Basic Matrix and vector computations

$$(\mathbf{AB})_{ij} = \sum_{s=1}^{n} a_{is} b_{sj} \qquad\qquad (\mathbf{Ax})_i = \sum_{j=1}^{n} a_{ij} x_j$$

Theorem 1: The following results hold

i) Consider arbitrary $m \times j$, $j \times k$ and $k \times n$ matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. Then $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$.

ii) Consider arbitrary $m \times n$ matrix $\mathbf{A}$, n-dimensional vectors $\mathbf{x}$ and $\mathbf{y}$ and scalars a, b. We then have $\mathbf{A}(a\mathbf{x} + b\mathbf{y}) = a\mathbf{Ax} + b\mathbf{Ay}$

iii) Consider arbitrary $m \times n$ matrices $\mathbf{A}$, $\mathbf{B}$ and an arbitrary n-dimensional vectors $\mathbf{x}$. We then have that $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$ and $(\mathbf{Ax})^T = \mathbf{x}^T\mathbf{A}^T$.

Linearly independent vectors

Definition 3: The vectors $\mathbf{x}_1, \ldots, \mathbf{x}_K$ in $\mathbb{R}^n$ are **linearly independent** if the expression

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \ldots + c_K\mathbf{x}_K = 0$$

implies that

$$c_1 = c_2 = \ldots = c_K = 0$$

Definition 4: Let $\mathbf{A}$ be an arbitrary $m \times n$ matrix where $m \geq n$. Now perceive the columns of $\mathbf{A}$ as vectors in $\mathbb{R}^m$. The maximum number of linearly independent columns in $\mathbf{A}$ is denoted the **rank** of $\mathbf{A}$. Furthermore, if the n columns of $\mathbf{A}$ all are linearly independent $\mathbf{A}$ is said to have full rank, i.e. the rank is n.

If $\mathbf{A}$ does not have have full rank it is said to be **singular**.

Orthogonality and Orthonormality

Definition 5: The vectors $\mathbf{x}_1, \mathbf{x}_2$ in a vectorspace $\mathbb{R}^n$ are said to be **orthogonal** if $\mathbf{x}_1 \cdot \mathbf{x}_2 = 0$. Furthermore, if $\mathbf{x}_1 \cdot \mathbf{x}_1 = \mathbf{x}_2 \cdot \mathbf{x}_2 = 1$ then $\mathbf{x}_1, \mathbf{x}_2$ are said to be **orthonormal**.

Theorem 3: Orthonormal vectors are always linearly independent.

Definition 6: An $m \times n$ matrix $\mathbf{U}$ $(m \geq n)$ is said to be **column orthonormal** if it consists of n pairwise orthonormal columns (the columns are perceived as vectors in $\mathbb{R}^n$).

If $\mathbf{U}$ is an $m \times n$ column orthonormal matrix then

$$(\mathbf{U}^T\mathbf{U})_{ij} = \sum_{k=1}^{m} \mathbf{U}_{ki}\mathbf{U}_{kj} = \begin{cases} 1 \text{ if } i = j \\ 0 \text{ if } i \neq j \end{cases}$$

$$\mathbf{U}^T\mathbf{U} = \mathbf{I}_n$$

$\mathbf{I}_n$ is the $n \times n$ identitymatrix

Definition 7: An $n \times n$ matrix $\mathbf{V}$ is said to be **orthonormal** if it consists of n pairwise orthonormal columns (the columns are perceived as vectors in $\mathbb{R}^n$).

Subspace and Dimension

Definition 8: Consider $S \subseteq \mathbb{R}^n$. If for arbitrary elements $\mathbf{x}_1, \mathbf{x}_2 \in S$ and arbitrary scalars $c_1, c_2 \in \mathbb{R}$ applies that $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 \in S$ then S is said to be a *vector space*. The vector space S is also said to be a **subspace** of $\mathbb{R}^n$.

Notice: $\mathbb{R}^n$ is thus a vector space.

Definition 9: Let $S \subseteq \mathbb{R}^n$ be a vector space. The **dimension** of the vector space is the maximum number of linearly independent vectors, that can be found in S. Or stated differently: if $\mathbf{u}_1, \ldots, \mathbf{u}_K \in S$ are linearly independent and meet the condition that every $y \in S$ can be written as $y = c_1\mathbf{u}_1 + \ldots + c_K\mathbf{u}_K$ then S has dimension K. The vectors $\mathbf{u}_1, \ldots, \mathbf{u}_K$ is then said to form a **basis** for S. If $\mathbf{u}_1, \ldots, \mathbf{u}_K$ are orthonormal, $\mathbf{u}_1, \ldots, \mathbf{u}_K$ is said to form an orthonormal **basis** for S.

Notice: Hence $\mathbb{R}^n$ has dimension n.

Coordinates and Distance computations

Assume that the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ form an orthonormal basis for $\mathbb{R}^n$. We know then that an arbitrary $\mathbf{x}$ can be written as

$$\mathbf{x} = \alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \ldots \alpha_n \mathbf{u}_n$$

We call $(\alpha_1, \ldots, \alpha_n)$ the **coordinates** of $\mathbf{x}$ w.r.t. the base $\mathbf{u}_1, \ldots, \mathbf{u}_n$ We now immediately get for all i that

$$\mathbf{x} \cdot \mathbf{u}_i = \mathbf{x} \cdot (\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \ldots \alpha_n \mathbf{u}_n) = \alpha_i$$

Theorem 4: Let $\mathbf{u}_1 \ldots, \mathbf{u}_K$ be orthonormal vectors and let

$$\mathbf{x} = \sum_{k=1}^{K} \alpha_k \mathbf{u}_k$$

Then $\|\mathbf{x}\|^2 \equiv \mathbf{x} \cdot \mathbf{x} = \sum_{k=1}^{K} \alpha_k^2$. ”generalized Pythagorean Theorem”

Gram-Schmidt method

Assume that the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_k$ are linearly independent and let S be the subspace for which these vectors form a basis. Now we can construct a set of orthonormal vectors $\mathbf{e}_1, \ldots, \mathbf{e}_k$, which form an orthonormal basis for S in this way:

The Gram-Schmidt method:

$\mathbf{e}_1 := \mathbf{x}_1 / \|\mathbf{x}_1\|$

For $i := 2, \ldots, k$ do {

$\quad \mathbf{e}_i := \mathbf{x}_i - \sum_{j=1}^{i-1} (\mathbf{x}_i \cdot \mathbf{e}_j)\mathbf{e}_j$

$\quad \mathbf{e}_i := \mathbf{e}_i / \|\mathbf{e}_i\|$

}

Range, Null Space and the Least Squares solution

Definition 10: Let $\mathbf{A}$ be an arbitrary $m \times n$ $(m \geq n)$ matrix. The function $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is then said to be a **linear mapping** from $\mathbb{R}^n \to \mathbb{R}^m$. The **range** of the linear mapping is the set $B(\mathbf{A}) \subseteq \mathbb{R}^m$, that meets the condition that for any $\mathbf{y} \in B(\mathbf{A})$ exists an $\mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{A}\mathbf{x} = \mathbf{y}$. **The null space** of the linear mapping is the set $N(\mathbf{A}) \subseteq \mathbb{R}^n$, that meets the condition that for any $\mathbf{x} \in N(\mathbf{A})$ it applies that $\mathbf{A}\mathbf{x} = 0$.

Theorem 5 Let $\mathbf{u}_1 \ldots, \mathbf{u}_K$ be an arbitrary orthonormal basis for $B(\mathbf{A})$. Then the least squares solution $\mathbf{x}$ that minimizes $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$ satisfies

$$\mathbf{A}\mathbf{x} = \sum_{k=1}^{K} (\mathbf{b} \cdot \mathbf{u}_k)\mathbf{u}_k \equiv \mathbf{b}_{LS}$$

Singular Value Decomposition

Theorem 6 Consider an arbitrary $m \times n$ matrix $\mathbf{A}$. Then we can write $\mathbf{A}$ as $\mathbf{A} = \mathbf{U}\mathbf{W}\mathbf{V}^T$, where $\mathbf{U}$ is an $m \times n$ column othonormal matrix, $\mathbf{V}$ is an $n \times n$ orthonormal matrix and $\mathbf{W}$ is an $n \times n$ diagonal matrix having non-negative diagonal elements $w_1, \ldots w_n$ ordered such that $w_1 \geq w_2 \geq \ldots \geq w_n$. This is said to be a **Singular Value Decomposition** (SVD) of $\mathbf{A}$.

$$\begin{pmatrix} & & \\ & \mathbf{A} & \\ & & \end{pmatrix} = \begin{pmatrix} & & \\ & \mathbf{U} & \\ & & \end{pmatrix} \cdot \begin{pmatrix} w_1 & & & 0 \\ & w_2 & & \\ & & \ddots & \\ 0 & & & w_n \end{pmatrix} \cdot \begin{pmatrix} & \mathbf{V}^T & \end{pmatrix}$$

$\mathbf{A} = \mathbf{U}\mathbf{W}\mathbf{V}^T$, where $\mathbf{U}$ is an $m \times n$ column othonormal matrix, $\mathbf{V}$ is an $n \times n$ orthonormal matrix and $\mathbf{W}$ is an $n \times n$ diagonal matrix having non-negative diagonal elements $w_1, \ldots w_n$ ordered such that $w_1 \geq w_2 \geq ... \geq w_n$.

Theorem 7 Consider an arbitrary $m \times n$ matrix $\mathbf{A}$ and assume that for $\mathbf{W}$ it applies that $w_1, \ldots, w_K$ are positive and $w_{K+1}, \ldots, w_n$ are equal to zero. Then it applies that

i) $N(\mathbf{A})$ has dimension $n - K$ and the last $n - K$ columns of $\mathbf{V}$ form an orthonormal basis for $N(\mathbf{A})$.

ii) $B(\mathbf{A})$ has dimension K and the first K columns of $\mathbf{U}$ form an orthonormal basis for $B(\mathbf{A})$.

iii) The SVD solution $\mathbf{x} = \mathbf{V}\tilde{\mathbf{W}}^{-1}\mathbf{U}^T\mathbf{b}$, where $[\tilde{\mathbf{W}}^{-1}]_{jj} = 0$ if $\mathbf{W}_{jj} = 0$, otherwise $[\tilde{\mathbf{W}}^{-1}]_{jj} = 1/\mathbf{W}_{jj}$, is the least squares solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$. (Notice that it then follows that if all of the $\mathbf{W}_{jj}$'s are positive, i.e. $\mathbf{A}$ has full rank it applies that $\mathbf{x} = \mathbf{V}\mathbf{W}^{-1}\mathbf{U}^T\mathbf{b}$ is the least squares solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$).

$\mathbf{A} = \mathbf{UWV}^T$, where $\mathbf{U}$ is an $m \times n$ column othonormal matrix, $\mathbf{V}$ is an $n \times n$ orthonormal matrix and $\mathbf{W}$ is an $n \times n$ diagonal matrix having non-negative diagonal elements $w_1, \ldots w_n$ ordered such that $w_1 \geq w_2 \geq \ldots \geq w_n$.

IF !!! $w_1, \ldots, w_K$ are positive and $w_{K+1}, \ldots, w_n$ are equal to zero

i) $N(\mathbf{A})$ has dimension $n - K$ and the last $n - K$ columns of $\mathbf{V}$ form an orthonormal basis for $N(\mathbf{A})$.

ii) $B(\mathbf{A})$ has dimension K and the first K columns of $\mathbf{U}$ form an orthonormal basis for $B(\mathbf{A})$.

Proof: As $\mathbf{V}$ is orthonormal the columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are orthonormal and thus form an orthonormal basis for $\mathbb{R}^n$. Then we can write an arbitrary $\mathbf{x} \in \mathbb{R}^n$ as $\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n$. Now we can calculate

$$\begin{aligned}\mathbf{Ax} &= \mathbf{UWV}^T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n) \\ &= c_1\mathbf{UWV}^T\mathbf{v}_1 + c_2\mathbf{UWV}^T\mathbf{v}_2 + \ldots c_n\mathbf{UWV}^T\mathbf{v}_n\end{aligned}$$

$$\mathbf{A}(a\mathbf{x} + b\mathbf{y}) = a\mathbf{Ax} + b\mathbf{Ay}$$

We notice that $\mathbf{V}^T\mathbf{v}_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ where the 1 is at the i'th position. Insertion gives us

$$\mathbf{Ax} = c_1 w_1 \mathbf{u}_1 + c_2 w_2 \mathbf{u}_2 + \ldots c_n w_n \mathbf{u}_n$$

$\mathbf{A} = \mathbf{U}\mathbf{W}\mathbf{V}^T$, where $\mathbf{U}$ is an $m \times n$ column othonormal matrix, $\mathbf{V}$ is an $n \times n$ orthonormal matrix and $\mathbf{W}$ is an $n \times n$ diagonal matrix having non-negative diagonal elements $w_1, \ldots w_n$ ordered such that $w_1 \geq w_2 \geq \ldots \geq w_n$.

iii) The SVD solution $\mathbf{x} = \mathbf{V}\tilde{\mathbf{W}}^{-1}\mathbf{U}^T\mathbf{b}$, where $[\tilde{\mathbf{W}}^{-1}]_{jj} = 0$ if $\mathbf{W}_{jj} = 0$, otherwise $[\tilde{\mathbf{W}}^{-1}]_{jj} = 1/\mathbf{W}_{jj}$, is the least squares solution to $\mathbf{A}\mathbf{x} = \mathbf{b}$.

In order to show iii), i.e. that the SVD solution is the same as the least squares solution we exploit that we know that the first K columns of $\mathbf{U}$ form an orthonormal basis for the range of $\mathbf{A}$. I.e. that the nearest point in $B(\mathbf{A})$ (the least squares mapping) according to Theorem 5 is given by $\mathbf{b}_{LS} = \sum_{j=1}^{K}(\mathbf{u}_j \cdot \mathbf{b})\mathbf{u}_j$. Now let $\mathbf{x}$ be the SVD solution. I.e.

$$\mathbf{A}\mathbf{x} = \sum_{k=1}^{K}(\mathbf{b} \cdot \mathbf{u}_k)\mathbf{u}_k \equiv \mathbf{b}_{LS}$$

$$\begin{aligned}
\mathbf{A}\mathbf{x} &= (\mathbf{U}\mathbf{W}\mathbf{V}^T)(\mathbf{V}[\tilde{\mathbf{W}}^{-1}]\mathbf{U}^T b) = \mathbf{U}(\mathbf{W}[\tilde{\mathbf{W}}^{-1}]\mathbf{U}^T b) \\
&= [\mathbf{u}_1 \ldots \mathbf{u}_K | \mathbf{u}_{K+1} \ldots \mathbf{u}_n] \begin{bmatrix} \mathbf{u}_1 \cdot \mathbf{b} \\ \mathbf{u}_2 \cdot \mathbf{b} \\ \cdots \\ \mathbf{u}_K \cdot \mathbf{b} \\ --- \\ 0 \\ \cdots \\ 0 \end{bmatrix} = \sum_{j=1}^{K}(\mathbf{u}_j \cdot \mathbf{b})\mathbf{u}_j = \mathbf{b}_{LS}
\end{aligned}$$