Chapter 1: Mathematical Background: Linear Algebra

Konrad Koniarski

2016/07/09

1 Vector Spaces

1.1 Vector space

A set V is called a **linear space** or a **vector space over the field** $\mathbb{R}$ if it is closed under vector summation

$$+ : V \times V \to V \tag{1}$$

and under scalar multiplication

$$\cdot : \mathbb{R} \times V \to V \tag{2}$$

i.e. $\alpha v_1 + \beta v_2 \in V, \forall v_1, v_2 \in V, \forall \alpha, \beta \in \mathbb{R}$. With respect to addition $(+)$ it forms a commutative group (existence of neutral element 0, inverse element $-v$). Scalar multiplication respects the structure of $\mathbb{R} : \alpha(\beta u) = (\alpha\beta)u$. Multiplication and addition respect the distributive law:

$$(\alpha + \beta)v = \alpha v + \beta v \text{ and } \alpha(v + u) = \alpha v + \alpha u \tag{3}$$

Example: $V = \mathbb{R}^n, v = (x_1, ..., x_n)^T$.

A subset $W \subset V$ of a vector space V is called **subspace** if $0 \in W$ and W is closed under $+$ and $\cdot$ (for all $\alpha \in \mathbb{R}$).

1.2 Linear Independence and Basis

The spanned subspace of a set of vectors $S = \{v_1, ..., v_k\} \subset V$ is the subspace formed by all linear combinations of these vectors:

$$span(S) = \{v \in V | v = \sum_{i=1}^{k} \alpha_i v_i\} \tag{4}$$

The set S is called **linearly independent** if:

$$\sum_{i=1}^{k} \alpha_i v_i = 0 \Rightarrow \alpha_i = 0 \forall i, \tag{5}$$

in other words: if none of the vectors can be expressed as a linear combination of the remaining vectors. Otherwise the set is called **linearly dependent**.

A set of vectors $B = \{v_1, ..., v_n\}$ is called a basis of V if it is linearly independent and if it spans the vector space V. A basis is a maximal set of linearly independent vectors.

1.3 Properties of Basis

Let B and B' be two bases of a linear space V.

- B and B' contains the same number of vectors. This number n is called the **dimension of the space V**.
- Any vector $v \in V$ can be uniquely expressed as a linear combination of the basis vectors in $B = \{b_1, ..., b_n\}$:

$$v = \sum_{i=1}^{n} \alpha_i b_i \tag{6}$$

- In particular, all bectors of B can be expressed as linear combinations of vectors of another basis $b'_i \in B'$:

$$b'_i = \sum_{j=1}^{n} \alpha_(ji) b_j \tag{7}$$

  The coefficients $\alpha_(ji)$ for this **basis transform** can be combined in a matrix A. Set $B \equiv (b_1, ..., b_n)$ and $B' \equiv (b'_1, ..., b'_n)$ as the matirces of basis vectors, we can write: $B' = BA \Leftrightarrow B = B'A^-1$.

1.4 Inner Product

On a vector space one can define an **inner product (dot product)**

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R} \tag{8}$$

which is defined by three properties:

- $\langle u, \alpha v + \beta w \rangle = \alpha \langle u, v \rangle + \beta \langle u, w \rangle$ (linear)
- $\langle u, v \rangle = \langle v, u \rangle$ (symmetric)
- $\langle v, v \rangle \geq 0$ and $\langle v, v \rangle = 0 \Leftrightarrow v = 0$ (positive definite)

The scalar product induces a **norm**

$$| \cdot | : V \to \mathbb{R}, |v| = \sqrt{\langle v, v \rangle} \tag{9}$$

and a **metric**

$$d : V \times V \to \mathbb{R}, d(v, w) = |v - w| = \sqrt{v - w, v - w} \tag{10}$$

for measuring lengths and distances, making V a **metric space**. Since the metric is inducted by a scalar product V is called a **Hilbert space**.

1.5 Canonical and Induced Inner Product

ON $V = \mathbb{R}^n$, one can define the canonical inner product for the canonical basis $B = I_n$ as

$$\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i \tag{11}$$

which induces the standard L_2-norm or Euclidean norm

$$|x|_2 = \sqrt{x^T y} = \sqrt{x_1^2 + ... + x_n^2} \tag{12}$$

With a basis transform A to the new basis B' given by $I = B'A^-1$ the canonical inner product in the new coordinates x',y' is given by:

$$\langle x, y \rangle = x^T y = (Ax')^T (Ay') = x'^T A^T Ay' \equiv \langle x', y' \rangle_{A^T A} \tag{13}$$

The latter product is called the inducted inner product from the matrix A.

Two vectors v and w are orthogonal iff $\langle v, w \rangle = 0$.

2 Properties and Matrices

2.1 Kronecker Product and Stack of a Matrix

Given two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{k \times l}$, one can define their **Kronecker product** $A \otimes B$ by :

$$A \otimes B \equiv \begin{pmatrix} a_{1,1}B & \cdots & a_{1,n}B \\ \vdots & \ddots & \vdots \\ a_{m,1}B & \cdots & a_{m,n}B \end{pmatrix} \in \mathbb{R}^{mn} \tag{14}$$

In Matlab this can be implemented by C=kron(A,B). Given a matrix $A \in \mathbb{R}^{m \times n}$, its **stack** A^s is obtained by stacking its n columns vectors $a_1, ..., a_n in \mathbb{R}^m$:

$$A^s \equiv \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{R}^{mn} \tag{15}$$

These notations allows to rewrite albegebraic expressions, for example

$$u^T A v = (v \otimes u)^T A^s \tag{16}$$

2.2 Linear Transformation and Matrices

Linear algebra studies the properties of linear transformations between linear spaces. Since these can be represented by matrices, linear algebra studies the properties of matrices. A **linear transformation** L between two linear spaces V and W is a map $L : V \to W$ such that:

- $L(x + y) = L(x) + L(y) \forall x, y \in V$
- $L(\alpha x) = \alpha L(x) \forall x \in V, \alpha \in \mathbb{R}$.

Due to the linearity, the action of L on the space V is uniquely defined by its action on the basis vectors of V. In the canonical basis $e_1, ..., e_n$ we have

$$L(x) = Ax \forall x \in V \tag{17}$$

where

$$A = (L(e_1), ..., L(e_n)) \in \mathbb{R}^{m \times n} \tag{18}$$

The set of all real $m \times n$-matrices is denoted by $\mathcal{M}(m,n)$. In the case that $m = n$, the set $\mathcal{M}(m,n) \equiv \mathcal{M}(n)$ form a **ring** over the field $\mathbb{R}$, i.e. it is closed under matrix multiplication and summation.

2.3 The Linear Groups $GL(n)$ and $SL(n)$

There exist certain sets of linear trasnforamations which form a group. A group is a set G with an operation $\circ : G \times G \to G$ such that:

- $g_1 \circ g_2 \in G \forall g_1, g_2 \in G$ (closed)
- $(g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3) \forall g_1, g_2, g_3 \in G$ (assoc.)
- $\exists e \in G : e \circ g = g \circ e = e \forall g \in G$ (neutral)
- $\exists g_{-1} \in G : g \circ g^{-1} = g^{-1} \circ g = e \forall g \in G$ (inverse)

Example: All invertable (non-singular) real $x \times n$-matrices form a group with respect to matrix multiplication. The group is called the **general linear group** $GL(n)$. It consists of all $A \in \mathcal{M}(n)$ for which

$$det(A) \neq 0 \tag{19}$$

All matrices $A \in GL(n)$ for which det $det(A) = 1$ form a group called the **special linear group** $SL(n)$. The inverse of A is also in this group, as $det(A^{-1}) = det(A)^{-1}$

2.4 Matrix Representation of Groups

A group G has a **matrix representation** of can be realized as a matrix group if there exists an injective transformation:

$$R : G \to GL(n) \tag{20}$$

which **preserves the group structure** of G, that is inverse and composition are preserved by the map:

$$R(e) = I_{n \times n}, R(g \circ h) = R(g)R(h) \forall g, h \in G \tag{21}$$

Such a map R is called a **group homomorphism**. The idea of matrix representations of a group is that they allow to analyze more abstract groups by looking at the properties of the respective matrix group.
Example: The rotations of an object form a group, as there exists a neutral elements (no rotation) and an inverse (the inverse rotation) and any concatenation of rotations in again a rotation (ariybd a different axis). Studying the properties of the rotation group is easier if rotation are represented by respective matrices.

2.5 The Affine Group $A(n)$

An affine transformation $L : \mathbb{R}^n \to \mathbb{R}^n$ is defined by a matrix $A \in GL(n)$ and a vector $b \in \mathbb{R}^n$ such that:

$$L(x) = Ax + b \tag{22}$$

The set of all such affine transformations is called the **affine group of dimension n**, denoted by $A(n)$. L defined above is not a linear map unless $b = 0$. By introducing **homogeneous coordinates** to represent $x \in \mathbb{R}^n$ by $\begin{pmatrix} x \\ 1 \end{pmatrix} \in \mathbb{R}^{n+1}$, L becomes a linear mapping from

$$L : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}; \begin{pmatrix} x \\ 1 \end{pmatrix} \to \begin{pmatrix} A & b \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} \tag{23}$$

A matrix $\begin{pmatrix} A & b \\ 0 & 1 \end{pmatrix}$ with $A \in GL(n)$ and $b \in \mathbb{R}^n$ is called an **affine matrix**. It is an element of $GL(n+1)$. The affine matrices form a subgroup of $GL(n+1)$.

2.6 The Orthogonal Group $O(n)$

A matrix $A \in \mathcal{M}(n)$ is called orthogonal if it preserves the inner product, i.e.:

$$\langle Ax, Ay \rangle = \langle x, y \rangle, \forall x, y \in \mathbb{R}^n \tag{24}$$

The set of all orthogonal matrices forms the orthogonal group O(n), which is a subgroup of GL(n). For an orthogonal matrix R we have

$$\langle Rx, Ry \rangle = x^T R^T R y = x^T y, \forall x, y \in \mathbb{R}^n \tag{25}$$

Therefore we must have $R^T R = R R^T = I$, in other words:

$$O(n) = \{R \in GL(n) | R^T R = I\} \tag{26}$$

The above identity shows that for an orthogonal matrix R, we have $det(R^T R) = (det(R))^2 = det(I) = 1$, such that $det(R) \in \{\pm 1\}$.

The subgroup of $O(n)$ with $det(R) = \pm 1$ is called the **special orthogonal group** $SO(n)$. $SO(n) = O(n) \cap SL(n)$. In particular, $SO(3)$ is the group of all 3-dimensional rotation matrices.

2.7 The Euclidean Group $E(n)$

A Euclidean transformation L from $\mathbb{R}^n$ to $\mathbb{R}^n$ is defined by an orthogonal matrix $\mathbb{R} \in O(n)$ and a vector $T \in \mathbb{R}^n$:

$$L : \mathbb{R}^n \to \mathbb{R}^n; x \to Rx + T \tag{27}$$

The set of all such transformations is called the Euclidean group E(n). It is a subgroup of the affine group A(n). Embedded by homogeneous coordinates, we get:

$$E(n) = \{ \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} | R \in O(n), T \in \mathbb{R}^n \} \tag{28}$$

if $R \in SO(n)(i.e. det(R) = 1)$. then we have the **special Euclidean group** $SE(n)$. In particular, $SE(3)$ represents the rigid-body motions in $\mathbb{R}^3$.

In Summary:

$$\boxed{SO(N) \subset O(n) \subset GL(n), SE(n) \subset E(n) \subset A(n) \subset GL(n+1).} \tag{29}$$

3 Linear Transformations and Matrices

3.1 Range, Span, Null Space and Kernel

Let $A \in \mathbb{R}^{m \times n}$ be a matrix defining a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$. The **range** or **span** of A is defined as subspace of $\mathbb{R}^m$ which can be 'reached' by A:

$$range(A) = y \in \mathbb{R}^m | \exists x \in \mathbb{R}^n : Ax = y \tag{30}$$

The range of a matrix A is given by the span of its column vectors. The **null space** or **kernel** of a matrix A is given by the subset of vectors $x \in \mathbb{R}^n$ which are mapped to zero:

$$null(A) \equiv ker(A) = \{x \in \mathbb{R}^n | Ax = 0\} \tag{31}$$

The null space of a matrix A is given by the vectors orthogonal to its row vectors. Matlab: Z = null(A). The concepts of range and null space are useful when studing the **solution of linear equations**. The system $Ax = b$ will have a solution $x \in \mathbb{R}^n$ if and only if $b \in range(A)$. Moreover, this solution will be unique only if $ker(A) = \{0\}$. Indeed, if x_s is a solution of $Ax = b$ and $x_0 \in ker(A)$, then $x_s + x_0$ is also a solution:$A(x_s + x_0) = Ax_s + Ax_0 = b$.

3.2 Rank of a Matrix

The rank of a matrix is the dimension of its range:

$$rank(A) = dim(range(A)) \tag{32}$$

The rank of a matrix $A \in \mathbb{R}m \times n$ has the following properties:

- $rank(A) = n - dim(ker(A))$.
- $0 \geq rank(A) \geq min\{m, n\}$.
- $rank(A)$ is equal to the maximum number of linearly independent row (or column) vector of A
- $rank(A) is the highest order of a non-zero minor of A, where a minor of order k is the determinant of a k \times k submatrix of A$
- Sylvester's inequality: Let $B \in \mathbb{R}^{n \times k}$. Then $AB \in \mathbb{R}^{m \times k}$ and $rank(A) + rank(B) - n \leq rank(AB) \leq min\{rank(A), rank(B)\}$.
- For any nonsingular matrices $C \in \mathbb{R}^{m \times m}$ and $D \in \mathbb{R}^{n \times n}$, we have : $rank(A) = rank(CAD)$.

3.3 Eigenvalues and Eigenvectors

Let $A \in \mathbb{C}^{n\times n}$ be a complex matrix, A non-zero vector $v \in \mathbb{C}^n$ is called a **(right)eigenvector of A** if :

$$Av = \lambda v, \text{ with } \lambda \in \mathbb{C} \tag{33}$$

λ is called an **eigenvalue of** A. Similarly v is called a **left eigenvector** of A, if $v^T A = \lambda v^T$ for some $\lambda \in \mathbb{C}$. The **spectrum** $\sigma(A)$ **of a matrix A** is the set of all its eigenvalues.

$$\textbf{Matlab: [V,D] = eig(A);} \tag{34}$$

where D is a diagonal matrix containing the eigenvalues and V is a matrix whose columns are the correponding eigenvectors such that $AV = VD$

3.4 Properties of Eigenvalues and Eigenvectors

Let $A \in \mathbb{R}^{n\times n}$ be a square matrix. Then:

- If $Av = \lambda v$ for some $\lambda \in \mathbb{R}$, then there also exists a left-eigenvector $\eta \in \mathbb{R} : \eta^T A = \lambda \eta^T$
- The eigenvectors of a matrix A associated with different eigenvalues are linearly independent.
- All eigenvalues $\sigma(A)$ are the root of the characteristic polynomial equation $det(\lambda I - A) = 0$. Therefore det(A) is equal to the product of all eigenvalues (some of which may appear multiple times).
- if $B = PAP^{-1}$ for some nonsingular matrix P, then $\sigma(B) = \sigma(A)$.
- if $\lambda \in \mathbb{C}$ is an eigenvalues, then its conjugate $\bar{\lambda}$ is also an eigenvalue. Thus $\sigma(A) = \overline{\sigma(A)}$ for real matrices A.

3.5 Symetric Matrices

A matrix $S \in \mathbb{R}^{n\times n}$ is called **symmetric** if $S^T = S$. A symmetric matrix S is called **positive semi-defined (denoted by** $S \geq 0$ **or** $S \succeq 0$**)** if $x^T Sx \geq 0$. S is called **positive defined (denoted by** $S > 0$ **or** $S \succ 0$**)** if $x^T Sx > 0 \forall x \neq 0$. **Let** $S \in \mathbb{R}^{n\times n}$ **be a real symmetric matrix**. Then:

- All eigenvalues of S are real, i.e. $\sigma(S) \subset \mathbb{R}$.
- Eigenvectors v_i and v_j of S corresponding to distinct eigenvalues $\lambda_i \neq \lambda_j$ are orthogonal.
- There always exist n ortonormal eigen vectors of S which form a basis of $\mathbb{R}^n$. Let $V = (v_1, ..., v_n) \in O(n)$ be the orthogonal matrix of these eigenvectors and $\Lambda = diag\{\lambda_1, ..., \lambda_n\}$ the diagonal matrix of eigenvalues. Then we have $S = V\Lambda V^T$.
- S is positive (semi-)definite, if all eigenvalues are positive (nonnegative).
- Let S be positive semi-defined and λ_1, λ_n the largest and smallest eigenvalue. Then $\lambda_1 = max_{|x|=1} \rangle x, Sx \langle$ and $\lambda_n = min_{|x|=1} \langle x, Sx \rangle$.

3.6 Norms and Matrices

There are many ways to define norms on the space of matrices $A \in \mathbb{R}^{m\times n}$. They canbe defined based on norms on the domain or codomain spaces on which A operates. In particular, the **induced 2-norm of a matrix A** is defined as

$$||A||_2 \equiv \max_{|x|_2=1} |Ax|_2 = \max_{|x|_2=1} \sqrt{\langle x, A^T Ax, \rangle} \tag{35}$$

Alternatively, one can define the **Frobenius norm of A** as:

$$||A||_f \equiv \sqrt{\sum_{i,j} a_i j^2} = \sqrt{trace(A^T A)} \tag{36}$$

Note that these norms are in general not the same. Since the matrix $A^T A$ is symmetric and pos. semi-definite, we can diagonalize it as $A^T A = V diag\{\sigma_1^2, ..., \sigma_n^2\} V^T$ with $\sigma_1^2 \geq \sigma_i^2 \geq 0$. This leads to:

$$||A||_2 = \sigma_1, \text{ and } ||A||_f = \sqrt{trace(A^T A)} = \sqrt{\sigma_1^2 + ... + \sigma_n^2} \tag{37}$$

3.7 Skew-symmetric Matrices

A matrix $A \in \mathbb{R}^{n \times n}$ is called **skew-symetric** or **anti-symetric** if $A^T = -A$

If A is a real skew-symetric matrix, then:

- All eigenvalues of A are either zero or purely imaginary, i.e. of the form $i\omega$ with $i^2 = -1, \omega \in \mathbb{R}$.
- There exists an orthogonal matrix V such that

$$A = V \Lambda V^T \tag{38}$$

  where Λ is a block-diagonal matrix $\Lambda = diag\{A_1, ..., A_m, 0, ..., 0\}$, with real skew-symmetric matrices A_i of the form:

$$A_i = \begin{pmatrix} 0 & a_i \\ -a_i & 0 \end{pmatrix} \in \mathbb{R}^{2 \times 2}, i = 1, ..., m \tag{39}$$

In particular, the rank of any skew-symmetric matrix is even.

3.8 Example of Skew-symetric matrices

In Computer Vision, a common skew-symmetric matrix is given by the **hat operator** of a vector $u \in \mathbb{R}^3$ is :

$$\hat{u} = \begin{pmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{pmatrix} \in \mathbb{R}^{3 \times 3} \tag{40}$$

This is a linear operator from the space of vectors $\mathbb{R}^3$ to the space of skew-symetric matrices in $\mathbb{R}^{3 \times 3}$. In particular, the matrix $\hat{u}$ has the property that

$$\hat{u}v = u \times v \tag{41}$$

where $\times$ denotes the standard vector cross product in $\mathbb{R}^3$. For $u \neq 0$, we have $rank(\hat{u}) = 2$ and the null space of $\hat{u}$ is spanned by u, because $\hat{u}u = u^T \hat{u} = 0$.

3.9 The Singular Value Decomposition (SVD)

Many of matrices properties such as rank, range, null space, and induced norms of matrices can be captured by the so-called singular value decomposition (SVD).

SVD can be seen as a generalization of eigenvalues and eigenvectors to non-square matrices. The computation of SVD is numericall well-conditioned. It is very usefull for solving linear-algebraic problems such as matric inversion, rank, computation, linear least-squares estimation, projections, and fixed-rank approximations. In practice, both singular value decomposition and eigenvalues decompositions are used quite extensively.

3.10 Algebraic Definition of SVD

Let $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ be a matrix of $rank(A) = p$. Then there exist:

- $U \in \mathbb{R}^{m \times p}$ whose columns are orthonormal
- $V \in \mathbb{R}^{n \times p}$ whose columns are orthonormal, and
- $\Sigma \in \mathbb{R}^{p \times p}, \Sigma = diag \sigma_1, ..., \sigma_p$, with $\sigma_1 \geq ... \geq \sigma_p$, such that

$$A = U \Sigma V^T \tag{42}$$

Note that this generalizes the eigenvalue decomposition. While the latter decomposes a symmetric square matrix A with an orthogonal transformation V as :

$$A = V \Lambda V^T, \text{ with } V \in O(n), \Lambda = diag\{\lambda_1, ..., \lambda_n\} \tag{43}$$

SVD allows to decompose an arbitrary (non-square) matrix A of rank p with two transformations U and V with orthonormal columns as shown above. Nevertheless, we will see that SVD is based on the eigenvalue decomposition of symmetric square matrices.

3.11 Proof of SVD Decomposition 1

Geven a matrix $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ and $rank(A) = p$, the matrix

$$A^T A \in \mathbb{R}^{n \times n} \tag{44}$$

is symmetric and positive semi-definite. Therefore it can be decomposed with non-negative eigenvalues $\sigma_1^2 \geq ... \geq \sigma_n^2 \geq 0$ with ortonormal eigenvectors $v_1, ..., v_n$. The σ_i are called **singular values**. Since

$$ker(A^T A) = ker(A) \text{ and } range(A^T A) = range(A^T), \tag{45}$$

we have $span\{v_1, ..., v_p\} = range(A^T)$ and $span\{v_{p+1}, ..., v_n\} = ker(A)$. Let

$$u_i \equiv \frac{1}{\sigma_i} A v_i \Leftrightarrow A v_i = \sigma_i u_i, i = 1, ..., p \tag{46}$$

then the $u_i \in \mathbb{R}^m$ are orthonormal"

$$\langle u_i, u_j \rangle = \frac{1}{\sigma_i \sigma_j} \langle A v_i, A v_j \rangle = \frac{1}{\sigma_i \sigma_j} \langle v_i, A^T A v_j \rangle = \delta_{ij} \tag{47}$$

3.12 Proof of SVD Decomposition 2

Complete $\{u_i\}_{i=0}^{p}$ to a basis $\{u_i\}_{i=1}^{m}$ of $\mathbb{R}^m$. Since $Av_i = \sigma_i u_i$, we have

$$A(v_1, ..., v_n) = (u_1, ..., u_m) \begin{pmatrix} \sigma_1 & 0 & 0 & \dots & 0 \\ 0 & \ddots & 0 & \vdots & 0 \\ 0 & \dots & \sigma_p & \vdots & 0 \\ \vdots & \dots & \dots & \vdots & 0 \\ 0 & \dots & \dots & 0 & 0 \end{pmatrix} \tag{48}$$

which is one of the form $A\tilde{V} = \tilde{U}\tilde{\Sigma}$, thus

$$A = \tilde{U}\tilde{\Sigma}\tilde{V}^T \tag{49}$$

Now simply delete all columns of $\tilde{U}$ and the rows of $\tilde{V}^T$ which are multiplied by zero singular values and we obtain the form $A = U\Sigma V^T$, with $U \in \mathbb{R}^{m\times p}$ and $V \in \mathbb{R}^{n\times p}$.

In Matlab: [U,S,V] = svd(A).

3.13 A Geometric Interpretation of SVD

For $A \in \mathbb{R}^{n\times n}$, the singular value decomposition $A = U\Sigma V^T$ is such that the columns $U = (u_1, ..., u_n)$ and $V = (v_1, ..., v_n)$ form orthonormal bases on $\mathbb{R}^n$. If a point $x \in \mathbb{R}^n$ is mapped to a point $y \in \mathbb{R}^n$ by the tranformation A, then the coordinated of y is basis U are related to the coordinates of x in basis V by the diagonal matrix Σ: each coordinate is merely scaled by the corresponding singular value:

$$y = Ax = U\Sigma V^T x \Leftrightarrow U^T y = \Sigma V^T x. \tag{50}$$

The matrix A maps the unit sphere into an ellipsoid with semi-axes $\sigma_i u_i$. To see this, we call $\alpha \equiv V^T x$ the coefficients of the point x in the basis V and those of y in basis U shall be called $\beta \equiv U^T y$. All points of the circle fulfill $|x|_2^2 = \sum_i^{\alpha_i^2} = 1$. The above statement says that $\beta_i = \sigma_i \alpha_i$. Thus for the points on the sphere we have

$$\sum_i \alpha_i^2 = \sum_i \beta_i^2/\sigma_i^2 = 1 \tag{51}$$

which states that the transformated points lie on an ellipsoid oriented along the axes of the basis U.

3.14 The Generalized (More Penrose) Inverse

For certain quadratic matrices one can define an inverse matrix, if $det(A) \neq 0$. The set of all invertible matrices forms the group GL(n). One can also define a **(generalized) inverse** (also called **pseudo inverse**) for an arbitraty (non-quadratic) matrix $A \in \mathbb{R}^{m\times n}$. if its SVD is $A = U\Sigma V^T$ the pseudo inverse is defined as:

$$A^\dagger = V\Sigma^\dagger U^T, \text{ where } \Sigma^\dagger = \begin{pmatrix} \Sigma_1^{-1} & 0 \\ 0 & 0 \end{pmatrix}_{n\times m} \tag{52}$$

where Σ_1 is the diagonal matrix of non-zero singular values. In **Matlab: X = pinv(A)**. In particular, the pseudo inverse can be employed in a similar fashion as the inverse of quadratic invertible matrices:

$$AA^\dagger A = A, A^\dagger AA^\dagger = A^\dagger \tag{53}$$

The linear system $Ax = b$ with $A \in \mathbb{R}^{m\times n}$ of rank $r \leq min(m, n)$ can have multiple or no solutions. $x_{min} = A^\dagger b$ **is among all minimizers of** $|Ax - b|^2$ **the only with the smallest norm** $|x|$**.**