DATA PARTITIONING

1. Mappings as data

1.1. Vector spaces of mappings and matrix representations

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

$$\mathscr{L} = \{L, S, +, \cdot\}, \ \ L = \{\boldsymbol{f} | \boldsymbol{f}: U \to V, \boldsymbol{f}(a\boldsymbol{u} + b\boldsymbol{v}) = af(\boldsymbol{u}) + bf(\boldsymbol{v})\},$$

with addition and scaling of linear mappings defined by $(\boldsymbol{f} + \boldsymbol{g})(\boldsymbol{u}) = \boldsymbol{f}(\boldsymbol{u}) + \boldsymbol{g}(\boldsymbol{u})$ and $(a\boldsymbol{f})(\boldsymbol{u}) = a\boldsymbol{f}(\boldsymbol{u})$. Let $B = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots\}$ denote a basis for the domain U of linear mappings within $\mathscr{L}$, such that the linear mapping $\boldsymbol{f} \in \mathscr{L}$ is represented by the matrix

$$\boldsymbol{A} = [\,\boldsymbol{f}(\boldsymbol{u}_1) \ \ \boldsymbol{f}(\boldsymbol{u}_2) \ \ \ldots\,].$$

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, $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. 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,

$$\int : \mathscr{C}^\infty \to \mathscr{C}^\infty \ \ v(x) = \int u(x)\,\mathrm{d}x$$

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

$$\boldsymbol{A} = \left[\, x \ \ \tfrac{1}{2}x^2 \ \ \tfrac{1}{3}x^3 \ \ \ldots \,\right].$$

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 $u \in \mathscr{C}^\infty$ at points $x_1, \ldots, x_m$, forms a standard matrix of real numbers

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

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

$$\mathscr{M} = (M, S, +, \cdot) \ \ M = \{\boldsymbol{A} | \boldsymbol{A} = [\,\boldsymbol{f}(\boldsymbol{u}_1) \ \ \boldsymbol{f}(\boldsymbol{u}_2) \ \ \ldots\,]\}\,.$$

The addition $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B}$ and scaling $\boldsymbol{S} = a\boldsymbol{R}$ of matrices is given in terms of the matrix components by

$$c_{ij} = a_{ij} + b_{ij}, s_{ij} = a r_{\mathrm{ij}}.$$

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 $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ that represent linear mappings between real vector spaces $\boldsymbol{f}: \mathbb{R}^m \to \mathbb{R}^n$. The columns $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ could be placed into a single column vector $\boldsymbol{c}$ with mn components

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

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

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

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 $f: \mathbb{R} \to \mathbb{R}$, 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 $\boldsymbol{f}: \mathbb{R}^2 \to \mathbb{R}^2$ within the plane, arguments that satisfy $\|\boldsymbol{x}\|_2 = 1$ are on the unit circle with components $\boldsymbol{x} = [\ \cos\theta \ \ \sin\theta\ ]$ have images through $\boldsymbol{f}$ given analytically by

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

and correspond to ellipses.

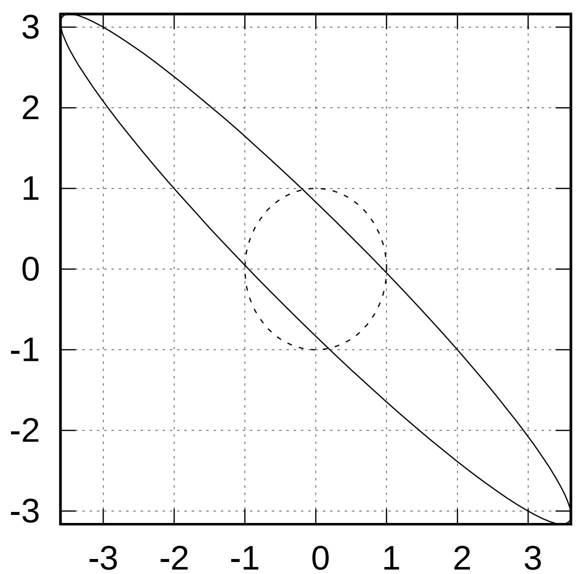

Figure 1. Mapping of unit circle by $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$, $\boldsymbol{A} = \begin{bmatrix} 2 & 3 \\ -1 & -3 \end{bmatrix}$.

From the above the mapping associated $\boldsymbol{A}$ 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 $\| \|_U: U \to \mathbb{R}_+$, $\| \|_V: V \to \mathbb{R}_+$, the induced norm of $\boldsymbol{f}: U \to V$ is*

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

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

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

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 $\boldsymbol{f}: U \to V$. 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 $\boldsymbol{A}$ as

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

The dimension of the column and row spaces $r = \dim C(\boldsymbol{A}) = \dim C(\boldsymbol{A}^T)$ is the rank of the matrix, $n - r$ is the nullity of $\boldsymbol{A}$, 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 faciliatate 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

$$\boldsymbol{U} = [\ \boldsymbol{u}_1 \ \ \boldsymbol{u}_2 \ \ \ldots \ \ \boldsymbol{u}_m \ ] \in \mathbb{R}^{m\times m}, \boldsymbol{V} = [\ \boldsymbol{v}_1 \ \ \boldsymbol{v}_2 \ \ \ldots \ \ \boldsymbol{v}_n \ ] \in \mathbb{R}^{n\times n},$$

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

$$\boldsymbol{U}^T \boldsymbol{U} = \boldsymbol{I}_m,$$

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

$$[\ \boldsymbol{u}_1 \ \ \boldsymbol{u}_2 \ \ \ldots \ \ \boldsymbol{u}_m \ ]^T [\ \boldsymbol{u}_1 \ \ \boldsymbol{u}_2 \ \ \ldots \ \ \boldsymbol{u}_m \ ] = \begin{bmatrix} \boldsymbol{u}_1^T \\ \boldsymbol{u}_2^T \\ \vdots \\ \boldsymbol{u}_m^T \end{bmatrix} [\ \boldsymbol{u}_1 \ \ \boldsymbol{u}_2 \ \ \ldots \ \ \boldsymbol{u}_m \ ] = \begin{bmatrix} \boldsymbol{u}_1^T \boldsymbol{u}_1 & \boldsymbol{u}_1^T \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_1^T \boldsymbol{u}_m \\ \boldsymbol{u}_2^T \boldsymbol{u}_1 & \boldsymbol{u}_2^T \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_2^T \boldsymbol{u}_m \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{u}_m^T \boldsymbol{u}_1 & \boldsymbol{u}_m^T \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_m^T \boldsymbol{u}_m \end{bmatrix} = \boldsymbol{I}_m.$$

It is useful to determine if a matrix $\boldsymbol{X}$ exists such that $\boldsymbol{U}\boldsymbol{X} = \boldsymbol{I}_m$, or

$$\boldsymbol{U}\boldsymbol{X} = \boldsymbol{U} [\ \boldsymbol{x}_1 \ \ \boldsymbol{x}_2 \ \ \ldots \ \ \boldsymbol{x}_m \ ] = [\ \boldsymbol{e}_1 \ \ \boldsymbol{e}_2 \ \ \ldots \ \ \boldsymbol{e}_m \ ].$$

The columns of $\boldsymbol{X}$ are the coordinates of the column vectors of $\boldsymbol{I}_m$ in the basis $\boldsymbol{U}$, and can readily be determined

$$\boldsymbol{U}\boldsymbol{x}_j = \boldsymbol{e}_j \Rightarrow \boldsymbol{U}^T \boldsymbol{U}\boldsymbol{x}_j = \boldsymbol{U}^T \boldsymbol{e}_j \Rightarrow \boldsymbol{I}_m \boldsymbol{x}_j = \begin{bmatrix} \boldsymbol{u}_1^T \\ \boldsymbol{u}_2^T \\ \vdots \\ \boldsymbol{u}_m^T \end{bmatrix} \boldsymbol{e}_j \Rightarrow \boldsymbol{x}_j = (\boldsymbol{U}^T)_j,$$

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

$$\boldsymbol{U}^T \boldsymbol{U} = \boldsymbol{I} = \boldsymbol{U}\boldsymbol{U}^T.$$

Note that the second equality

$$[\ \boldsymbol{u}_1 \ \ \boldsymbol{u}_2 \ \ \ldots \ \ \boldsymbol{u}_m \ ][\ \boldsymbol{u}_1 \ \ \boldsymbol{u}_2 \ \ \ldots \ \ \boldsymbol{u}_m \ ]^T = [\ \boldsymbol{u}_1 \ \ \boldsymbol{u}_2 \ \ \ldots \ \ \boldsymbol{u}_m \ ] \begin{bmatrix} \boldsymbol{u}_1^T \\ \boldsymbol{u}_2^T \\ \vdots \\ \boldsymbol{u}_m^T \end{bmatrix} = \boldsymbol{u}_1 \boldsymbol{u}_1^T + \boldsymbol{u}_2 \boldsymbol{u}_2^T + \cdots + \boldsymbol{u}_m \boldsymbol{u}_m^T = \boldsymbol{I}$$

acts as normalization condition on the matrices $\boldsymbol{U}_j = \boldsymbol{u}_j \boldsymbol{u}_j^T$.

DEFINITION. *A square matrix* $\boldsymbol{U}$ *is said to be orthogonal if* $\boldsymbol{U}^T \boldsymbol{U} = \boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{I}$.

2.2. Intrinsic bases of a linear mapping

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

$$\boldsymbol{\Sigma} = \left[\begin{array}{ccccccc} \sigma_1 & 0 & \ldots & 0 & 0 & \ldots & 0\\ 0 & \sigma_2 & \ldots & 0 & 0 & \ldots & 0\\ \vdots & \vdots & \ddots & 0 & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & \sigma_r & 0 & \ldots & 0\\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \end{array}\right].$$

Imposing the condition that $\boldsymbol{U}, \boldsymbol{V}$ are orthogonal leads to

$$\boldsymbol{U}\boldsymbol{s} = \boldsymbol{y} \Rightarrow \boldsymbol{s} = \boldsymbol{U}^T\boldsymbol{y}, \boldsymbol{V}\boldsymbol{r} = \boldsymbol{x} \Rightarrow \boldsymbol{r} = \boldsymbol{V}^T\boldsymbol{x},$$

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

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

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

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

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T,$$

with properties:

1. *$\boldsymbol{U} \in \mathbb{R}^{m \times m}$ is an orthogonal matrix, $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}_m$;*
2. *$\boldsymbol{V} \in \mathbb{R}^{m \times m}$ is an orthogonal matrix, $\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_n$;*
3. *$\boldsymbol{\Sigma} \in \mathbb{R}^{m \times n}$ is diagonal, $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_p)$, $p = \min(m, n)$, and $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_p \geqslant 0$.*

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

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

The domain $\|\boldsymbol{x}\|_2 = 1$ is compact (closed and bounded), and the extreme value theorem implies that $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$ attains its maxima and minima, hence there must exist some vectors $\boldsymbol{u}_1, \boldsymbol{v}_1$ of unit norm such that $\sigma_1\boldsymbol{u}_1 = \boldsymbol{A}\boldsymbol{v}_1 \Rightarrow \sigma_1 = \boldsymbol{u}_1^T\boldsymbol{A}\boldsymbol{v}_1$. Introduce orthogonal bases $\boldsymbol{U}_1$, $\boldsymbol{V}_1$ for $\mathbb{R}^m, \mathbb{R}^n$ whose first column vectors are $\boldsymbol{u}_1, \boldsymbol{v}_1$, and compute

$$\boldsymbol{U}_1{}^T\boldsymbol{A}\boldsymbol{V}_1 = \left[\begin{array}{c} \boldsymbol{u}_1^T\\ \vdots\\ \boldsymbol{u}_m^T \end{array}\right] [\, \boldsymbol{A}\boldsymbol{v}_1 \; \ldots \; \boldsymbol{A}\boldsymbol{v}_n \,] = \left[\begin{array}{cc} \sigma_1 & \boldsymbol{w}^T\\ \boldsymbol{0} & \boldsymbol{B} \end{array}\right] = \boldsymbol{C}.$$

In the above $\boldsymbol{w}^T$ is a row vector with $n-1$ components $\boldsymbol{u}_1^T\boldsymbol{A}\boldsymbol{v}_j$, $j = 2, \ldots, n$, and $\boldsymbol{u}_i^T\boldsymbol{A}\boldsymbol{v}_1$ must be zero for $\boldsymbol{u}_1$ to be the direction along which the maximum norm $\|\boldsymbol{A}\boldsymbol{v}_1\|$ is obtained. Introduce vectors

$$\boldsymbol{y} = \left[\begin{array}{c} \sigma_1\\ \boldsymbol{w} \end{array}\right], z = \boldsymbol{C}\boldsymbol{y} = \left[\begin{array}{c} \sigma_1^2 + \boldsymbol{w}^T\boldsymbol{w}\\ \boldsymbol{B}\boldsymbol{w} \end{array}\right],$$

and note that $\|z\|_2 \geqslant \|\boldsymbol{y}\|_2^2 = \sigma_1^2 + \boldsymbol{w}^T\boldsymbol{w}$. From $\|\boldsymbol{U}_1{}^T\boldsymbol{A}\boldsymbol{V}_1\| = \|\boldsymbol{A}\| = \sigma_1 = \|\boldsymbol{C}\| \geqslant \sigma_1^2 + \boldsymbol{w}^T\boldsymbol{w}$ it results that $\boldsymbol{w} = \boldsymbol{0}$. By induction, assume that $\boldsymbol{B}$ has a singular value decomposition, $\boldsymbol{B} = \boldsymbol{U}_2\boldsymbol{\Sigma}_2\boldsymbol{V}_2^T$, such that

$$\boldsymbol{U}_1{}^T\boldsymbol{A}\boldsymbol{V}_1 = \left[\begin{array}{cc} \sigma_1 & \boldsymbol{0}^T\\ \boldsymbol{0} & \boldsymbol{U}_2\boldsymbol{\Sigma}_2\boldsymbol{V}_2^T \end{array}\right] = \left[\begin{array}{cc} 1 & \boldsymbol{0}^T\\ \boldsymbol{0} & \boldsymbol{U}_2 \end{array}\right] \left[\begin{array}{cc} \sigma_1 & \boldsymbol{0}^T\\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \end{array}\right] \left[\begin{array}{cc} 1 & \boldsymbol{0}^T\\ \boldsymbol{0} & \boldsymbol{V}_2^T \end{array}\right],$$

and the orthogonal matrices arising in the singular value decomposition of $\boldsymbol{A}$ *are*

$$\boldsymbol{U}=\boldsymbol{U}_1\begin{bmatrix} 1 & \boldsymbol{0}^T \\ \boldsymbol{0} & \boldsymbol{U}_2 \end{bmatrix}, \boldsymbol{V}^T=\begin{bmatrix} 1 & \boldsymbol{0}^T \\ \boldsymbol{0} & \boldsymbol{V}_2^T \end{bmatrix}\boldsymbol{V}_1^T.$$

□

The scaling coefficients σ_j are called the *singular values* of $\boldsymbol{A}$. The columns of $\boldsymbol{U}$ are called the *left singular vectors*, and those of $\boldsymbol{V}$ are called the *right singular vectors*.
The fact that the scaling coefficients are norms of $\boldsymbol{A}$ and submatrices of $\boldsymbol{A}$, $\sigma_1=\|\boldsymbol{A}\|$, is crucial importance in applications. Carrying out computation of the matrix products

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

leads to a representation of $\boldsymbol{A}$ as a sum

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

Each product $\boldsymbol{u}_i\boldsymbol{v}_i^T$ 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 $\boldsymbol{A}$

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

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

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 $\boldsymbol{b}$ in the canonical basis $\boldsymbol{I}\in\mathbb{R}^{m\times m}$ to coordinates $\boldsymbol{x}$ in some other basis $\boldsymbol{A}\in\mathbb{R}^{m\times m}$, a solution to the equation $\boldsymbol{I}\boldsymbol{b}=\boldsymbol{A}\boldsymbol{x}$ can be found by the following steps.

1. Compute the SVD, $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T=\boldsymbol{A}$;

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

Best 2-norm approximation. In the above $\boldsymbol{A}$ was assumed to be a basis, hence $r = \operatorname{rank}(\boldsymbol{A}) = m$. If columns of $\boldsymbol{A}$ do not form a basis, $r < m$, then $\boldsymbol{b} \in \mathbb{R}^m$ might not be reachable by linear combinations within $C(\boldsymbol{A})$. The closest vector to $\boldsymbol{b}$ 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, $y_i = c_i / \sigma_i$, $i = 1, \ldots, r$.

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

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

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

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

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