \times n}$  is of the same dimensions as  $A$  and given by

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_r & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix}.$$

Imposing the condition that  $U, V$  are orthogonal leads to

$$Us = y \Rightarrow s = U^T y, Vr = x \Rightarrow r = V^T x,$$

which can be replaced into  $s = \Sigma r$  to obtain

$$U^T y = \Sigma V^T x \Rightarrow y = U \Sigma V^T x.$$

From the above the orthogonal bases  $U, V$  and scaling coefficients  $\Sigma$  that are sought must satisfy  $A = U \Sigma V^T$ .

**THEOREM.** Every matrix  $A \in \mathbb{R}^{m \times n}$  has a *singular value decomposition (SVD)*

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

with properties:

1.  $U \in \mathbb{R}^{m \times m}$  is an orthogonal matrix,  $U^T U = I_m$ ;
2.  $V \in \mathbb{R}^{n \times n}$  is an orthogonal matrix,  $V^T V = I_n$ ;
3.  $\Sigma \in \mathbb{R}^{m \times n}$  is diagonal,  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_p)$ ,  $p = \min(m, n)$ , and  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$ .

**Proof.** The proof of the SVD makes use of properties of the norm, concepts from analysis and complete induction. Adopting the 2-norm set  $\sigma_1 = \|A\|_2$ ,

$$\sigma_1 = \sup_{\|x\|_2=1} \|Ax\|_2.$$

The domain  $\|x\|_2 = 1$  is compact (closed and bounded), and the extreme value theorem implies that  $f(x) = Ax$  attains its maxima and minima, hence there must exist some vectors  $u_1, v_1$  of unit norm such that  $\sigma_1 u_1 = Av_1 \Rightarrow \sigma_1 = u_1^T Av_1$ . Introduce orthogonal bases  $U_1, V_1$  for  $\mathbb{R}^m, \mathbb{R}^n$  whose first column vectors are  $u_1, v_1$ , and compute

$$U_1^T A V_1 = \begin{bmatrix} u_1^T \\ \vdots \\ u_m^T \end{bmatrix} [Av_1 \dots Av_n] = \begin{bmatrix} \sigma_1 & w^T \\ \mathbf{0} & B \end{bmatrix} = C.$$

In the above  $w^T$  is a row vector with  $n-1$  components  $u_1^T Av_j$ ,  $j=2, \dots, n$ , and  $u_i^T Av_1$  must be zero for  $u_1$  to be the direction along which the maximum norm  $\|Av_1\|$  is obtained. Introduce vectors

$$y = \begin{bmatrix} \sigma_1 \\ w \end{bmatrix}, z = Cy = \begin{bmatrix} \sigma_1^2 + w^T w \\ Bw \end{bmatrix},$$

and note that  $\|z\|_2 \geq \|y\|_2^2 = \sigma_1^2 + w^T w$ . From  $\|U_1^T A V_1\| = \|A\| = \sigma_1 = \|C\| \geq \sigma_1^2 + w^T w$  it results that  $w = \mathbf{0}$ . By induction, assume that  $B$  has a singular value decomposition,  $B = U_2 \Sigma_2 V_2^T$ , such that

$$U_1^T A V_1 = \begin{bmatrix} \sigma_1 & \mathbf{0}^T \\ \mathbf{0} & U_2 \Sigma_2 V_2^T \end{bmatrix} = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & U_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & \mathbf{0}^T \\ \mathbf{0} & \Sigma_2 \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & V_2^T \end{bmatrix},$$

and the orthogonal matrices arising in the singular value decomposition of  $A$  are

$$U = U_1 \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & U_2 \end{bmatrix}, V^T = \begin{bmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & V_2^T \end{bmatrix} V_1^T.$$

□

The scaling coefficients  $\sigma_j$  are called the *singular values* of  $A$ . The columns of  $U$  are called the *left singular vectors*, and those of  $V$  are called the *right singular vectors*.

The fact that the scaling coefficients are norms of  $A$  and submatrices of  $A$ ,  $\sigma_1 = \|A\|$ , is crucial importance in applications. Carrying out computation of the matrix products

$$A = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_r \ \mathbf{u}_{r+1} \ \dots \ \mathbf{u}_m] \begin{bmatrix} \sigma_1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_r & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_r^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} = [\mathbf{u}_1 \ \mathbf{u}_2 \ \dots \ \mathbf{u}_r \ \mathbf{u}_{r+1} \ \dots \ \mathbf{u}_m] \begin{bmatrix} \sigma_1 \mathbf{v}_1^T \\ \sigma_2 \mathbf{v}_2^T \\ \vdots \\ \sigma_r \mathbf{v}_r^T \\ \vdots \\ 0 \end{bmatrix}$$

leads to a representation of  $A$  as a sum

$$A = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^T, r \leq \min(m, n).$$

Each product  $\mathbf{u}_i \mathbf{v}_i^T$  is a matrix of rank one, and is called a rank-one update. Truncation of the above sum to  $p$  terms leads to an approximation of  $A$

$$A \cong A_p = \sum_{i=1}^p \sigma_i \mathbf{u}_i \mathbf{v}_i^T.$$

In very many cases the singular values exhibit rapid, exponential decay,  $\sigma_1 \gg \sigma_2 \gg \dots$ , such that the approximation above is an accurate representation of the matrix  $A$ .



**Figure 1.** Successive SVD approximations of Frida Kahlo's (1907-1954) painting, *Portrait of a Lady in White* (1929), with  $k = 10, 20, 40$  rank-one updates.

### 2.3. SVD solution of linear algebra problems

The SVD can be used to solve common problems within linear algebra.

**Change of coordinates.** To change from vector coordinates  $\mathbf{b}$  in the canonical basis  $\mathbf{I} \in \mathbb{R}^{m \times m}$  to coordinates  $\mathbf{x}$  in some other basis  $\mathbf{A} \in \mathbb{R}^{m \times m}$ , a solution to the equation  $\mathbf{I}\mathbf{b} = \mathbf{A}\mathbf{x}$  can be found by the following steps.

1. Compute the SVD,  $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \mathbf{A}$ ;
2. Find the coordinates of  $\mathbf{b}$  in the orthogonal basis  $\mathbf{U}$ ,  $\mathbf{c} = \mathbf{U}^T \mathbf{b}$ ;
3. Scale the coordinates of  $\mathbf{c}$  by the inverse of the singular values  $y_i = c_i / \sigma_i$ ,  $i = 1, \dots, m$ , such that  $\mathbf{\Sigma}\mathbf{y} = \mathbf{c}$  is satisfied;

4. Find the coordinates of  $\mathbf{y}$  in basis  $\mathbf{V}^T$ ,  $\mathbf{x} = \mathbf{V}\mathbf{y}$ .

**Best 2-norm approximation.** In the above  $\mathbf{A}$  was assumed to be a basis, hence  $r = \text{rank}(\mathbf{A}) = m$ . If columns of  $\mathbf{A}$  do not form a basis,  $r < m$ , then  $\mathbf{b} \in \mathbb{R}^m$  might not be reachable by linear combinations within  $C(\mathbf{A})$ . The closest vector to  $\mathbf{b}$  in the norm is however found by the same steps, with the simple modification that in Step 3, the scaling is carried out only for non-zero singular values,  $y_i = c_i / \sigma_i$ ,  $i = 1, \dots, r$ .

**The pseudo-inverse.** From the above, finding either the solution of  $\mathbf{A}\mathbf{x} = \mathbf{I}\mathbf{b}$  or the best approximation possible if  $\mathbf{A}$  is not of full rank, can be written as a sequence of matrix multiplications using the SVD

$$(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)\mathbf{x} = \mathbf{b} \Rightarrow \mathbf{U}(\mathbf{\Sigma}\mathbf{V}^T\mathbf{x}) = \mathbf{b} \Rightarrow (\mathbf{\Sigma}\mathbf{V}^T\mathbf{x}) = \mathbf{U}^T\mathbf{b} \Rightarrow \mathbf{V}^T\mathbf{x} = \mathbf{\Sigma}^+\mathbf{U}^T\mathbf{b} \Rightarrow \mathbf{x} = \mathbf{V}\mathbf{\Sigma}^+\mathbf{U}^T\mathbf{b},$$

where the matrix  $\mathbf{\Sigma}^+ \in \mathbb{R}^{n \times m}$  (notice the inversion of dimensions) is defined as a matrix with elements  $\sigma_i^{-1}$  on the diagonal, and is called the pseudo-inverse of  $\mathbf{\Sigma}$ . Similarly the matrix

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

that allows stating the solution of  $\mathbf{A}\mathbf{x} = \mathbf{b}$  simply as  $\mathbf{x} = \mathbf{A}^+\mathbf{b}$  is called the *pseudo-inverse* of  $\mathbf{A}$ . Note that in practice  $\mathbf{A}^+$  is not explicitly formed. Rather the notation  $\mathbf{A}^+$  is simply a concise reference to carrying out steps 1-4 above.