Mathematical Foundations of Computer Vision
(Revised Script)

Michael Breuß

Contents

Chapter 1

Three-Dimensional Space

We identify a *point* p in the three-dimensional (3-D) space with the *coordinates*

$$\vec{X} \; := \; (X_1, X_2, X_3)^\top \; = \; \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \; = \; \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \in \mathbb{R}^3 \tag{1.1}$$

We will talk about points and coordinates as if they were the same thing.

The 3-D space can be represented by a *Cartesian coordinate frame* which means that the coordinate axes are pairwise right-angled.

We are now heading for measuring distances and angles. To this end, we need vectors and an inner product defining a metric.

1.1 Vectors

We need some definitions.

Definition 1.1.1 (vector, base point) *A vector v is determined by a pair of points p, q, and is defined as the directed arrow connecting p to q, denoted $v = \vec{pq}$. The point p is the base point of v.*

If p and q have coordinates $\vec{X}$ and $\vec{Y}$, respectively, then v has coordinates

$$v \; := \; \vec{Y} - \vec{X} \; \in \mathbb{R}^3 \tag{1.2}$$

We distinguish the two concepts of a *bound vector* (as above) and a *free vector*, a vector which does not depend on its base point.

To make the latter precise, if we have two pairs of points (p, q) and (p', q') with coordinates satisfying

$$\vec{Y} - \vec{X} \;=\; \vec{Y}' - \vec{X}' \tag{1.3}$$

they define the same free vector. For a free vector v, one may safely assume that the base point is the origin of the Cartesian frame:

$$\vec{X} \;=\; \vec{0} \qquad \text{and} \qquad \vec{Y} \;=\; v \tag{1.4}$$

The set of all free vectors forms a *vector space*. As an example for this, one may think of the $\mathbb{R}^3$, where the linear combination of two vectors $v := (v_x, v_y, v_z)^\top$ and $u := (u_x, u_y, u_z)^\top$ using scalars α, β in $\mathbb{R}$ is defined by

$$\alpha v + \beta u \;=\; \begin{pmatrix} \alpha v_x + \beta u_x \\ \alpha v_y + \beta u_y \\ \alpha v_z + \beta u_z \end{pmatrix} \tag{1.5}$$

For convenience, we recall the underlying definition of a vector space.

Definition 1.1.2 (vector space) *A set of vectors V is a vector space over $\mathbb{R}$ if the following structural requirements are met:*

Closedness

(i) *Given any v_1, v_2 in V and any α, β in $\mathbb{R}$,*
the linear combination $v = \alpha v_1 + \beta v_2$ is again a vector in V.

Addition of vectors

(ii) *The addition of vectors is commutative*

(iii) *and associative;*

(iv) *there is an identity element $\vec{0}$ w.r.t. addition, i.e. $v + \vec{0} = v$;*

(v) *there is an inverse element '$-v$', such that $v + (-v) = \vec{0}$ for any v.*

Multiplication with a scalar respects the structure of $\mathbb{R}$

(vi) $\alpha\,(\beta v) \;=\; (\alpha\beta)\,v$

(vii) $1v = v$

(viii) $0v = \vec{0}$

Addition and multiplication with a scalar are related by the distributive laws

(ix) $(\alpha + \beta)\,v \;=\; \alpha v + \beta v$

(x) $\alpha\,(v + u) \;=\; \alpha v + \alpha u$

1.2 Linear Independence and Change of Basis

We now review important basic notions associated with a vector space V. We may silently assume $V = \mathbb{R}^n$.

Definition 1.2.1 (subspace) *A subset W of a vector space V is called a subspace if*

(i) *the zero vector $\vec{0}$ is in W and*

(ii) *it holds $\alpha w_1 + \beta w_2 \in W$ for all α, β in $\mathbb{R}$ and any w_1, w_2 in W.*

Definition 1.2.2 (span) *Given a set of vectors $S = \{v_i\}_{i=1}^m$, the subspace spanned by S is the set of all linear combinations*

$$\text{span}\,(S) \;:=\; \sum_{i=1}^{m} \alpha_i v_i \tag{1.6}$$

for $\alpha_1, \ldots, \alpha_m \in \mathbb{R}$.

Example 1.2.1 The two vectors $v_1 = (1,0,0)^\top$ and $v_2 = (0,1,0)^\top$ span a subspace of $\mathbb{R}^3$ whose vectors are of the general form $v = (x,y,0)^\top$.

Definition 1.2.3 (linear dependence/independence) *A set of vectors $S = \{v_i\}_{i=1}^m$ is linearly independent if the equation*

$$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_m v_m \;=\; \vec{0} \tag{1.7}$$

implies $\alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$. The set S is said to be linearly dependent if there exist $\alpha_1, \alpha_2, \ldots, \alpha_m$ not all zero such that (1.7) is satisfied.

Definition 1.2.4 (basis) *A set of vectors $B = \{b_i\}_{i=1}^n$ of a vector space V is said to be a basis if B is a linearly independent set and if* $\text{span}(B) = V$.

Let us now suppose that B and B' are two bases of a linear space V. Then:

1. B and B' contain exactly the same number of linearly independent vectors. This number, say n, is the *dimension* of V.

2. Let $B = \{b_i\}_{i=1}^n$ and $B' = \{b'_i\}_{i=1}^n$. Then each basis vector of B can be expressed as

$$b_j \;=\; a_{1j}b'_1 + a_{2j}b'_2 + \ldots + a_{nj}b'_n \;=\; \sum_{i=1}^{n} a_{ij}b'_i \tag{1.8}$$

with some $a_{ij} \in \mathbb{R}$.

3. Any vector $v \in V$ can be written in terms of either of the bases:

$$v \;=\; x_1b_1 + x_2b_2 + \ldots + x_nb_n \;=\; x'_1b'_1 + x'_2b'_2 + \ldots + x'_nb'_n \tag{1.9}$$

where the coefficients $x_i, x'_i \in \mathbb{R}$ are the coordinates of v w.r.t. each basis. We denote by $x := (x_1, x_2, \ldots, x_n)^\top$ and $x' := (x'_1, x'_2, \ldots, x'_n)^\top$ the corresponding *coordinate vectors.*

We may arrange the basis vectors $B = \{b_i\}_{i=1}^n$ and $B' = \{b'_i\}_{i=1}^n$ as columns of two $n \times n$ matrices and also call them B and B', respectively:

$$B \;:=\; [b_1, b_2, \ldots, b_n] \in \mathbb{R}^{n\times n} \quad \text{and} \quad B' \;:=\; [b'_1, b'_2, \ldots, b'_n] \in \mathbb{R}^{n\times n} \tag{1.10}$$

Then we can express as by (1.8) the *basis transformation* between B and B' via

$$[b_1, b_2, \ldots, b_n] \;=\; [b'_1, b'_2, \ldots, b'_n] \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \text{ i.e. } B = B'A \tag{1.11}$$

Let us comment, that the practical role of the basis transformation is to encode how coordinates change when going over from one basis to another.

In this context, let us elaborate on the role of the matrix $A = (a_{ij}) \in \mathbb{R}^{n\times n}$. For this, let us apply both sides of the equation in (1.11) at a coordinate vector x given in the 'old' basis B:

$$Bx \;=\; B'Ax \tag{1.12}$$

To emphasize the rough idea of what follows, let us understand B and B' as two different parameterizations of the same space, and x is given in terms of the 'old' parameterization B. We highlight making use of two brackets the usual ordering in which operations are carried out on the right hand side:

$$Bx \;=\; B'\,(Ax) \tag{1.13}$$

This means, the matrix A takes the coordinate vector x given in the parameterization B and makes it readable in terms of the new parameterization B'.

Consequently, the role of the matrix A is to transform B to B'.

Since this transformation must go either way, A is invertible, so that

$$B' \;=\; BA^{-1} \tag{1.14}$$

In analogy to the above interpretation, we note that A^{-1} transforms B' to B.

For a vector $v = Bx = B'x'$ as in (1.9) we have

$$v \;=\; Bx \;=\; B'Ax \tag{1.15}$$

Consequently, the vector Ax in the latter formula gives us the coordinates of v in terms of the basis B'. Therefore we obtain by (1.15) the *transformation of coordinates* of a vector, here in terms of the coordinates x in the basis B into the coordinates x' in the basis B':

$$x' \;=\; Ax \tag{1.16}$$

The transformation of coordinates going the inverse way is given by substituting B' as by (1.14) and following the same procedure as discussed.

1.3 Inner Product and Orthogonality

We now deal with vectors in $\mathbb{R}^n$.

Definition 1.3.1 (inner product) *A function*

$$\langle \cdot , \cdot \rangle \;:\; \mathbb{R}^n \times \mathbb{R}^n \;\to\; \mathbb{R} \tag{1.17}$$

is an inner product if

(i) $\langle u, \alpha v + \beta w \rangle \;=\; \alpha \langle u, v \rangle + \beta \langle u, w \rangle \quad \textit{for all} \quad \alpha, \beta \in \mathbb{R}$

(ii) $\langle u, v \rangle \;=\; \langle v, u \rangle$

(iii) $\langle v, v \rangle \;\geq\; 0 \qquad \textit{and} \qquad \langle v, v \rangle = 0 \;\Leftrightarrow\; v = \vec{0}$

For a vector v, $\sqrt{\langle v, v \rangle}$ is called its *norm*, thus the inner product induces a *metric*.

A *standard basis* of the $\mathbb{R}^n$ is the set of vectors

$$e_1 := (1, 0, 0, \ldots, 0)^\top ,\; e_2 := (0, 1, 0, \ldots, 0)^\top ,\; \ldots, e_n := (0, 0, \ldots, 0, 1)^\top \tag{1.18}$$

The identity matrix $I := [e_1, e_2, \ldots, e_n]$ encodes these vectors as columns.

Definition 1.3.2 (canonical inner product) *For vectors* $x := (x_1, x_2, \ldots, x_n)^\top$ *and* $y := (y_1, y_2, \ldots, y_n)^\top$ *in* $\mathbb{R}^n$*, the canonical inner product is given by*

$$\langle x, y\rangle \; := \; x^\top y \; = \; x_1y_1 + x_2y_2 + \ldots + x_ny_n \tag{1.19}$$

This inner product induces the *Euclidean norm*, or *2-norm*, which measures the length of a vector as

$$\|x\|_2 \; = \; \sqrt{x^\top x} \; = \; \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2} \tag{1.20}$$

If we employ another basis B' instead of $B = I$ so that $x' = Ax$ and $y' = Ay$, then the inner product in terms of the new coordinates is

$$\langle x, y\rangle \; = \; x^\top y \; = \; \left(A^{-1}x'\right)^\top \left(A^{-1}y'\right) \; = \; (x')^\top A^{-\top} A^{-1} y' \tag{1.21}$$

with $A^{-\top} := \left(A^{-1}\right)^\top$. We denote the expression of the canonical inner product with respect to the new basis B' by

$$\langle x', y'\rangle_{A^{-\top}A^{-1}} \; := \; (x')^\top A^{-\top} A^{-1} y' \tag{1.22}$$

Remark 1.3.1 *The aim in defining a canonical inner product as well as of its representation (1.22) is, that angles and distances between points (as measured using the canonical inner product) stay the same when we transform bases and point coordinates.*

We conclude the presentation of the inner product by defining orthogonality.

Definition 1.3.3 (orthogonality) *Two vectors* x*,* y *are said to be orthogonal if their inner product is zero:*

$$x \perp y \; \Leftrightarrow \; \langle x, y\rangle = 0 \tag{1.23}$$

1.4 The Cross Product

While the inner product of two vectors is a scalar number, the cross product of two vectors is a vector as defined below.

Definition 1.4.1 (cross product) *Given two vectors* $u := (u_1, u_2, u_3)^\top$ *and* $v := (v_1, v_2, v_3)^\top$ *in* $\mathbb{R}^3$*, their cross product is given by*

$$u \times v \ := \ \begin{pmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{pmatrix} \in \mathbb{R}^3 \tag{1.24}$$

It is easy to verify that

$$\langle u \times v, u \rangle \ = \ \langle u \times v, v \rangle \ = \ 0 \tag{1.25}$$

and

$$u \times v \ = \ -v \times u \tag{1.26}$$

Therefore, the cross product of two vectors is orthogonal to each of its factors, and the order of factors defines an *orientation*.

Example 1.4.1 For $e_1 = (1, 0, 0)^\top$ and $e_2 = (0, 1, 0)^\top$, we compute

$$e_1 \times e_2 \ = \ \begin{pmatrix} 0 \cdot 0 - 0 \cdot 1 \\ 0 \cdot 0 - 1 \cdot 0 \\ 1 \cdot 1 - 0 \cdot 0 \end{pmatrix} \ = \ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \ = \ e_3 \tag{1.27}$$

This means, for a standard Cartesian frame, the cross product of the principal axes X and Y gives the principal axis Z. This is the *right-hand rule*.

If we fix u in (1.24), the cross product defines a linear mapping $v \mapsto u \times v$ which can be written in terms of the matrix $\hat{u} \in \mathbb{R}^{3\times3}$ defined as

$$\hat{u} \ := \ \begin{pmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{pmatrix} \tag{1.28}$$

so that

$$u \times v \ = \ \hat{u}v \tag{1.29}$$

Let us comment on the structure of the matrix $\hat{u}$.

Definition 1.4.2 (skew symmetric matrix) *A matrix $M \in \mathbb{R}^{3\times3}$ is skew symmetric, i.e. $M^\top = -M$, if and only if $M = \hat{u}$ for some $u \in \mathbb{R}^3$.*

By the structure of entries in $\hat{u}$, it is evident that there is a ono-to-one mapping between $\mathbb{R}^3$ and the set of skew symmetric 3×3 matrices. Moreover, the addition of skew symmetric matrices as well as the multiplication with a scalar preserves skew symmetry. This leads to to the following conclusion.

Proposition 1.4.1 *The set of skew symmetric 3×3 matrices has the structure of a vector space. This space of skew symmetric matrices is isomorphic ('of the same shape as') the $\mathbb{R}^3$.*

The corresponding isomophism is defined via the *hat operator* and its inverse, the *vee operator*:

$$\wedge(u) \;:=\; \hat{u} \quad \text{and} \quad \vee(\hat{u}) \;:=\; u \tag{1.30}$$

1.5 Excursion on Vector Products

The aim of this paragraph is to recall and illuminate some useful properties of the inner product and the cross product.

The basic tool we will rely on is the *Theorem of Pythagoras*

$$a^2 + b^2 \;=\; c^2 \tag{1.31}$$

for a right-angled triangle defined by points (p_1, p_2, p_3) where c is the length of the hypotenuse.

Let us recall the geometrical definition of *sine* and *cosine* using the unit circle, i.e.

$$p_1 \;:=\; \vec{0}, \quad p_2 \;:=\; (\cos\alpha, 0)^\top, \quad p_3 \;:=\; (\cos\alpha, \sin\alpha)^\top \tag{1.32}$$

where α is the angle between $p_1\vec{p}_2$ and $p_1\vec{p}_3$, and

$$\sin\alpha \;:=\; \frac{A}{H} \;:=\; \frac{|\text{opposite}|}{|\text{hypotenuse}|}, \quad \cos\alpha \;:=\; \frac{G}{H} \;:=\; \frac{|\text{adjacent}|}{|\text{hypotenuse}|} \tag{1.33}$$

Let us stress that the definitions in (1.33) are independent of a scaling paramater s, e.g.

$$\cos\alpha \;=\; \frac{G}{H} \;=\; \frac{G \cdot s}{H \cdot s} \tag{1.34}$$

This implies:

- One may scale the *complete triangle* (p_1, p_2, p_3) to an arbitrary multiple of the vector $\vec{p_1 p_3}$ without changing the value of $\cos\alpha$.
- When scaling the vector $\vec{p_1 p_3}$ to obtain a new point $p'_3 := s' \cdot \vec{p_1 p_3}$, the point $p'_2 := s' \cdot \vec{p_1 p_2}$ is always given by letting fall a perpendicular from p'_3 in the direction of $\vec{p_1 p_2}$.

More generally, we obtain, where $\|.\|$ without index denotes from now on the Euclidean distance:

Proposition 1.5.1 *Let $\vec{a}$ and $\vec{b}$ be two free vectors in $\mathbb{R}^2$. Then the projection of $\vec{a}$ on the direction of $\vec{b}$ is given by*

$$\mathrm{proj}_{\vec{b}}(\vec{a}) \;=\; \underbrace{\|\vec{a}\| \cdot \cos\alpha}_{\text{projected length of } \vec{a}} \cdot \underbrace{\frac{\vec{b}}{\|\vec{b}\|}}_{\text{direction}} \tag{1.35}$$

where α is the angle between $\vec{b}$ and $\vec{a}$.

Let us now generalize the Theorem of Pythagoras:

Theorem 1.5.1 (of Ghiyath Al-Kashi (1380-1429)) *Consider a triangle with vertices of lengths a, b and c, and where γ is the angle opposite of the side corresponding to c. Then it holds:*

$$c^2 \;=\; a^2 + b^2 - 2ab\cos\gamma \tag{1.36}$$

Proof 1 Without loss of generality, let us place the triangle in a Cartesian frame by

$$p_1 \;:=\; (b\cos\gamma, b\sin\gamma)^\top, \quad p_2 \;:=\; (a, 0)^\top, \quad p_3 \;:=\; \vec{0} \tag{1.37}$$

where γ is the angle between $\vec{p_3 p_2}$ and $\vec{p_3 p_1}$.

Letting fall a perpendicular from p_1 onto the direction $\vec{p_3 p_2}$ gives the point p_4 with

$$p_4 \;:=\; (b\cos\gamma, 0)^\top \tag{1.38}$$

By the coordinates in (1.37), we obviously have

$$b' \;:=\; \|\vec{p_4 p_1}\| \;=\; b\sin\gamma \quad \text{and} \quad a' \;:=\; \|\vec{p_4 p_2}\| \;=\; |a - b\cos\gamma| \tag{1.39}$$

We apply the Theorem of Pythagoras for the triangle (p_1, p_4, p_2) with $c = \|\vec{p_2 p_3}\|$ to obtain

$$\begin{aligned} c^2 &= (a')^2 + (b')^2 \\ \stackrel{(1.39)}{\Leftrightarrow} \quad c^2 &= \underbrace{(a - b\cos\gamma)^2}_{|a-b\cos\gamma|^2=(\pm1)^2(a-b\cos\gamma)^2} + (b\sin\gamma)^2 \\ \Leftrightarrow \quad c^2 &= a^2 - 2ab\cos\gamma + \underbrace{b^2\cos^2\gamma + b^2\sin^2\gamma}_{=b^2(\cos^2\gamma+\sin^2\gamma)} \\ \Leftrightarrow \quad c^2 &= a^2 + b^2 - 2ab\cos\gamma \qquad (1.40) \end{aligned}$$

since $\cos^2\gamma + \sin^2\gamma = 1$. ■

We now show how the canonical inner product is related to the angle between vectors.

Theorem 1.5.2 *Let $\vec{a}$, $\vec{b}$ be two free vectors in $\mathbb{R}^n$, $n \geq 2$, and let θ be the angle between them. Then it holds*

$$\langle \vec{a}, \vec{b} \rangle = \|\vec{a}\| \cdot \|\vec{b}\| \cdot \cos\theta \qquad (1.41)$$

Proof 2 We define the difference vector

$$\vec{c} := \vec{a} - \vec{b} \qquad (1.42)$$

creating a triangle by $\vec{a}, \vec{b}, \vec{c}$ in a twodimensional subspace of $\mathbb{R}^n$. By the Theorem of Al-Kashi we have

$$\|\vec{c}\|^2 = \|\vec{a}\|^2 + \|\vec{b}\|^2 - 2\|\vec{a}\|\|\vec{b}\|\cos\gamma \qquad (1.43)$$

We also have

$$\|\vec{c}\|^2 = \langle \vec{a} - \vec{b}, \vec{a} - \vec{b} \rangle = \langle \vec{a}, \vec{a} \rangle - \langle \vec{b}, \vec{a} \rangle - \langle \vec{a}, \vec{b} \rangle + \langle \vec{b}, \vec{b} \rangle \qquad (1.44)$$

so that

$$\|\vec{c}\|^2 = \|\vec{a}\|^2 + \|\vec{b}\|^2 - 2\langle \vec{a}, \vec{b} \rangle \qquad (1.45)$$

Equating the right-hand sides of (1.43) and (1.45) gives the desired result. ■

Let us emphasize that by (1.41) the inner product gieves a measure for the *parallelity* of two vectors, since the cosine becomes larger the more two vectors are in parallel.

Let us now stay in $\mathbb{R}^3$ and consider also the cross product.

One may easily validate the *Lagrange identity*

$$\|\vec{a} \times \vec{b}\|^2 \;=\; \|\vec{a}\|^2 \cdot \|\vec{b}\|^2 - \langle \vec{a}, \vec{b} \rangle^2 \tag{1.46}$$

which we use in the following.

Theorem 1.5.3 *For $\vec{a}, \vec{b} \in \mathbb{R}^3$, the number $\|\vec{a} \times \vec{b}\|$ is the area of the parallelogram spanned by $\vec{a}$ and $\vec{b}$.*

Proof 3 Let θ be the angle between $\vec{b}$ and $\vec{a}$, so that $\langle \vec{a}, \vec{b} \rangle = \|\vec{a}\| \cdot \|\vec{b}\| \cdot \cos\theta$. By (1.46) we then have

$$\|\vec{a} \times \vec{b}\|^2 \;=\; \|\vec{a}\|^2 \cdot \|\vec{b}\|^2 - \|\vec{a}\|^2 \cdot \|\vec{b}\|^2 \cdot \cos^2\theta \;=\; \|\vec{a}\|^2 \cdot \|\vec{b}\|^2 \cdot \big(1 - \cos^2\theta\big) \tag{1.47}$$

which implies

$$\|\vec{a} \times \vec{b}\|^2 \;=\; \|\vec{a}\|^2 \cdot \|\vec{b}\|^2 \cdot \sin^2\theta \tag{1.48}$$

Since $\|\vec{a} \cdot \sin\theta\|$ is the height of the parallelogram above the base segment $\vec{b}$ and $\|\vec{b}\|$ is the length of the latter, $\|\vec{a}\| \cdot \|\vec{b}\| \cdot |\sin\theta|$ is the sought area. ∎

Let us consider an additional vector $\vec{c}$. By computing the projection onto the vector $\vec{a} \times \vec{b}$, i.e.

$$\mathrm{proj}_{\vec{a} \times \vec{b}}(\vec{c}) \;=\; \|\vec{c}\| \cdot \cos\alpha \cdot \frac{\vec{a} \times \vec{b}}{\|\vec{a} \times \vec{b}\|} \;\overset{(1.41)}{=}\; \|\vec{c}\| \cdot \frac{\langle \vec{c}, \vec{a} \times \vec{b} \rangle}{\|\vec{c}\| \cdot \|\vec{a} \times \vec{b}\|} \cdot \frac{\vec{a} \times \vec{b}}{\|\vec{a} \times \vec{b}\|} \tag{1.49}$$

we can compute the height of the parallelepipedon over the base area spanned by $\vec{a}$ and $\vec{b}$:

$$\left|\mathrm{proj}_{\vec{a} \times \vec{b}}(\vec{c})\right| \;=\; \left| \frac{\langle \vec{c}, \vec{a} \times \vec{b} \rangle}{\|\vec{a} \times \vec{b}\|} \right| \cdot \underbrace{\left\| \frac{\vec{a} \times \vec{b}}{\|\vec{a} \times \vec{b}\|} \right\|}_{=1} \tag{1.50}$$

By multiplying height with base area $\|\vec{a} \times \vec{b}\|$ follows

Proposition 1.5.2 *The volume V of the parallelepipedon given by $(\vec{a}, \vec{b}, \vec{c})$ is given by the scalar triple product*

$$V = \left|\langle \vec{c}, \vec{a} \times \vec{b} \rangle\right| \tag{1.51}$$

The latter expression for V can algebraically be made identical to

$$V = \left|\det\left[\vec{a}, \vec{b}, \vec{c}\right]\right| \tag{1.52}$$

To conclude, let us emphasize that while the inner product gives a measure for *parallelity* of two vectors, the cross product enables to measure *orthogonality* of two vectors: As observable via (1.48), the cross product becomes larger if the absolute of the sine function i.e. the orthogonality of input vectors grows. Information on *orientation* is encoded in addition via the ordering of the input vectors.

1.6 Excursion on Complex Eigenvalues and Eigenvectors of a Matrix

The aim of this paragraph is to elaborate on the role of complex eigenvalues of a matrix and the associated eigenvectors. In preparation of a later application, we make use of an example concerned with a rotation matrix.

We consider a rotation in 2-D as modeled by a *rotation matrix*

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{1.53}$$

where the direction of rotation is, in a Cartesian frame, counterclockwise for positive θ.

Let us briefly discuss eigenvalues and eigenvectors of $R(\theta)$. Concerning the eigenvalues, an easy computation gives

$$\det(R(\theta) - \lambda I) = \begin{vmatrix} \cos\theta - \lambda & -\sin\theta \\ \sin\theta & \cos\theta - \lambda \end{vmatrix} = 0 \quad \Leftrightarrow \quad \lambda^2 - 2\lambda\cos\theta + 1 = 0 \tag{1.54}$$

so that one can verify that the two eigenvalues $\lambda_{1,2}$ are

$$\lambda_{+,-} = \cos\theta \pm \mathrm{i}\sin\theta \quad \text{where} \quad \mathrm{i}^2 = -1 \tag{1.55}$$

Can we interpret this result? The fact that we have two complex eigenvalues shows, that there is no vector except $\vec{0}$ in 2-D whose *orientation* in space is kept invariant (for general θ!) and which is *stretched* in space by an eigenvalue λ. This is a natural result since $R(\theta)$ describes a rotation where only $\vec{0}$ is kept fixed.

We now turn to the eigenvectors. Plugging the concrete $\lambda_{+,-}$ into the constituting equation yields

$$(R(\theta) - \lambda_{+,-} I)\begin{pmatrix} x \\ y \end{pmatrix} = \vec{0} \quad \Leftrightarrow \quad \begin{pmatrix} \mp \mathrm{i}\sin\theta & -\sin\theta \\ \sin\theta & \mp \mathrm{i}\sin\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \vec{0} \tag{1.56}$$

Dividing the latter equations by $\sin\theta$ results in the set of equations

$$\mp \mathrm{i}x = y \quad \text{and} \quad \pm \mathrm{i}y = x \tag{1.57}$$

Fixing $x = 1$ – which is just chosen but not required – we obtain via the two sign combinations in (1.57) the corresponding two (unnormalized) eigenvectors

$$v_+ := \begin{pmatrix} 1 \\ -\mathrm{i} \end{pmatrix} \quad \text{and} \quad v_- := \begin{pmatrix} 1 \\ \mathrm{i} \end{pmatrix} \tag{1.58}$$

Let us stress that one could have fixed alternatively $x = 1 + i$ (just as an example for *some* number) leading to two different eigenvectors as in (1.58). The particular choice $x = 1$ has the property that we can interprete the situation geometrically, since the resulting entries of v_+, v_- are *either* real *or* imaginary.

For a geometric interpretation, consider the following. Understanding the entries of v_+, v_- as coordinates $(\mathbb{R}, \mathrm{i}\mathbb{R})$, then the corresponding coordinate vectors are

$$\tilde{v}_+ := \begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \text{and} \quad \tilde{v}_- := \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{1.59}$$

for which we have

$$\langle \tilde{v}_+, \tilde{v}_- \rangle = \langle \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rangle = 1 - 1 = 0 \tag{1.60}$$

We observe that the two (unnormalized) eigenvectors are orthogonal. This is what we expect from the eigenvectors since the matrix $R(\theta)$ is orthogonal.

Now we elaborate a bit more on the eigenvectors with complex entries. Since $R(\theta)$ is orthogonal, we generally expect that there is a complete system of eigenvectors. Since we have complex eigenvalues and the matrix $R(\theta)$ is not symmetric, one can expect that the complete system of eigenvectors spans $\mathbb{C}^2$, not $\mathbb{R}^2$.

What is actually done when we multiply the eigenvectors by a complex eigenvalue is easily illustrated using the *polar form* of complex numbers. For $z = a + ib$ with $a, b \in \mathbb{R}$, one may write alternatively

$$z \;=\; r \cdot e^{\mathrm{i}\varphi} \;=\; r \cdot (\cos\varphi + \mathrm{i} \cdot \sin\varphi) \quad \text{with} \quad r, \varphi \in \mathbb{R} \tag{1.61}$$

with the relations $a = r \cdot \cos\varphi$ and $b = r \cdot \sin\varphi$. Multiplying two such complex numbers gives

$$r \cdot e^{\mathrm{i}\varphi} \,\cdot\, s \cdot e^{\mathrm{i}\psi} \;=\; r \cdot s \cdot e^{\mathrm{i}(\varphi+\psi)} \tag{1.62}$$

We observe that the multiplication of two complex numbers means to *(i)* add up the angles and *(ii)* to multiply the distances from the origin w.r.t. the polar form. Consequently, recalling that

$$\lambda_{+,-} \;=\; \cos\theta \pm i \sin\theta \;=\; e^{\pm \mathrm{i}\theta} \tag{1.63}$$

we see that by multiplying with $\lambda_{+,-}$ we only perform a *rotation* by the angle $\pm\theta$ while no stretching takes place because $|\lambda_{+,-}| = 1$.

Since in the constituting formula $R(\theta)x = \lambda x$ we multiply the eigenvalue with an eigenvector, we see that we rotate via $\lambda_{+,-}$ as above the individual components of x in polar form. Consequently, we perform a rotation in $\mathbb{C}^2$; note that this is natural since $R(\theta)$ also describes a rotation.

Let us now consider for a moment an eigenvector v

$$R(\theta)v \;=\; e^{\mathrm{i}\theta}v \tag{1.64}$$

Since $R(\theta)$ only has real entries, it holds

$$R(\theta)v = e^{\mathrm{i}\theta}v \quad \Rightarrow \quad \overline{R(\theta)v} = \overline{e^{\mathrm{i}\theta}v} \quad \Rightarrow \quad R(\theta)\bar{v} = e^{-\mathrm{i}\theta}\bar{v} \tag{1.65}$$

To fix the result, if we have two complex conjugate eigenvalues, also the eigenvectors are their complex conjugate, respectively.

Getting back to the rotation matrix, let us consider two eigenvectors of the eigenvalues $\lambda_{+,-} = e^{\pm \mathrm{i}\theta}$. Making use of our last result, these two eigenvectors can be written as $v_+ = v$ and $v_- = \bar{v}$.

Specifying the vectors

$$v_1 \;:=\; \frac{1}{2}\,(v + \bar{v}) \quad \text{and} \quad v_2 \;:=\; \frac{\mathrm{i}}{2}\,(v - \bar{v}) \tag{1.66}$$

we observe that they have only real-valued entries. We can compute

$$R(\theta)v_1 \;=\; R(\theta)\left[\frac{1}{2}\,(v+\bar{v})\right] \;=\; \frac{1}{2}\,(R(\theta)v+R(\theta)\bar{v}) \;=\; \frac{1}{2}\,(\lambda_+ v+\lambda_-\bar{v}) \tag{1.67}$$

Writing $\lambda_{+,-}$ in terms of sine and cosine, we have

$$\frac{1}{2}\,(\cos\theta v+\mathrm{i}\sin\theta v+\cos\theta\bar{v}+\mathrm{i}\sin\theta\bar{v}) \;=\; \cos\theta v_1+\sin\theta v_2 \tag{1.68}$$

An analogous computation can be done for $R(\theta)v_2$, so that we summarize

$$R(\theta)v_1=\cos\theta v_1+\sin\theta v_2 \quad \text{and} \quad R(\theta)v_2=-\cos\theta v_1+\sin\theta v_2 \tag{1.69}$$

Let us note that we may identify v_1, v_2 with their normalized versions without change in the above computations but division by $\|v_i\|$, $i=1,2$.

These developments show that by v_1, v_2 we have achieved:

- It is a real-valued basis of the rotation plane;
- vectors given in terms of this basis are rotated in the usual way in $\mathbb{R}^2$;
- generally speaking, a meaningful real-valued setting can sometimes be extracted from complex eigenvalues and eigenvectors.

Chapter 2

Rigid-Body Motion

Consider an object moving in front of a camera. As a first approach to describe the motion, one could track the points $\vec{X}$ on the object in time t (beginning at $t = 0$), writing $\vec{X}(t)$.

To make the *rigid* property concrete, if $\vec{X}(t)$ and $\vec{Y}(t)$ are two points on the object, the distance between them is constant:

$$\left\| \vec{X}(t) - \vec{Y}(t) \right\| \; = \; \text{constant } \forall t \tag{2.1}$$

A *rigid-body motion/transformation* is a family of mappings

$$g(t) \; : \; \mathbb{R}^3 \to \mathbb{R}^3; \quad \vec{X}(t) \; \mapsto \; g(t)\vec{X} \tag{2.2}$$

where $\vec{X} := \vec{X}(t = 0)$, describing how the points on the object move in time while satisfying (2.1).

If we just look for the mapping between initial and final state, this is a *rigid-body displacement*

$$g \; : \; \mathbb{R}^3 \to \mathbb{R}^3; \quad \vec{X} \; \mapsto \; g\left(\vec{X}\right) \tag{2.3}$$

Denoting $v := \vec{Y} - \vec{X}$ the difference vector between two points $\vec{X}$ and $\vec{Y}$ on the object, we obtain after transformation

$$u \; = \; g_*(v) \; := \; g\left(\vec{Y}\right) - g\left(\vec{X}\right) \tag{2.4}$$

Since g satisfies (2.1), it holds

$$\|g_*(v)\| \; = \; \|v\| \tag{2.5}$$

for all free vectors $v \in \mathbb{R}^3$.

Definition 2.0.1 *A mapping g_* satisfying (2.5) is called Euclidean transformation. In 3-D, the set of such transformations is denoted by $E(3)$.*

However, also operations describing that an object is mirrored fulfill (2.1). In order to rule out here this kind of mappings, we require Euclidean transformations to preserve orientations. This is encoded in the following definition.

Definition 2.0.2 *A mapping $g : \mathbb{R}^3 \to \mathbb{R}^3$ is a rigid-body motion also denoted as special Euclidean transformation if it preserves the norm and the cross product of any given vectors:*

$$\begin{aligned} \textit{(norm)} \qquad \|g_*(v)\| &= \|v\| && \forall v && (2.6) \\ \textit{(cross product)} \quad g_*(v) \times g_*(w) &= g_*(v \times w) && \forall v, w && (2.7) \end{aligned}$$

The set of all such transformations is denoted by $SE(3)$. Properties like (2.6) and (2.7) are called invariants of the special Euclidean transform.

Remark 2.0.1 *The intuition after (2.7) is as follows. Consider the three vectors v, w and $u := v \times w$. For given v and w, the vector u is orthogonal to them and satisfies a right-hand rule* before transformation. *Transforming then – as in the left-hand side of (2.7) – v and w via g_* and taking the cross product of the new vectors, we compare this to the transform of u, see the right-hand side of (2.7). Orientation is only preserved if equality is preserved for the three transformed vectors exactly.*

Instead of using the norm in the above definition, we may have chosen instead the condition that *angles* are preserved:

$$\langle u, v \rangle = \langle g_*(u), g_*(v) \rangle \quad \forall u, v \in \mathbb{R}^3 \tag{2.8}$$

Let us emphasize:

If we use a special Euclidean transform for representing rigid-body motion, then the preservation of form and orientation of an object is done automatically by the transform!

Therefore we can represent the *configuration* of a rigid body by specifying the motion of just one point, attaching a Cartesian coordinate frame on it for keeping track of orientation. We will look for the motion of this *coordinate frame* relative to a fixed *world/reference frame*.

For the coordinate frame, we consider three *orthonormal* vectors $\hat{e}_1, \hat{e}_2, \hat{e}_3 \in \mathbb{R}^3$ with

$$\langle \hat{e}_i, \hat{e}_j \rangle \;=\; \delta_{ij} \;:=\; \begin{cases} 1 & \text{for} \quad i = j \\ 0 & \text{for} \quad i \neq j \end{cases} \tag{2.9}$$

The vectors are ordered as to form a right-handed frame:

$$\hat{e}_1 \times \hat{e}_2 \;=\; \hat{e}_3 \tag{2.10}$$

Let us check what happens by applying a rigid-body motion g. We have

$$\langle g_* (\hat{e}_i), g_* (\hat{e}_j) \rangle \;\overset{(2.8)}{=}\; \langle \hat{e}_i, \hat{e}_j \rangle \;=\; \delta_{ij} \tag{2.11}$$

and

$$g_* (\hat{e}_1) \times g_* (\hat{e}_2) \;\overset{(2.7)}{=}\; g_* (\hat{e}_1 \times \hat{e}_2) \;=\; g_* (\hat{e}_3) \tag{2.12}$$

As expected, the resulting vectors still form a right-handed orthonormal frame.

Evidently, the configuration of a moving object can be entirely specified by two components of the motion of such an *object coordinate frame*:

1. The vector between the origin of the world frame and that of the object coordinate frame, called the *translational part* T;

2. the rotation of the object coordinate frame relative to the fixed world frame, called the *rotational part* R.

Besides the obvious need to study translations and rotations, let us also take a more general point of view on *geometric transforms*:

- The composition of two admissible transforms (one after the other) should define again an admissible transform;

- the identity transform (doing nothing) should be an admissible transform;

- transforms should be invertible;

- the composition of more than two transforms ϕ_i should behave associatively: $\phi_1 \circ (\phi_2 \circ \phi_3) = (\phi_1 \circ \phi_2) \circ \phi_3$

The structure just described is a *group structure*. Transformations as discussed up to now can be expressed by *matrices*. Therefore, we will study group properties and invariants of some useful sets of matrices.

2.1 Matrix Groups

We now study a general framework for interesting transformations. At the heart there is the *group* notion. It relies on a set and an operation defined on this set.

Definition 2.1.1 *Consider the set of invertible matrices*

$$A \in \mathbb{R}^{n\times n} \quad \textit{with} \quad \det(A) \neq 0 \tag{2.13}$$

together with the usual matrix multiplication '·', i.e.

$$G := \left(\left\{ A \in \mathbb{R}^{n\times n} \;\middle|\; \det(A) \neq 0 \right\}, \cdot\right) \tag{2.14}$$

G is a group:

1. *Closedness – For $A, B \in G$, it holds $A \cdot B \in G$*
2. *Associativity – $A \cdot (B \cdot C) = (A \cdot B) \cdot C$*
3. *Neutral element –*
   There exists $I \in G$, $I = [e_1, e_2, \ldots, e_n]$, with $IA = AI = A$ for all $A \in G$
4. *Inverse element –*
   For each $A \in G$, there exists $B = A^{-1} \in G$ with $AB = BA = I$

G is called general linear group $GL(n, \mathbb{R})$.

G represents all linear maps from $\mathbb{R}^n$ to $\mathbb{R}^n$ by which no geometric information of a configuration is lost. It is *not* a *commutative* group: In general $A \cdot B \neq B \cdot A$, e.g.

$$\begin{pmatrix} 1 & 3 \\ 2 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix} = \begin{pmatrix} 4 & 10 \\ 2 & 2 \end{pmatrix} \quad \text{but} \quad \begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix} \cdot \begin{pmatrix} 1 & 3 \\ 2 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 7 & 3 \end{pmatrix} \tag{2.15}$$

We will now discuss some specific groups, so-called *subgroups* of $GL(n, \mathbb{R})$. These are characterised as follows:

- Of interest is a subset S of $\{ A \in \mathbb{R}^{n\times n} \mid \det(A) \neq 0\}$ with specific properties;
- on this subset S, the operation '·' inherits the structure of $GL(n, \mathbb{R})$;
- especially, $(S, \cdot)$ is closed.

An obvious invariant of interest is given by mappings A that preserve *volume* and *orientation*. This gives rise to:

Definition 2.1.2 *The special linear group is given by*

$$SL(n, \mathbb{R}) \ := \ \left\{ A \in \mathbb{R}^{n \times n} \ \middle| \ \det(A) = 1 \right\} \tag{2.16}$$

A building block of Euclidean transformations is given by the *orthogonal group*:

Definition 2.1.3 *The orthogonal group consists of matrices with orthonormal columns or rows:*

$$O(n) \ := \ \left\{ A \in GL(n, \mathbb{R}) \mid \ A^\top A = I \right\} \tag{2.17}$$

This definition encodes

$$\langle Ax, Ay \rangle \ = \ (Ax)^\top Ay \ = \ x^\top A^\top Ay \ \overset{(2.17)}{=} \ x^\top I y \ = \ \langle x, y \rangle \tag{2.18}$$

for any $x, y \in \mathbb{R}^n$. Therefore, $O(n)$ represents linear maps that preserve inner products – and thus Euclidean norms, lengths and angles – of vectors in $\mathbb{R}^n$.

It also follows:

Theorem 2.1.1 *For* $A \in O(n)$ *it holds* $\det(A) = \pm 1$.

Proof 4

$$\begin{aligned} \det(A) &= \det\left(A^\top\right) && \textit{(elementary)} \\ &= \det\left(A^{-1}\right) && \textit{(since } A^\top A = I, \textit{ i.e. } A^\top = A^{-1}\textit{)} \\ &= \left(\det(A)\right)^{-1} && \textit{(elementary)} \end{aligned}$$

Multiplication with $\det(A)$ gives $\det^2(A) = 1$, i.e. $\det(A) = \pm 1$. ■

The possibility $\det(A) = -1$ implies that mirror operations are allowed, e.g. the matrix

$$A \ = \ \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \text{with } \det(A) = -1 \tag{2.19}$$

realizes in $\mathbb{R}^2$ a mirror operation with the x-axis as reflection axis.

In order to rule out such operations, we define:

Definition 2.1.4 *The special orthogonal group $SO(n)$ consists of all orthogonal matrices A with* $\det(A) = 1$*, i.e.*

$$SO(n) \;:=\; O(n) \cap SL(n, \mathbb{R}) \tag{2.20}$$

The special orthogonal group exactly describes *rotations* around the origin.

For realizing *translations*, we have to add a translation vector. This is done via a special construction:

Definition 2.1.5 *The affine group $Aff(n)$ consists of all matrices in $\mathbb{R}^{(n+1)\times(n+1)}$ of the form*

$$A \;=\; \begin{pmatrix} B & \vec{b} \\ \vec{0}^\top & 1 \end{pmatrix} \tag{2.21}$$

with $B \in GL(n, \mathbb{R})$ and $\vec{b}, \vec{0} \in \mathbb{R}^n$.

Let us briefly elaborate on this special construction. For $\tilde{v} = A\tilde{u}$ where $\tilde{v} := (v, 1)^\top$ and $\tilde{u} := (u, 1)^\top$ with $v, u \in \mathbb{R}^n$, it holds

$$\tilde{v} \;=\; \begin{pmatrix} v \\ 1 \end{pmatrix} \;=\; A\tilde{u} \;=\; \begin{pmatrix} B & \vec{b} \\ \vec{0}^\top & 1 \end{pmatrix} \begin{pmatrix} u \\ 1 \end{pmatrix} \;=\; \begin{pmatrix} Bu + \vec{b} \\ 1 \end{pmatrix} \tag{2.22}$$

which implies

$$v \;=\; Bu + \vec{b} \tag{2.23}$$

While the transform is affine, we have nevertheless a linear mapping plus a group structure. Thus, we may e.g. apply two such transforms one after the other just by matrix multiplication. The construction in (2.21) enables to include a translation $\vec{b}$ whereas structural group properties are encoded via B.

Let us now include previous concepts into $Aff(n)$.

Definition 2.1.6 *The similarity group $Sim(n)$ consists of all matrices in $\mathbb{R}^{(n+1)\times(n+1)}$ of the form*

$$A \;=\; \begin{pmatrix} \alpha B & \vec{b} \\ \vec{0}^\top & 1 \end{pmatrix} \tag{2.24}$$

with $\alpha \in \mathbb{R}_+$, $B \in O(n)$ and $\vec{b}, \vec{0} \in \mathbb{R}^n$.

The name of this group arises as it describes exactly *similarity mappings* in classical geometry: Translations, rotations, reflections as well as scalings.

Euclidean and special Euclidean transforms can also be put in the matrix group framework:

Definition 2.1.7 *The Euclidean group $E(n)$ where $B \in O(n)$ is given by*

$$E(n) \; := \; \{A \in Sim(n) \mid \det(A) = \pm 1\} \tag{2.25}$$

Definition 2.1.8 *The special Euclidean group $SE(n)$ where $B \in SO(n)$ is given by*

$$SE(n) \; := \; \{A \in Sim(n) \mid \det(A) = 1\} \tag{2.26}$$

2.2 Homogeneous Coordinates

Let us consider rigid-body motion $g(R,T)$ as an affine mapping defined by a rotational part R and a translational part T.

By the special construction of $Aff(n)$, we may write $g(R,T)$ in its *homogeneous representation*

$$g_h \; = \; \begin{pmatrix} R & T \\ \vec{0}^\top & 1 \end{pmatrix} \tag{2.27}$$

Useful coordinates in the context of this representation are *homogeneous coordinates.*

Appending a '1' to the coordinates $\vec{X} = (X_1, X_2, X_3)^\top \in \mathbb{R}^3$ of a point p gives its *homogeneous coordinates*

$$\vec{X}_h \; := \; \begin{pmatrix} \vec{X} \\ 1 \end{pmatrix} \; = \; \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{pmatrix} \in \mathbb{R}^4 \tag{2.28}$$

Effectively, concerning the coordinates we have embedded the $\mathbb{R}^3$ into a hyperplane in $\mathbb{R}^4$.

Homogeneous coordinates of a vector $v = \vec{X}(q) - \vec{X}(p)$, where q and p are points in $\mathbb{R}^3$, are defined as the difference between homogeneous coordinates of the two points:

$$v_h := \begin{pmatrix} v \\ 1 \end{pmatrix} = \begin{pmatrix} \vec{X}(q) \\ 1 \end{pmatrix} - \begin{pmatrix} \vec{X}(p) \\ 1 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ 0 \end{pmatrix} \in \mathbb{R}^4 \tag{2.29}$$

Since in the latter format the zero vector $\vec{0} \in \mathbb{R}^4$ is included, this gives rise to a subspace structure as also all linear operations in $\mathbb{R}^3$ are preserved.

In summary, rigid-body motion is described via

$$SE(3) = \left\{ g_h = \begin{pmatrix} R & T \\ \vec{0}^\top & 1 \end{pmatrix} \middle| R \in SO(3), T \in \mathbb{R}^3 \right\} \subset \mathbb{R}^{4\times 4} \tag{2.30}$$

Applying a rigid-body motion at a point $p \in \mathbb{R}^3$ with homogeneous coordinates $\vec{X}_h(p)$ we obtain

$$g_h \vec{X}_h(p) = \begin{pmatrix} R & T \\ \vec{0}^\top & 1 \end{pmatrix} \vec{X}_h(p) = \begin{pmatrix} R\vec{X}(p) + T \\ 1 \end{pmatrix} \tag{2.31}$$

The action of a rigid-body motion on a vector $v = \vec{X}(q) - \vec{X}(p) \in \mathbb{R}^3$ becomes

$$g_{h,*}(v_h) = g_h \vec{X}_h(q) - g_h \vec{X}_h(p) = g_h v_h = \begin{pmatrix} Rv \\ 0 \end{pmatrix} \tag{2.32}$$

Thus, in 3-D we have $g_*(v) = Rv$ since only rotational parts affect vectors.

In summary, rigid-body motion acts differently on points and vectors.

2.3 The Theorem of Euler

Before we discuss the theorem itself, let us investigate its mathematical building blocks.

We consider again the rotation in 2-D represented by a *rotation matrix*

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{2.33}$$

where the direction of rotation is, in a Cartesian frame, counterclockwise for positive θ.

The rotation matrix $R(\theta)$ from (2.33) is easily generalized to 3-D via

$$R_x(\theta) := \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}, \qquad R_y(\theta) := \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}$$
$$\text{and} \quad R_z(\theta) := \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.34}$$

where the lower index of R indicates the axis which is kept fixed. General rotations can then be obtained by combining these basic rotations:

$$R := R(\alpha, \beta, \gamma) := R_z(\gamma) R_x(\beta) R_y(\alpha) \tag{2.35}$$

where the $y - x - z$ convention for the corresponding *Euler angles* are used.

Inspecting $R(\theta)$ from (2.33) and recalling what we learned on matrix groups, it is apparent that $R(\theta)$ and its generalizations are orthogonal matrices with orthonormal column vectors.

Let us now tackle the main part of this paragraph.

Theorem 2.3.1 (of Leonhard Euler (1707-1783)) *Any displacement of a rigid body such that a point $\vec{0}$ on the rigid body remains fixed is equivalent to a rotation about a fixed axis through the point $\vec{0}$.*

Proof 5 We consider the displacement of any point $p \in \mathbb{R}^3$ to some point $P \in \mathbb{R}^3$ on the rigid body. Employing a rotation $R \in SO(3)$ this reads as

$$P = Rp \tag{2.36}$$

Let us compute the eigenvalues and eigenvectors of R. The constituting equation

$$Rp = \lambda p \tag{2.37}$$

leads to the characteristic equation

$$\det(R - \lambda I) = 0 \tag{2.38}$$

Denoting $R = (r_{ij})$, the latter is in detail – after straight forward computation using e.g. the rule of Sarrus (due to the french mathematician Pierre Frédéric Sarrus (1798-1861)) – identical to

$$\begin{aligned} -\lambda^3 &+ \lambda^2 (r_{11} + r_{22} + r_{33}) \\ &-\lambda \left[(r_{22}r_{33} - r_{32}r_{23}) + (r_{11}r_{33} - r_{31}r_{13}) + (r_{11}r_{22} - r_{21}r_{12}) \right] = 0 \end{aligned} \tag{2.39}$$

Because $R \in SO(3)$ it holds $\det(R) = 1$. As $R^{-1} = R^\top$, it follows directly by Cramer's rule $A^{-1} = \mathrm{adj}(A)/\det(A)$ – due to the swiss mathematician Gabriel Cramer (1704-1752) – that

$$\left\{ \begin{array}{rcl} r_{11} & = & r_{22}r_{33} - r_{32}r_{23} \\ r_{22} & = & r_{11}r_{33} - r_{31}r_{13} \\ r_{33} & = & r_{11}r_{22} - r_{21}r_{12} \end{array} \right. \tag{2.40}$$

Making use of these relations in (2.39) leads to

$$-\lambda^3 + \lambda^2 (r_{11} + r_{22} + r_{33}) - \lambda (r_{11} + r_{22} + r_{33}) + 1 \ = \ 0 \tag{2.41}$$

We observe that $\lambda = 1$ is an eigenvalue of R.

In other words, there exists a vector v, namely the eigenvector corresponding to the eigenvalue $\lambda = 1$, such that all points on the line αv, $\alpha \in \mathbb{R}$, stay *invariant* under the rotation R, thus defining the sought rotation axis. ■

Let us now consider the rotation of a rigid body about v through an angle φ.

For points p and P with $P = Rp$ as in (2.36), there is a point $Q \in \mathrm{span}(v)$ so that the triangle defined by the vertices (p, P, Q) lives in a plane perpendicular to v.

We may write

$$P \ = \ p + a_1 + a_2 \tag{2.42}$$

where a_1 and a_2 are orthogonal vectors in the latter plane: The vector a_1 is assumed to point from p to Q, and it gives the base point for the vector a_2 such that (2.42) holds.

Our aim is now to specify a_1 and a_2 in terms of the rotation axis v and the angle of rotation φ.

The equation (2.42) gives rise to

$$Rp \ = \ p + a_1 + a_2 \tag{2.43}$$

Since Q is on the rotation axis, it is the center of the rotation. Thus, it holds for θ being the angle between $0\vec{Q}$ and $0\vec{p}$

$$\|\vec{Qp}\| \ = \ \|\vec{QP}\| \ = \ \|p\| \sin\theta \ = \ |v \times p| \tag{2.44}$$

where we used $|v \times p| = \|v\|\|p\| \sin\theta$ and $\|v\| = 1$. Analogously follows

$$\|a_2\| \ = \ \|\vec{QP}\| \sin\varphi \ = \ |v \times p| \sin\varphi \tag{2.45}$$

where φ is the angle between $\vec{Qp}$ and $\vec{QP}$ in the triangle (p, P, Q). Since a_2 is perpendicular to v and $\vec{Qp}$, it also holds

$$a_2 \;=\; \sin\varphi\,(v \times p) \tag{2.46}$$

Having described a_2, we now turn to a_1.

The vector $\vec{Qp}$ can be expressed via the projection

$$\vec{Qp} \;=\; p - \langle p, v\rangle v \tag{2.47}$$

Therefore

$$-a_1 \;=\; \vec{Qp} - \cos\varphi\vec{Qp} \;=\; (1-\cos\varphi)\,(p - \langle p, v\rangle v) \tag{2.48}$$

Subsituting a_1 and a_2 in (2.43) gives

$$Rp \;=\; p+(\cos\varphi-1)\,(p-\langle p, v\rangle v)+\sin\varphi\,(v\times p) \;=\; \cos\varphi\cdot p-\cos\varphi\langle p, v\rangle v+\langle p, v\rangle v+\sin\varphi\hat{v} \tag{2.49}$$

with $\hat{v}$ defined in accordance to (1.28). Making use of the relation

$$\left(vv^\top\right)p \;=\; v\left(v^\top\right)p \;=\; v\langle v, p\rangle \;=\; \langle v, p\rangle v \tag{2.50}$$

we obtain the *formula of Rodrigues* (after the french mathematician Benjamin Olinde Rodrigues (1795-1851))

$$R \;=\; I\cos\varphi + \hat{v}\sin\varphi + vv^\top\,(1-\cos\varphi) \tag{2.51}$$

By Rodrigues' formula one can easily compute the angle φ and the axis of rotation v from R. The corresponding equations derived from (2.51) are

$$\cos\varphi \;=\; \frac{1}{2}\,(\mathrm{trace}(R)-1) \tag{2.52}$$

where $\mathrm{trace}(R) = r_{11} + r_{22} + r_{33}$, and

$$\hat{v} \;=\; \frac{1}{2\sin\varphi}\left(R - R^\top\right) \tag{2.53}$$

where one can extract v from $\hat{v}$.

Let us now derive a *canonical representation* of any rotation matrix which allows us to view it as a rotation through an angle ϕ about the $z-axis$.

To this end, let us recall that we found for $r := r_{11}+r_{22}+r_{33}$ by the formula (2.41)

$$-\lambda^3 + \lambda^2 r - \lambda r + 1 \;=\; 0 \tag{2.54}$$

the eigenvalue $\lambda_3 = 1$ and the normalized eigenvector $p_3 = v$ for the rotation axis.

Factoring (2.54) and multiplying with -1 one obtains

$$(\lambda - 1)\left(\lambda^2 + \lambda(1 - r) + 1\right) \tag{2.55}$$

Defining then

$$\cos\phi \;:=\; \frac{1}{2}(r - 1) \tag{2.56}$$

one may easily check that the remaining eigenvalues are

$$\lambda_1 \;=\; e^{\mathrm{i}\phi} \quad \text{and} \quad \lambda_1 \;=\; e^{-\mathrm{i}\phi} \tag{2.57}$$

Following the extensive discussion in Section 1.6, we can construct a real-valued basis of the corresponding rotation plane, composed of normalized vectors v_1, v_2. Considering then the triad of orthonormal vectors (v_1, v_2, v), it follows by the interpretation of v as the 'z-axis' that R can be written as

$$R \;=\; Q\Lambda Q^\top \quad \text{where} \quad Q = [v_1, v_2, v] \quad \text{and} \quad \Lambda = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.58}$$

If we view the rotation from a new frame whose *orientation* is given by the coordinate transfrom Q, it is clear that in this new frame, the *displacement* is a rotation about the z-axis.

In formulae, we have after the coordinate transform

$$Qx' \;=\; x\,, \quad \text{i.e. } x' \;=\; Q^\top x \tag{2.59}$$

a rotation in this new frame about the z-axis performed by Λ, and a final back-transform mapping to the old coordinate system:

$$Rx \;=\; Q\Lambda Q^\top x \;=\; \underbrace{Q \qquad \underbrace{\Lambda x'}_{\text{rotation about 'z-axis'}}}_{\text{gives back rotated points in old coordinate system}} \tag{2.60}$$

2.4 The Theorem of Chasles

As an extension of the Theorem of Euler, we now discuss the following assertion:

Theorem 2.4.1 (of Michel Chasles (1793-1880)) *A general rigid-body displacement can be produced by a translation along a line followed (or preceded) by a rotation about that line.*

Proof 6 We consider a general homogeneous transfer matrix

$$g_h = \begin{pmatrix} R & T \\ \vec{0}^\top & 1 \end{pmatrix} \tag{2.61}$$

We consider the following *similarity transform* of g_h:

$$\Lambda = \begin{pmatrix} Q^\top & -Q^\top c \\ \vec{0}^\top & 1 \end{pmatrix} \begin{pmatrix} R & T \\ \vec{0}^\top & 1 \end{pmatrix} \begin{pmatrix} Q & c \\ \vec{0}^\top & 1 \end{pmatrix} \tag{2.62}$$

which computes to

$$\Lambda = \begin{pmatrix} Q^\top RQ & Q^\top Rc - Q^\top c + Q^\top T \\ \vec{0}^\top & 1 \end{pmatrix} \tag{2.63}$$

Let us comment that a similarity transform gives an equivalent mapping to the original mapping, just using a different basis. In accordance, the role of the matrix Q is that of a basis transform.

We choose Q as for the canonical representation of the rotation matrix, i.e.

$$Q = [v_1, v_2, v] \tag{2.64}$$

where v_1, v_2, v are the eigenvectors of R. As by (2.58), this leads to

$$Q^\top RQ = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{2.65}$$

so that the upper left 3×3 submatrix of Λ reduces to the *rotation matrix* corresponding to a rotation about the z-axis.

Defining

$$c' := Q^\top c \quad \text{and} \quad T' := Q^\top T \tag{2.66}$$

the *translation part* of Λ can be formulated as

$$\begin{aligned} & Q^\top Rc - Q^\top c + Q^\top T \\ = \; & Q^\top R \underbrace{QQ^\top}_{=I} c - IQ^\top c + Q^\top T \\ = \; & \left(Q^\top RQ - I\right) Q^\top c + Q^\top T \\ = \; & \begin{pmatrix} \cos\phi - 1 & -\sin\phi & 0 \\ \sin\phi & \cos\phi - 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} c'_1 \\ c'_2 \\ c'_3 \end{pmatrix} + \begin{pmatrix} T'_1 \\ T'_2 \\ T'_3 \end{pmatrix} \end{aligned} \tag{2.67}$$

Let us have a closer look. Provided the upper left 2×2 submatrix of $Q^\top RQ - I$ is invertible, we can find c_1', c_2' such that

$$\begin{pmatrix} \cos\phi - 1 & -\sin\phi \\ \sin\phi & \cos\phi - 1 \end{pmatrix} \begin{pmatrix} c_1' \\ c_2' \end{pmatrix} = - \begin{pmatrix} T_1' \\ T_2' \end{pmatrix} \tag{2.68}$$

In other words, we can always solve the first two equations of

$$\left(Q^\top RQ - I\right) c' = -T' \tag{2.69}$$

for the first two components of c', and by (2.67) the third component of c' can be set simply to zero.

Putting these pieces together, Λ has the form

$$\Lambda = \begin{pmatrix} \cos\phi & -\sin\phi & 0 & 0 \\ \sin\phi & \cos\phi & 0 & 0 \\ 0 & 0 & 1 & T_3' \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2.70}$$

so that the displacement can be described by a rotation about the z-axis through an angle ϕ and a simultaneous translation along the z-axis through a distance T_3'.
■

Let us comment on two points within the proof:

- As the displacement reminds on the motion of a screw, it is sometimes called *screw displacement* and the axis is called *screw axis*.
- If the upper left submatrix of $Q^\top RQ - I$ is not invertible, then $Q^\top RQ = I$ must hold. This means that Λ describes a pure translation.

The question arises how to find a normalized vector v along the screw axis and the angle of rotation ϕ about the screw axis.

To this end one can employ the formulae of Rodrigues, see (2.52) and (2.53). However, in order to apply them in the current context of the screw displacement, one needs to find the position vector of one point on the screw axis and the translation of a point on the rigid body along the screw axis.

If T_p denotes the projection of the vector T onto a plane perpendicular to v, i.e.

$$T_p := T - \langle T, v\rangle v \tag{2.71}$$

then we may choose

$$Q \;:=\; [a, b, v] \tag{2.72}$$

where

$$a \;:=\; \frac{T_p}{\|T_p\|} \quad \text{and} \quad b := a \times v \tag{2.73}$$

Let us for convenience recall the equation (2.68)

$$\underbrace{\begin{pmatrix} \cos\phi - 1 & -\sin\phi \\ \sin\phi & \cos\phi - 1 \end{pmatrix}}_{=:M} \begin{pmatrix} c_1' \\ c_2' \end{pmatrix} \;=\; -\begin{pmatrix} T_1' \\ T_2' \end{pmatrix} \quad \text{with } \; c' = Q^\top c \text{ and } T' = Q^\top T \tag{2.74}$$

With Q as in (2.72) we have

$$Q^\top T \;=\; \begin{pmatrix} a^\top T \\ b^\top T \\ v^\top T \end{pmatrix} \tag{2.75}$$

and consequently

$$\begin{pmatrix} T_1' \\ T_2' \end{pmatrix} \;=\; \begin{pmatrix} a^\top T \\ b^\top T \end{pmatrix} \tag{2.76}$$

Making use of the standard formula

$$\begin{pmatrix} e & f \\ g & h \end{pmatrix}^{-1} \;=\; \frac{1}{eh - fg} \begin{pmatrix} e & -f \\ -g & h \end{pmatrix} \tag{2.77}$$

we immediately compute M^{-1} as by (2.74):

$$M^{-1} \;=\; \frac{1}{(\cos\phi - 1)^2 + \sin^2\phi} \begin{pmatrix} \cos\phi - 1 & \sin\phi \\ -\sin\phi & \cos\phi - 1 \end{pmatrix} \tag{2.78}$$

Employing (2.76) and (2.78) we obtain from (2.74)

$$\begin{pmatrix} c_1' \\ c_2' \end{pmatrix} \;=\; \frac{-1}{(\cos\phi - 1)^2 + \sin^2\phi} \begin{pmatrix} \cos\phi - 1 & \sin\phi \\ -\sin\phi & \cos\phi - 1 \end{pmatrix} \begin{pmatrix} a^\top T \\ b^\top T \end{pmatrix} \tag{2.79}$$

Let us now simplify the terms occuring in (2.79).

By construction of b in (2.73), it is built as the cross product of two vectors a and v with $T \in \text{span}(a, v)$, so that

$$b^\top T \;=\; \langle b, T \rangle \;=\; 0 \tag{2.80}$$

Furthermore

$$(\cos\phi - 1)^2 + \sin^2\phi \;=\; 1 - 2\cos\phi + \underbrace{\cos^2\phi + \sin^2\phi}_{=1} \;=\; 2\,(1 - \cos\phi) \tag{2.81}$$

and

$$\begin{aligned}
a^\top T &= \langle T - \langle T, v\rangle v, T\rangle \\
&= \langle T - \langle T, v\rangle v, T - \langle T, v\rangle v + \langle T, v\rangle v\rangle \\
&= \|T_p\| + \langle \underbrace{T - \langle T, v\rangle v}_{\perp v}, \underbrace{\langle T, v\rangle v}_{\| v}\rangle \\
&= \|T_p\|
\end{aligned} \tag{2.82}$$

By (2.80) and (2.82) follows

$$\begin{pmatrix} a^\top T \\ b^\top T \end{pmatrix} \;=\; \begin{pmatrix} 1 \\ 0 \end{pmatrix} a^\top T \;=\; \begin{pmatrix} 1 \\ 0 \end{pmatrix} \|T_p\| \tag{2.83}$$

Putting things together we obtain

$$\begin{aligned}
\begin{pmatrix} c'_1 \\ c'_2 \end{pmatrix} &= \frac{-\|T_p\|}{2\,(1 - \cos\phi)} \begin{pmatrix} \cos\phi - 1 & \sin\phi \\ -\sin\phi & \cos\phi - 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\
&= \frac{\|T_p\|}{2\,(1 - \cos\phi)} \begin{pmatrix} 1 - \cos\phi \\ \sin\phi \end{pmatrix}
\end{aligned} \tag{2.84}$$

Letting as before the third component of c' be zero we conclude

$$c' \;=\; \frac{\|T_p\|}{2} \begin{pmatrix} 1 \\ \frac{\sin\phi}{1-\cos\phi} \\ 0 \end{pmatrix} \tag{2.85}$$

so that

$$c \;=\; Qc' \;=\; [a, b, c] \cdot \frac{\|T_p\|}{2} \begin{pmatrix} 1 \\ \frac{\sin\phi}{1-\cos\phi} \\ 0 \end{pmatrix} \;=\; \frac{\|T_p\|}{2} \left(a + \frac{\sin\phi}{1 - \cos\phi} b \right) \tag{2.86}$$

Recalling the interpretation of Λ, we see that by c we obtained the position vector of a point on the screw axis. Furthermore, the displacement of a point on this axis is given by the third component of T', i.e.

$$T'_3 \;=\; \langle v, T\rangle \tag{2.87}$$

We can also compute the so-called *pitch* of the screw, i.e. the translation along the axis as

$$h \;=\; \frac{T_3}{\phi} \tag{2.88}$$

2.5 Exponential Coordinates

The aim of this paragraph is to study a convenient way to *parameterize* rigid-body motions.

An important tool will be the *matrix exponential*

$$e^M := \sum_{k=0}^{\infty} \frac{M^k}{k!} \quad \text{for } M \in \mathbb{R}^{n\times n} \tag{2.89}$$

We will also make use of

$$\begin{aligned} \frac{d}{dt}e^{tM} &= \frac{d}{dt}\sum_{k=0}^{\infty}\frac{(tM)^k}{k!} \\ &= \frac{d}{dt}\left(\frac{1}{1}t^0M^0 + \frac{1}{1}t^1M^1 + \frac{1}{2}t^2M^2 + \frac{1}{6}t^3M^3 + \dots\right) \\ &= M\left(I + 2\frac{1}{2}tM + 3\frac{1}{6}t^2M^2 + \dots\right) \end{aligned} \tag{2.90}$$

so that

$$\frac{d}{dt}\underbrace{e^{tM}}_{=:A(t)} = \dot{A}(t) = Me^{tM} = MA(t) \tag{2.91}$$

Let us consider first *rotations*.

Given a *trajectory* $R(t) : \mathbb{R} \to SO(3)$ that describes continuous rotational motion, we must have by definition of $SO(3)$

$$R(t)R^\top(t) = I \quad \forall t \tag{2.92}$$

Computing the derivative w.r.t. the time t, we obtain

$$\dot{R}(t)R^\top(t) + R(t)\dot{R}^\top(t) = 0 \quad \forall t \tag{2.93}$$

so that

$$\dot{R}(t)R^\top(t) = -\left(\dot{R}R^\top(t)\right)^\top \quad \forall t \tag{2.94}$$

This reflects the fact that the matrix $\dot{R}(t)R^\top(t)$ is *skew symmetric*. Thus, there must exist a vector $w(t) \in \mathbb{R}^3$ such that

$$\dot{R}(t)R^\top(t) = \hat{w}(t) \tag{2.95}$$

Multiplication with $R(t)$ from the right gives

$$\dot{R}(t) = \hat{w}(t)R(t) \tag{2.96}$$

Let us observe, that if $R(t_0) = I$ for $t = t_0$, we have $\dot{R}(t_0) = \hat{w}(t_0)$. Hence, around the identity matrix I, the skew symmetric matrix $\hat{w}(t_0)$ gives a first-order approximation to a rotation matrix:

$$\dot{R}(t = t_0) \;\approx\; \frac{R(t_0 + dt) - R(t_0)}{dt} \qquad \Rightarrow \qquad R(t_0 + dt) \;\approx\; I + dt\hat{w}(t_0) \tag{2.97}$$

We formalize this observation as follows.

Definition 2.5.1 *Let the space of all skew symmetric matrices be denoted by*

$$so(3) \;:=\; \left\{\hat{w} \in \mathbb{R}^{3\times 3} \mid w \in \mathbb{R}^3\right\} \tag{2.98}$$

It is called the tangent space at the identity of the rotation group $SO(3)$.

As a remark, this definition is easily generalized: For $R(t)$ not the identity matrix, the tangent space is again $so(3)$, transported to the location $R(t)$, so that $R(t + dt) \approx R(t) + dt\hat{w}(t)$.

Comparing (2.96) with (2.91) and assuming $R(t_0 = 0) = I$ we must have

$$R(t) \;=\; e^{\hat{w}t} \tag{2.99}$$

In general, t can be absorbed into w, so that we have

$$R \;=\; e^{\hat{w}} \tag{2.100}$$

In conclusion, the matrix exponential defines a mapping from $so(3)$ to $SO(3)$, the so-called *exponential map*

$$\exp \;:\; so(3) \rightarrow SO(3)\,, \quad \hat{w} \mapsto e^{\hat{w}} \tag{2.101}$$

While one can show that any given rotation matrix R can thus be expressed by *exponential coordinates* $w \in \mathbb{R}^3$, these coordinates are unfortunately not uniquely defined.

Let us consider now the continuous motion of a rigid body described by a trajectory in $SE(3)$ in homogeneous representation:

$$g(t) \;:=\; g_h(t) \;=\; \begin{pmatrix} R(t) & T(t) \\ \vec{0}^\top & 1 \end{pmatrix} \in \mathbb{R}^{4\times 4} \tag{2.102}$$

In analogy to the case of pure rotation, let us first compute $\dot{g}(t)g(t)$ and have a look at its structure. We compute (valid for any t)

$$g^{-1} = \begin{pmatrix} R & T \\ \vec{0}^\top & 1 \end{pmatrix}^{-1} = \begin{pmatrix} R^\top & -R^\top T \\ \vec{0}^\top & 1 \end{pmatrix} \tag{2.103}$$

so that for any t

$$\dot{g}g^{-1} = \begin{pmatrix} \dot{R}R^\top & \dot{T} - \dot{R}R^\top T \\ \vec{0}^\top & 0 \end{pmatrix} \tag{2.104}$$

We already know that there exists $w \in \mathbb{R}^3$ such that $\hat{w}(t) = \dot{R}R^\top$. Let us define a vector $v := v(t)$ with

$$v := \dot{T} - \hat{w}T \tag{2.105}$$

Then (2.104) can be written as

$$\dot{g}g^{-1} = \begin{pmatrix} \hat{w} & v \\ \vec{0}^\top & 0 \end{pmatrix} =: \hat{\xi} \in \mathbb{R}^{4\times 4} \tag{2.106}$$

with $\hat{\xi} = \hat{\xi}(t)$, and we have

$$\dot{g} = \hat{\xi} \cdot g \tag{2.107}$$

where $\hat{\xi}$ can be understood as the 'tangent vector' along the curve of $g(t)$. It can be used to approximate $g(t)$ locally:

$$g(t + dt) \approx g(t) + \hat{\xi}(t)g(t)dt = \left(I + \hat{\xi}(t)dt\right) g(t) \tag{2.108}$$

A matrix of the form $\hat{\xi}$ is called a *twist*.

Definition 2.5.2 *The set of all twists is denoted by*

$$se(3) := \left\{ \hat{\xi} = \begin{pmatrix} \hat{w} & v \\ \vec{0}^\top & 0 \end{pmatrix} \mid \hat{w} \in so(3),\ v \in \mathbb{R}^3 \right\} \tag{2.109}$$

The set $se(3)$ is called the tangent space *of the matrix group $SE(3)$.*

We may also define two operators '$\vee$' and '$\wedge$' to convert between a twist $\hat{\xi} \in se(3)$ and its *twist coordinates* $\xi \in \mathbb{R}^6$ as follows:

$$\begin{pmatrix} \hat{w} & v \\ \vec{0}^\top & 0 \end{pmatrix}^\vee := \begin{pmatrix} v \\ w \end{pmatrix} \in \mathbb{R}^6, \quad \begin{pmatrix} v \\ w \end{pmatrix}^\wedge := \begin{pmatrix} \hat{w} & v \\ \vec{0}^\top & 0 \end{pmatrix} \in \mathbb{R}^{4\times 4} \tag{2.110}$$

In the twist coordinates, we may referto v as the *linear velocity* and to w as the *angular velocity.*

The time t is typically used to index camera motion. The transform $g(t) \in SE(3)$ denotes the relative displacement between some fixed world frame W and the camera frame at time t.

By default, we assume $g(0) = I$, i.e. at time $t = 0$ the camera frame coincides with W.

So, if the coordinates of a point p relative to W are $X_0 := X(0)$, its coordinates relative to the camera at time t are

$$X(t) = R(t)X_0 + T(t) \tag{2.111}$$

or in homogeneous representation

$$X(t) = g(t)X_0 \tag{2.112}$$

When the starting time is not $t = 0$, the relative motion between the camera at times t_1 and t_2 will be denoted by $g(t_1, t_2) \in SE(3)$. We have then the following relation between coordinates of the same point p at different times:

$$X(t_2) = g(t_2, t_1)X(t_1) \tag{2.113}$$

Invoking the group properties of $SE(3)$, we have the *composition rule*

$$g(t_3, t_1) = g(t_3, t_2)g(t_2, t_1) \tag{2.114}$$

when considering another camera frame, and also the *rule of inverse*

$$g^{-1}(t_2, t_1) = g(t_1, t_2) \tag{2.115}$$

Let us not only consider point coordinates under camera motion, but also point velocities.

We know that the coordinates $X(t)$ ofa point p relative to a moving camera frame are a function of time

$$X(t) = g(t)X_0 \tag{2.116}$$

Then the velocity of p relative to the camera frame is

$$\dot{X}(t) = \dot{g}(t)X_0 \tag{2.117}$$

Using the notion of twist, we obtain

$$\dot{X}(t) \;=\; \dot{g}(t)X_0 \;=\; \dot{g}(t)g^{-1}(t)X(t) \;=:\; \hat{V}X(t) \tag{2.118}$$

with $\hat{V}(t) \in se(3)$. Since $\hat{V}(t)$ is of the form

$$\hat{V}(t) \;=\; \begin{pmatrix} \hat{w} & v \\ \vec{0}^\top & 0 \end{pmatrix} \tag{2.119}$$

we may also write the velocity of p relative to the camera frame in 3-D as

$$\dot{X}(t) \;=\; \hat{w}(t)X(t) + v(t) \tag{2.120}$$

Suppose now that the viewer is in a new coordinate frame displaced relative to a diferent camera frame (relative to which we have coordinates $X(t)$) by a rigid-body transform $g^* \in SE(3)$. Then the coordinates of the same point p relative to this new frame are

$$Y(t) = g^* X(t)\;, \quad \textit{i.e. } X(t) = (g^*)^{-1}\, Y(t) \tag{2.121}$$

We compute the velocity in the new frame:

$$\dot{Y}(t) \;=\; g^* \dot{X}(t) \;\overset{(2.118)}{=}\; g^* \dot{g}(t) g^{-1}(t) X(t) \;=\; g^* \underbrace{\dot{g}(t) g^{-1}(t)}_{=\hat{\xi}(t)} (g^*)^{-1}\, Y(t) \tag{2.122}$$

Consequently, the new twist is

$$\hat{V} \;=\; g^* \hat{\xi}(t)\, (g^*)^{-1} \tag{2.123}$$

This is the same quantity as before, but observed from a different point. The two velocities are related via a mapping defined by the relative motion g^*, the so-called *adjoint map*

$$\mathrm{ad}_{g^*} \;:\; se(3) \;\to\; se(3)\;; \quad \hat{\xi} \;\mapsto\; g^* \hat{\xi}\, (g^*)^{-1} \tag{2.124}$$

It transforms velocity from one frame to another.

Chapter 3

Mathematical Camera Model

We consider in the following an ideal perspective camera. A popular set-up is the *pinhole camera model* for image formation, of which we consider its frontal variant.

The frontal pinhole model is given by

- the optical center 0 of the camera;
- the camera frame attached to its origin 0, yielding a reference frame (X, Y, Z) where the Z-axis is also called *optical axis*;
- an *image plane* (x, y) given by the coordinate $Z = f$ where f is the *focal length*, i.e. it is parallel to the (X, Y) plane and perpendicular to the optical axis;
- the origin of *principal point* of the image plane is located where the optical axis meets the image plane;

Given a point p with world coordinates $\vec{X}_w = (X_w, Y_w, Z_w)^\top$ on a photographed object surface, the corresponding irradiance in the image pixel $\vec{x} = (x, y)^\top$ is found by intersecting $\vec{0p}$ with the image plane.

Let us consider the coordinates of a pixel $\vec{x} = (x, y)^\top$, i.e. more specifically

$$\begin{pmatrix} x \\ y \\ f \end{pmatrix} \sim \begin{pmatrix} x \\ y \end{pmatrix} = \vec{x} \tag{3.1}$$

for f fixed. As we map $\vec{x}$ to the corresponding object point p by stretching the ray $\vec{0x}$ we find

$$p = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \\ f \end{pmatrix} \tag{3.2}$$

Let us point out the relations

$$\frac{1}{f}\left\|\begin{pmatrix} x \\ y \end{pmatrix}\right\| = \frac{1}{\lambda f}\left\|\begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix}\right\| = \frac{1}{Z}\left\|\begin{pmatrix} X \\ Y \end{pmatrix}\right\| \tag{3.3}$$

which holds for

$$\frac{x}{f} = \frac{X}{Z} \quad \text{i.e.} \quad x = f\frac{X}{Z} \tag{3.4}$$

and

$$\frac{y}{f} = \frac{Y}{Z} \quad \text{i.e.} \quad y = f\frac{Y}{Z} \tag{3.5}$$

We may write the underlying projection as

$$\pi : \mathbb{R}^3 \to \mathbb{R}^2 , \quad \vec{X} \mapsto \pi\left(\vec{X}\right) = \vec{x} \tag{3.6}$$

Note that not only p but also any other point along the ray $\lambda \cdot \vec{0x}$ projects onto the same image point $\vec{x} = (x, y)^\top$.

In order to establish a precise correspondence between *(i)* point coordinates in 3-D space w.r.t. fixed world coordinates, and *(ii)* their projected points in a 2-D image w.r.t. a local coordinate frame, one must take into account:

- coordinate transforms between camera frame and world frame;
- projection of 3-D space onto 2-D images;
- coordinate transforms between possible choices of image coordinate frames.

The inversion of a corresponding chain of mappings is sometimes called camera calibration.

Let us get more specific.

3.1 Perspective Camera Desricption

Let us connsider a point p with world coordinates $\vec{X}_w = (X_w, Y_w, Z_w)^\top$. Its coordinates relative to the camera frame are in homogeneous form

$$\begin{pmatrix} \vec{X} \\ 1 \end{pmatrix} = g\begin{pmatrix} \vec{X}_w \\ 1 \end{pmatrix} = \begin{pmatrix} R & T \\ \vec{0}^\top & 1 \end{pmatrix}\begin{pmatrix} \vec{X}_w \\ 1 \end{pmatrix} \tag{3.7}$$

with $g \in SE(3)$.

As by (3.4), (3.5), we see that $\vec{X}$ is projected onto the image plane via

$$\vec{x} = \begin{pmatrix} x \\ y \end{pmatrix} = \frac{f}{Z} \begin{pmatrix} X \\ Y \end{pmatrix} \tag{3.8}$$

In homogeneous coordinates, this can be formulated via the intermediate step

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \frac{f}{Z} \begin{pmatrix} X \\ Y \\ Z/f \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{3.9}$$

as

$$Z \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{3.10}$$

Since the Z coordinate is usually assumed to be the unknown, we may write it as $\lambda \in \mathbb{R}$.

We may rewrite the latter equation as

$$Z\vec{x} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \vec{X} \tag{3.11}$$

where $\vec{x} := \vec{x}_h = (x, y, 1)^\top$ and $\vec{X} := \vec{X}_h = (X, Y, Z, 1)^\top$ are now in homogeneous representation.

One can decompose the matrix in (3.11) as by

$$\begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \underbrace{\begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{=:K_f} \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}}_{\Pi} \tag{3.12}$$

The matrix Π is often denoted as the *canonical projection matrix*.

To summarize, we have the *perspective projection equation*

$$\lambda\vec{x} = \Pi\vec{X} = \Pi g \vec{X}_w \tag{3.13}$$

We need now to take care of the pixel array one obtains from a digital camera.

3.2 Intrinsic Camera Parameters

Capturing digital images with a digital camera, the origin of the ordering of pixels (i, j) is typically in the upper left corner of the image. Also, the size of the pixels must be considered.

Starting with the latter issue, if $(x, y)^\top$ are specified in terms of metric units (e.g. millimeters), and if $(x_s, y_s)^\top$ are scaled versions corresponding to pixel coordinates, then the *scaling* reads as

$$\begin{pmatrix} x_s \\ y_s \end{pmatrix} = \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{3.14}$$

As indicated, we also need to translate the prinicipal point to the upper left corner:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} x + o_x \\ y + o_y \end{pmatrix} \tag{3.15}$$

where $(o_x, o_y)^\top$ are the pixel coordinates of the principla point. In homogeneous coordinates, the above steps give the pixel coordinates $\vec{x}'$ with

$$\vec{x}' = \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} s_x & 0 & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{3.16}$$

In case the pixels are not rectangular, a more general *scaling matrix*

$$\begin{pmatrix} s_x & s_\theta \\ 0 & s_y \end{pmatrix} \tag{3.17}$$

with the *skew factor* s_θ can be employed.

Combining these developments with the projection model (3.13), we have

$$\lambda \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} s_x & s_\theta & o_x \\ 0 & s_y & o_y \\ 0 & 0 & 1 \end{pmatrix}}_{=:K_s} \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{3.18}$$

Let us stress that (3.18) contains two main stages:

1. A perspective projection Π to a normalized coordinate system (as if $f = 1$).
2. A second stage depending on the *intrinsic parameters* of the camera: f, s_x, s_y, s_θ, and o_x, o_y.

With

$$K := K_s K_f = \begin{pmatrix} fs_x & fs_\theta & o_x \\ 0 & fs_y & o_y \\ 0 & 0 & 1 \end{pmatrix} \tag{3.19}$$

we may write the perspective projection equation as

$$\lambda \vec{x} = K\Pi\vec{X} = K\Pi g\vec{X}_w \tag{3.20}$$

Often, the 3×4 matrix $K\Pi g$ is called *general projection matrix.*

Let us stress that (3.20) is nonlinear. Using $\Pi^\top = [\pi_1, \pi_2, \pi_3]$, we see this by dividing by λ, thus obtaining for the entries of $\vec{x}'$

$$x' = \frac{\pi_1^\top \vec{X}_w}{\pi_3^\top \vec{X}_w}\ , \quad y' = \frac{\pi_2^\top \vec{X}_w}{\pi_3^\top \vec{X}_w}\ , \quad z' = 1 \tag{3.21}$$

The nonlinearity arises as by the division.

When the *calibration matrix* K is known, the *calibrated coordinates* $\vec{x}$ can be obtained via its inverse:

$$\lambda \vec{x} = \lambda K^{-1}\vec{x}' = \Pi\vec{X} = \Pi g\vec{X}_w \tag{3.22}$$

One may interprete the latter equation as an ideal perspective camera model.

3.3 Projective Geometry of the Perspective Camera

Let us point out an ambiguity in the perspective projection: Two vectors x and y in $\mathbb{R}^3$ may represent the same image point as long as $x = \alpha y$ for some $\alpha \in \mathbb{R}$. We now elaborate on useful concepts that help in avoiding confusion.

Consider the perspective projection of a straight line L living in 3-D space onto a 3-D image plane by our perspective camera.

To define L in 3-D, we specify a *base point* p_0 and a *free vector* v that indicates the direction of the line. With homogeneous coordinates

$$p_0 \ : \ \vec{X}_0 = (X_0, Y_0, Z_0, 1)^\top\ , \quad v \ : \ \vec{V} = (V_1, V_2, V_3, 0)^\top \tag{3.23}$$

the line L can be expressed as

$$\vec{X} = \vec{X}_0 + \lambda\vec{V}\ , \quad \lambda \in \mathbb{R} \tag{3.24}$$

Making use of the standard projection Π we obtain the image points

$$\vec{x} = \Pi\vec{X} = \Pi\vec{X}_0 + \lambda\Pi\vec{V} \tag{3.25}$$

Evidently, the set of all possible $\{\vec{x}\}$, understood as vectors in the camera frame with origin $\vec{o}$, spans a 2-D subspace of $\mathbb{R}^3$. Intersecting this subspace with the image plane gives the image of L.

This process is not injective:

Definition 3.3.1 *A* preimage *of a point or a line in the image plane is the set of 3-D points that give rise to an image equal to the given point or line.*

The preimage of L is a plane P through $\vec{o}$. Its intersection with the image plane is the image of L.

As an alternative representation, a plane may be described by its normal vector. This gives rise to:

Definition 3.3.2 *The* coimage *of a point or line is defined to be the subspace in* $\mathbb{R}^3$ *that is the orthogonal complement of its preimage.*

Let us stress that image, preimage and coimage are equivalent representations.

For

$$\vec{l} = (a, b, c)^\top \tag{3.26}$$

describing the coimage of L, we have for $\vec{x}$ being the image of a point p on L

$$\vec{l}^\top \vec{x} = 0 \tag{3.27}$$

Recall that $\hat{l} \in \mathbb{R}^{3\times3}$ denotes the skew symmetric matrix associated to $\vec{l} \in \mathbb{R}^3$. Since the column vectors of $\hat{l}$ span the subspace orthogonal to $\vec{l}$, they also span the preimage of L, i.e.

$$P = \text{span}\left(\hat{l}\right) \tag{3.28}$$

Similarly, if $\vec{x}$ is the image of a point p, its coimage is the plane orthogonal to $\vec{x}$ given by $\text{span}\,(\hat{x})$.

The relations between preimage or coimage of points or lines can be summarized as

$$\underbrace{\hat{x}\vec{x} \; = \; \vec{0}}_{\text{coimage of a point}} \qquad \text{and} \qquad \underbrace{\hat{l}\vec{l} \; = \; \vec{0}}_{\text{preimage of a plane}} \tag{3.29}$$

Turning to the use of homogeneous coordinates, we observe that the two vectors

$$(X, Y, Z, 1)^\top \quad \text{and} \quad (\lambda X, \lambda Y, \lambda Z, \lambda)^\top \tag{3.30}$$

represent the same point in $\mathbb{R}^3$. Similarly, $(x', y', z')^\top$ represents a point $(x, y, 1)^\top$ on the 2-D image plane as long as

$$x \cdot z' \; = \; x' \quad \text{and} \quad y \cdot z' \; = \; y' \tag{3.31}$$

Let us tackle the question what happens if the last entry in homogeneous coordinates is zero. To this end we define:

Definition 3.3.3 *An n-dimensional projective space $\mathbb{P}^n$ is the set of all 1-dimensional subspaces of the vector space $\mathbb{R}^{n+1}$. A point in $\mathbb{P}^n$ can then be assigned homogeneous coordinates $\vec{X} = (x_1, x_2, \ldots, x_{n+1})^\top$ where at least one entry is nonzero. For any $\lambda \neq 0$ the coordinates $\vec{Y} = (\lambda x_1, \lambda x_2, \ldots, \lambda x_{n+1})^\top$ represent the same point p in $\mathbb{P}^n$:*

$$\vec{X} \; \sim \; \vec{Y} \tag{3.32}$$

Consequently, $\mathbb{R}^n$ with its homogeneous representation is the subset of $\mathbb{P}^n$ that excludes $\vec{X} = (x_1, x_2, \ldots, x_n, 0)^\top$. In addition, we may always normalize vecX to $x_{n+1} = 1$.

Recalling the perspective camera model

$$\lambda \vec{x}' \; = \; K \Pi g \vec{X}_w \tag{3.33}$$

with the general projection matrix $\pi := K\Pi g \in \mathbb{R}^{4\times4}$, then the camera model is a projection from a 3-D projective space $\mathbb{P}^3$ to a 2-D projective space $\mathbb{P}^2$:

$$\pi \; : \; \mathbb{P}^3 \to \mathbb{P}^2 \; ; \quad \vec{X}_w \mapsto \vec{x}' \sim \pi \vec{X}_w \tag{3.34}$$

where we can omit λ now by use of the equivalence in the homogeneous case.

Let us now consider $\vec{X} = (x, y, z, \varepsilon)^\top$ with $\varepsilon \downarrow 0$.

If we would normalize for small ε to 1 in the last entry, we would get

$$\vec{X} = \left(\frac{x}{\varepsilon}, \frac{y}{\varepsilon}, \frac{z}{\varepsilon}, 1\right)^\top \tag{3.35}$$

so that we describe a point with 3-D coordinates $\vec{X} = \left(\frac{x}{\varepsilon}, \frac{y}{\varepsilon}, \frac{z}{\varepsilon}\right)^\top$ which is far away from the origin. Letting ε drop to zero, we interprete a point $\vec{X} = (x, y, z, 0)^\top$ as infinitely far away from the origin.

All such points form a *plane at infinity* called P_∞, described by

$$(0, 0, 0, 1)^\top \vec{X} = 0 \tag{3.36}$$

The model (3.34) is then well-defined on $\mathbb{P}^3$ including points at infinity.

A classical application of projective geometry is the computation of *vanishing points* which we will briefly describe.

Two parallel lines do not intersect in 3-D. Let us make use of the homogeneous free vector

$$\vec{V} = (V_1, V_2, V_3, 0)^\top \tag{3.37}$$

indicating the direction of two parallel lines L_1, L_2. As two base points that lines let us fix

$$\vec{X}_a = (X_a, Y_a, Z_a, 1)^\top \quad \text{and} \quad \vec{X}_b = (X_b, Y_b, Z_b, 1)^\top \tag{3.38}$$

respectively. The homogeneous coordinates of points on L_1, L_2 can be written as

$$\vec{X}_1(\eta) = \vec{X}_a + \eta\vec{V} \quad \text{and} \quad \vec{X}_x(\eta) = \vec{X}_b + \eta\vec{V} \tag{3.39}$$

for $\eta \in \mathbb{R}$. Letting $\eta \to \infty$ in $\vec{X}_1$ and $\vec{X}_2$ shows that asymptotically

$$\vec{X}_1(\infty) = \vec{X}_2(\infty) = \vec{V} \tag{3.40}$$

Thus, the two lines intersect at infinity at the point $\vec{V}$. The image of the intersections is given by

$$\vec{x}' = \Pi\vec{V} \tag{3.41}$$

which is exactly the vanishing point.

Practically, one may look for the images of $\vec{X}_1$ and $\vec{X}_2$ for $\eta \to \infty$. Then the vanishing point can easily be computed.

Chapter 4

Geometry of Two Views

In this chapter we discuss the basic set-up of stereo vision.

To this end, we assume that the *two cameras* used for acquiring a static scene are calibrated. We recall, that the calibration matrix K in the general projection matrix $K\Pi g$ is known in this case, and that the calibrated coordinates $\vec{x}$ are

$$\lambda\vec{x} \;=\; \Pi\vec{X} \;=\; \Pi g\vec{X}_w \tag{4.1}$$

4.1 Epipolar Geometry

We assume that we associate with each of the two cameras an orthonormal camera frame.

We denote their optical centers by $\vec{o}_1$, $\vec{o}_2$ (usually given in world coordinates) and align their Z-axes with their optical axes.

For simplicity we assume that the world frame is identical to the camera frame corresponding to $\vec{o}_1$.

If $\vec{X}_1$ and $\vec{X}_2$ are the coordinates of a point p in the two camera frames, then they relate as

$$\vec{X}_2 \;=\; R\vec{X}_1 + T \tag{4.2}$$

We now turn to the two images acquired by the two cameras. We are interested in the projection of one and the same point p onto the two image planes, and we consider homogeneous coordinates $\vec{x}_1$ and $\vec{x}_2$ of p.

Our goal is to build a direct relationship between $\vec{x}_1$ and $\vec{x}_2$. We have

$$\vec{X}_i \;=\; \lambda_i\vec{x}_i\,, \quad i = 1, 2 \tag{4.3}$$

which implies by (4.2) that there are λ_i with

$$\lambda_2 \vec{x}_2 \;=\; R\lambda_1 \vec{x}_1 + T \tag{4.4}$$

Now we simplify this expression. In a first step we want to get eliminate T. This is done by multiplication from the left with $\hat{T}$, compare (1.29):

$$\lambda_2 \hat{T}\vec{x}_2 \;=\; \lambda_1 \hat{T}R\vec{x}_1 + \underbrace{\hat{T}T}_{=T\times T=\vec{0}} \tag{4.5}$$

Taking into account

$$\hat{T}\vec{x}_2 \;=\; T \times \vec{x}_2 \;\perp\; \vec{x}_2 \tag{4.6}$$

we obtain after multiplying (4.5) from the left with $\vec{x}_2^\top$

$$\lambda_1 \vec{x}_2^\top \hat{T}R\vec{x}_1 \;=\; \lambda_2 \underbrace{\vec{x}_2^\top \hat{T}\vec{x}_2}_{=\langle \vec{x}_2, \hat{T}\vec{x}_2\rangle = 0} \;=\; 0 \tag{4.7}$$

Since $\lambda_1 \neq 0$ we have shown:

Theorem 4.1.1 *Consider two images $\vec{x}_1$, $\vec{x}_2$ of one and the same point p. Then the* epipolar constraint *or* essential constraint

$$\langle \vec{x}_2, \hat{T}R\vec{x}_1\rangle \;=\; 0 \tag{4.8}$$

holds, where $R \in SO(3)$ and $T \in \mathbb{R}^3$ are the relative orientation and the relative position of the cameras.

The matrix $\hat{T}R$ is said to encode the *relative pose* between the two cameras.

Definition 4.1.1 *The matrix*

$$E \;:=\; \hat{T}R \tag{4.9}$$

is called essential matrix*.*

The epipolar constraint can also be derived using geometric considerations.

Since $\vec{o}_1$, $\vec{o}_2$ and p define a triangle in a plane, also the following vectors are in this plane:

- $\vec{x}_1$, pointing from $\vec{o}_1$ to p;

- $\vec{x}_2$, pointing from $\vec{o}_2$ to p;
- $T := \vec{o_2 o_1}$, defining a direction that connexts the origins of the two camera frames.

Since $\vec{X}_2 = R\vec{X}_1 + T$, it obviously holds

$$R\vec{x}_1 \ \| \ \vec{x}_2 \tag{4.10}$$

i.e. the orientations are identical, they are only distinguished by a translation T and are a member of the same plane. As a remark, we must have by the latter equation

$$\langle \vec{x}_2, R\vec{x}_1 \rangle \ \neq \ 0 \tag{4.11}$$

As $T \in \mathbb{R}^3$ is (a free vector defining an orientation) in the same plane as $\vec{x}_2$ and $R\vec{x}_1$, we can apply the cross product to obtain

$$\langle \vec{x}_2, T \times R\vec{x}_1 \rangle \ = \ \langle \vec{x}_2, \hat{T}R\vec{x}_1 \rangle \ = \ 0 \tag{4.12}$$

which is exactly (4.8).

Let us define some other important geometric entities.

Definition 4.1.2 *In the epipolar geometry we have:*

- *The plane defined by $(\vec{o}_1, \vec{o}_2, p)$ is called epipolar plane: For a fixed camera configuration, there is one epipolar plane for each point p.*
- *The projection $\vec{e}_k$ of a camera center $\vec{o}_l$ onto the corresponding image plane of the other camera $\vec{o}_k$ is called an epipole.*
- *The intersection of the epipolar plane $(\vec{o}_1, \vec{o}_2, p)$ with the image plane belonging to $\vec{o}_k$ is a line $\vec{l}_k$ called the epipolar line of p.*

We will ususally use the normal vector $\vec{l}_k$ to the epipolar plane to represent the epipolar line. This means, the vector $\vec{l}_k$ will be the *coimage* of the epipolar line.

Let us remark, that one needs to take care of the different coordinate systems in which the described geometric entities are given. For instance, the epipole $\vec{e}_1$ denoting the projection of $\vec{o}_2$ onto the first image plane is given in the *first* camera frame. In contrast, $\vec{e}_2$ is given in the local coordinates of the second camera frame. Analogous assertions hold for $\vec{l}_1$ and $\vec{l}_2$.

The epipoles and the epipolar line have some interesting properties upon which we now elaborate.

First we consider the two epipoles $\vec{e}_i \in \mathbb{R}^3$ and their relation to the essential matrix E.

For what follows it is useful to clarify the meaning of the translation T. The vector T is $o_2\vec{o}_1$ in terms of the camera frame associated with $\vec{o}_2$, and one interprete the free vector T as originating at $\vec{o}_2$.

We recall that the meaning of $\vec{X}_2 = R\vec{X}_1 + T$ is that of a coordinate transform, thus we may ask for $\vec{X}_1$ that gives us $\vec{X}_2 = \vec{o}_2 = (0,0,0)^\top$ (in terms of the second coordinate frame), i.e.

$$\vec{0} \;=\; R\vec{X}_1 + T \tag{4.13}$$

This yields

$$R\vec{X}_1 \;=\; -T \quad \Leftrightarrow \quad \vec{X}_1 \;=\; -R^\top T \tag{4.14}$$

Because the sought vector $\vec{X}_1$ is a scalar multiple of the epipole vector $\vec{e}_1$ (where both are given in terms of the first coordinate frame) – this is obvious as $\vec{e}_1$ points from $\vec{o}_1$ to $\vec{o}_2$, meeting a corresponding point on the first image plane – we may just write

$$\vec{e}_1 \;\sim\; R^\top T \tag{4.15}$$

By (4.15) we can compute

$$E\vec{e}_1 \;=\; \hat{T}R\vec{e}_1 \;\sim\; \hat{T}\underbrace{RR^\top}_{=I}T \;=\; \hat{T}T \;=\; \vec{0} \tag{4.16}$$

which will be our first result of interest.

Because of the abovementioned interpretation of T we may write immediately

$$\vec{e}_2 \;\sim\; T \tag{4.17}$$

as these are both originating in $\vec{o}_2$ and are given in terms of the second camera frame. Thus follows our second result of interest

$$\vec{e}_2^\top E \;=\; \vec{e}_2^\top \hat{T}R \;\sim\; T^\top \hat{T}R \;=\; \left(-\hat{T}T\right)^\top R \;=\; -\vec{0}^\top R \;=\; \vec{0}^\top \tag{4.18}$$

Let us summarize our computations.

Theorem 4.1.2 *The epipoles $\vec{e}_i \in \mathbb{R}^3$ are the left and right kernels of E, respectively.*

We now focus on the epipolar lines.

We remember that for all the points in the epipolar plane holds by $\vec{X}_2 = R\vec{X}_1 + T$

$$\lambda_2 \vec{x}_2 = \lambda_1 R \vec{x}_1 + T \tag{4.19}$$

for some $\lambda_i \in \mathbb{R}$ which gives

$$\lambda_2 \hat{T} \vec{x}_2 = \lambda_1 \hat{T} R \vec{x}_1 + \underbrace{\hat{T} T}_{=\vec{0}} \tag{4.20}$$

Focusing on the right hand side of (4.20) we have

$$\hat{T} R \vec{x}_1 \sim E \vec{x}_1 \tag{4.21}$$

On the left hand side of (4.20) we obtain

$$\hat{T} \vec{x}_2 = T \times \vec{x}_2 \sim \vec{l}_2 \tag{4.22}$$

since T and $\vec{x}_2$ both hit the epipolar line and are in the epipolar plane: Therefore the cross product gives a vector perpendicular to the epipolar plane, and the definition of $\vec{l}_2$ meets exactly this requirement.

Similarly, we have by $\vec{X}_2 = R\vec{X}_1 + T$

$$\lambda_2 R^\top \hat{T}^\top \vec{x}_2 = \lambda_1 R^\top \hat{T}^\top R \vec{x}_1 + R^\top \underbrace{\hat{T}^\top T}_{=-\hat{T}T=\vec{0}} \tag{4.23}$$

For the remaining term on the right hand side holds

$$\lambda_1 R^\top \hat{T}^\top R \vec{x}_1 \sim R^\top \hat{T} R \vec{x}_1 \tag{4.24}$$

where R and $R^\top$ are rotations in the epipolar plane:

- $R\vec{x}_1 \sim \vec{v}_1$ where $\vec{v}_1$ points to a location on the epipolar line $\vec{l}_1$;
- $\hat{T}\vec{v}_1 \sim \vec{v}_2$ gives a vector perpendicular to the epipolar line $\vec{l}_1$ (so that $\sim \vec{l}_1$ if understanding $\vec{l}_1$ as the coimage of the epipolar line);
- $R^\top \vec{v}_2 \parallel \vec{v}_2$ since $R^\top$ is a rotation in the epipolar plane i.e. it does not change the orientation of $\vec{v}_2$.

By the above construction follows

$$R^\top \hat{T}^\top \vec{x}_2 \sim \vec{l}_1 \quad \Leftrightarrow \quad E^\top \vec{x}_2 \sim \vec{l}_1 \tag{4.25}$$

Let us summarize:

Theorem 4.1.3 *The coimages $\vec{l}_1$, $\vec{l}_2$ of the epipolar lines can be expressed as*

$$\vec{l}_2 \sim E\vec{x}_1 \quad and \quad \vec{l}_1 \sim E^\top \vec{x}_2 \tag{4.26}$$

Moreover, in the epipolar geometry the following properties are obvious:

Proposition 4.1.1 *In each image, both the image point and the epipole lie on the epipolar line, i.e.*

$$\langle \vec{l}_i, \vec{e}_i \rangle = 0 \quad and \quad \langle \vec{l}_i, \vec{x}_i \rangle = 0 \tag{4.27}$$

for $i = 1, 2$.

4.2 The Essential Matrix

In this paragraph we consider basic properties of the essential matrix.

An useful tool will be the *singular value decomposition (SVD)* which we now briefly review.

Theorem 4.2.1 *If $M \in \mathbb{R}^{m\times n}$ then there exist orthogonal matrices $U = [u_1, \ldots, u_m] \in \mathbb{R}^{m\times m}$ and $V = [v_1, \ldots, v_n] \in \mathbb{R}^{n\times n}$ such that*

$$U^\top M V = \Sigma = diag(\sigma_1, \ldots, \sigma_p) \tag{4.28}$$

with

$$\sigma_1 \geq \ldots \geq \sigma_p \geq 0, \qquad p = \min(m, n) \tag{4.29}$$

The decomposition (4.28), (4.29), is called SVD of M.

The numbers σ_i are the *singular values* of M, the set $\{\sigma_1, \ldots, \sigma_p\}$ is called the *singular value spectrum.*

A fundamental property of the SVD is as follows. If

$$\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_p = 0 \tag{4.30}$$

then $\text{rank}(M) = r$, and as by

$$U^\top M V = \Sigma \quad \Leftrightarrow \quad M = U\Sigma V^\top \tag{4.31}$$

we can represent M as by the *SVD expansion*

$$M = \sum_{i=1}^{r} \sigma_i u_i v_i^\top \tag{4.32}$$

The latter gives the basis for the following assertion:

Theorem 4.2.2 *[Eckart-Young-Mirsky matrix approximation theorem] Let the SVD of $M \in \mathbb{R}^{m\times n}$ be given by*

$$M \;=\; \sum_{i=1}^{r} \sigma_i u_i v_i^\top \tag{4.33}$$

with $r = rank(M)$. If $k < r$ and

$$M_k \;:=\; \sum_{i=1}^{k} \sigma_i u_i v_i^\top \tag{4.34}$$

then

$$\min_{\substack{N\in\mathbb{R}^{m\times n}\\ rank(N)=k}} \|M-N\|_2 \;=\; \|M-M_k\|_2 \;=\; \sigma_{k+1} \tag{4.35}$$

and

$$\min_{\substack{N\in\mathbb{R}^{m\times n}\\ rank(N)=k}} \|M-N\|_F \;=\; \|M-M_k\|_F \;=\; \sqrt{\sum_{i=k+1}^{p} \sigma_i^2}, \qquad p=\min(m,n) \tag{4.36}$$

The *2-norm of a matrix* is defined via

$$\|A\|_2 \;:=\; \sup_{\|\vec{x}\|_2=1} \|A\vec{x}\|_2 \tag{4.37}$$

which can be interpreted as the maximal stretching factor of the mapping $f(\vec{)} = A\vec{x}$, $\vec{x} \in \mathbb{R}^n$.

The *Frobenius-norm* $\|.\|_F$ has no geometric interpretation, but it is easier to compute than the 2-norm of a matrix:

$$\|A\|_F \;:=\; \sqrt{\sum_{i,j} (a_{ij})^2}, \tag{4.38}$$

Moreover, an there is an *inclusion property*

$$\|A\|_2 \;\le\; \|A\|_F \;\le\; \sqrt{p}\|A\|_2\,, \qquad p=\min(m,n) \tag{4.39}$$

by which one can infer from $\|A\|_F$ on the range of $\|A\|_2$.

Put into simple words, the above theorem states that the minimal distance between a given matrix M and a matrix of lower rank N with $\text{rank}(N) = k$, is computable by the $(k+1)$-th singular value.

In addition to the SVD, we will make use of a rotation matrix $R_Z(\theta)$ describing a rotation about the Z-axis by an angle of θ radians.

The latter matrix is conveniently written by using exponential coordinates as

$$R_Z(\theta) \;=\; e^{\hat{w}\theta} \tag{4.40}$$

with $\vec{w} = (0,0,1)^\top = \vec{e}_3$. Then

$$R_Z\left(+\frac{\pi}{2}\right) \;=\; \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.41}$$

For a given rotation matrix $R \in SO(3)$ and a translation $T \in \mathbb{R}^3$, the essential matrix E is determined by $E = \hat{T}R$. We will now follow the inverse route: Given a matrix E, can we see if it is essential, and what are R and T?

Theorem 4.2.3 *A nonzero matrix $E \in \mathbb{R}^{3\times 3}$ is an essential matrix if and only if E has a SVD of the form $E = U\Sigma^\top V^\top$ with*

$$\Sigma \;=\; \mathrm{diag}(\sigma, \sigma, 0) \tag{4.42}$$

for some $\sigma > 0$ and $U, V \in SO(3)$.

Let us comment on the proof on *sufficiency*, since it includes a constructive step.

If a given matrix $E \in \mathbb{R}^{3\times 3}$ has a non-standard SVD – i.e. U, V with $E = U\Sigma^\top V^\top$ are not only orthogonal, but in $SO(3)$ – where $\Sigma = \mathrm{diag}(\sigma, \sigma, 0)$, let us define (R_1, T_1) and (R_2, T_2) in $SE(3)$ to be

$$\begin{cases} R_1 := UR_Z^\top(+\frac{\pi}{2})V^\top, & \hat{T}_1 := UR_Z(+\frac{\pi}{2})\Sigma U^\top \\ R_2 := UR_Z^\top(-\frac{\pi}{2})V^\top, & \hat{T}_2 := UR_Z(-\frac{\pi}{2})\Sigma U^\top \end{cases} \tag{4.43}$$

One can validate that $\hat{T}_k$ as defined above are skew symmetric matrices fitting the hat notation.

We can confirm by a simple computation

$$\hat{T}_1 R_1 \;=\; U\Sigma^\top V^\top \;=\; \hat{T}_2 R_2 \tag{4.44}$$

Thus, $E = \hat{T}_1 R_1 = \hat{T}_2 R_2$ is an essential matrix.

The SVD has been used to construct *two* possible pairs of (R, T) for a given essential matrix E. Are these all possibilities?

We prepare the answer to this question by stating first the following

Lemma 4.2.1 *Consider an arbitrary nonzero skew symmetric matrix* $\hat{T} \in so(3)$ *for* $T \in \mathbb{R}^3$*. If for* $R \in SO(3)$*,* $\hat{T}R$ *is again a skew symmetric matrix, it follows* $R = I$ *or* $R = e^{\hat{v}\pi}$*, where* $\vec{v} = T/\|T\|$*.*

Proof. For proving the above lemma, let us assume that T (and later $\vec{v}$) is of unit length. Since $\hat{T}R$ shall be skew symmetric, it shall hold

$$\left(\hat{T}R\right)^\top = -\hat{T}R \tag{4.45}$$

This equation gives

$$\hat{T} = R\hat{T}R \tag{4.46}$$

Since R is a rotation matrix we can write it as

$$R = e^{\hat{w}\theta} \tag{4.47}$$

with some vector $\vec{w} \in \mathbb{R}^3$, $\|\vec{w}\| = 1$, performing as rotation axis, and with some rotation angle θ.

For $\theta = 0$, we have $R = I$.

For $\theta \neq 0$, we put (4.47) into (4.46) and apply the mappings at $\vec{w}$ to obtain

$$e^{\hat{w}\theta}\hat{T}e^{\hat{w}\theta}\vec{w} = \hat{T}\vec{w} \tag{4.48}$$

Since $e^{\hat{w}\theta} = \vec{w}$ for all θ, because $e^{\hat{w}\theta}$ rotates about $\vec{w}$, it follows

$$e^{\hat{w}\theta}\hat{T}\vec{w} = \hat{T}\vec{w} \tag{4.49}$$

We have that $\vec{w}$ is exactly the only eigenvector of $e^{\hat{w}\theta}$ with eigenvalue 1. Also, $\hat{T}\vec{w} = T \times \vec{w}$ must be orthogonal to $\vec{w}$, so that $e^{\hat{w}\theta}\hat{T}\vec{w}$ describes the rotation of the vector $\hat{T}\vec{w} \perp \vec{w}$ about the axis $\vec{w}$.

The result can only meet (4.49) if θ describes a full rotation (so that $R = I$), or if $\hat{T}\vec{w} = \vec{0}$.

The condition $\hat{T}\vec{w} = \vec{0}$ implies $\vec{w} = \pm T$. Thus it follows

$$e^{\pm\hat{T}\theta}\hat{T}e^{\pm\hat{T}\theta} = \hat{T} \tag{4.50}$$

By the geometrical interpretations of $\hat{T}$ and $e^{\pm\hat{T}\theta}$ it is obvious that

$$\hat{T}e^{\pm\hat{T}\theta} = e^{\pm\hat{T}\theta}\hat{T} \tag{4.51}$$

such that (4.50) rewrites to

$$e^{\pm 2\hat{T}\theta}\hat{T} = \hat{T} \tag{4.52}$$

Since this implies that $e^{\pm 2\hat{T}\theta}$ describes a full rotation about T, it follows $\theta = \pi$. ■

We are now ready to show

Theorem 4.2.4 *There exist only two relative poses (R, T), $R \in SO(3)$, $T \in \mathbb{R}$, corresponding to a nonzero essential matrix. These poses are given by (4.43).*

Proof. Assume that $(R_1, T_1) \in SE(3)$ and $(R_2, T_2) \in SE(3)$ are both solutions of $\hat{T}R = E$. Then it holds

$$\hat{T}_1 R_1 = \hat{T}_2 R_2 \tag{4.53}$$

yielding

$$\hat{T}_2 = \hat{T}_1 R_1 R_2^\top \tag{4.54}$$

Since $\hat{T}_1$ and $\hat{T}_2$ are both skew symmetric and $R := R_1 R_2^\top$ is a rotation matrix, we can apply the preceeding lemma.
We obtain that *either* $R = I$, i.e.

$$R_1 R_2^\top = I \quad \Leftrightarrow \quad R_1 = R_2 \tag{4.55}$$

and consequently also $T_1 = T_2$, *or*

$$R_1 R_2^\top = e^{\hat{T}_1 \pi} \tag{4.56}$$

i.e.

$$R_1 = R_2 e^{\hat{T}_1 \pi} \tag{4.57}$$

This equality implies

$$R_2 = e^{\hat{T}_1 \pi} R_1 = R_1 e^{\hat{T}_1 \pi} \tag{4.58}$$

and because of $\hat{T}_1 R_1 = \hat{T}_2 R_2$ follows

$$\begin{aligned} \hat{T}_2 e^{\hat{T}_1 \pi} R_1 &= \hat{T}_1 R_1 \\ \Leftrightarrow \hat{T}_2 e^{\hat{T}_1 \pi} &= \hat{T}_1 \\ \Leftrightarrow \hat{T}_2 &= \hat{T}_1 e^{\hat{T}_1 \pi} \end{aligned} \tag{4.59}$$

so that

$$\hat{T}_2 = -\hat{T}_1 \quad \Leftrightarrow \quad \hat{T}_1 = -\hat{T}_2 \tag{4.60}$$

by the geometric interpretation of $\hat{T}_1 e^{\hat{T}_1 \pi} \vec{a} = T_1 \times e^{\hat{T}_1 \pi} \vec{a}$ for $\vec{a} \in \mathbb{R}^3$.

In summary we have

$$(R_2, T_2) \;=\; (R_1 e^{\hat{T}_1 \pi}, -T_1) \tag{4.61}$$

so that there are only two poses as in (4.43). ■

Let us now come back to the epipolar constraint.

Having two possible poses at hand, let us also point out, that by the epipolar constraint

$$\langle \vec{x}_2, \hat{T} R \vec{x}_1 \rangle \;=\; 0 \tag{4.62}$$

as in (4.8), not only $\hat{T}R$ but also $-\hat{T}R$ is essential (assuming also that we normalise, else there is a more general scaling factor). This means, we have in fact two essential matrices $\pm E$, each of them leading to two possible poses.

For $E = \hat{T}R$ the entries can be written as

$$E \;=\; [E_1, E_2, E_3] \;=\; \begin{pmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{pmatrix} \in \mathbb{R}^{3\times 3} \tag{4.63}$$

Stacking them columnwise into a vector $E^s \in \mathbb{R}^9$, the *stacked version* of E, reads as

$$E^s \;:=\; \left[E_1^\top, E_2^\top, E_3^\top\right] \;=\; (e_{11}, e_{12}, e_{13}, e_{21}, e_{22}, e_{23}, e_{31}, e_{32}, e_{33})^\top \tag{4.64}$$

The inverse operation, building the matrix E from E^s is called *unstacking.*

We will make use of the *Kronecker product* $\otimes$ of the two vectors $\vec{x}_1 = (x_1, y_1, z_1)^\top$ and $\vec{x}_2 = (x_2, y_2, z_2)^\top$. We set

$$\vec{a} \;:=\; \vec{x}_1 \otimes \vec{x}_2 \tag{4.65}$$

where the Kronecker product of vectors is a special case of the Kronecker product of matrices:

Definition 4.2.1 *Given two matrices A, B with $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{k\times l}$, their Kronecker product $A \otimes B$ is the new matrix*

$$A \otimes B \;:=\; \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ \vdots & \ddots & \ddots & \vdots \\ a_{m1}B & a_{12}B & \cdots & a_{mn}B \end{pmatrix} \in \mathbb{R}^{mk\times nl} \tag{4.66}$$

Thus for vectors A and B with $n = l = 1$, the product $A \otimes B$ is also a vector in $\mathbb{R}^{mk}$.

As by (4.66) we compute

$$\vec{a} \;=\; \vec{x}_1 \otimes \vec{x}_2 \;=\; \left(x_1x_2, x_1y_2, x_1z_2, x_2x_2, x_2y_2, x_2z_2, x_3x_2, x_3y_2, x_3z_2\right)^\top \in \mathbb{R}^9 \quad (4.67)$$

It is easy to validate then that we can reformulate the epipolar constraint $\langle \vec{x}_2, E\vec{x}_1 \rangle = \vec{x}_2^\top E\vec{x}_1 = 0$ as

$$\vec{a}E^s \;=\; 0 \quad (4.68)$$

Let us note that this formulation emphasizes the linear dependence of the epipolar constraint on the entries of E.

We now consider a set of corresponding image points $(\vec{x}_1^i, \vec{x}_2^i)$, $i = 1, \ldots, k$, and we define a matrix $M \in \mathbb{R}^{k \times 9}$ associated with this set:

$$M \;:=\; \left[\vec{a}^1, \vec{a}^2, \ldots, \vec{a}^k\right]^\top \quad (4.69)$$

where the i-th row $\vec{a}^i$ is given as by

$$\vec{a}^i \;:=\; \vec{x}_1^i \otimes \vec{x}_2^i \quad (4.70)$$

Ideally, by (4.68) the vector E^s shall satisfy

$$ME^s \;=\; \vec{0} \quad (4.71)$$

In order to obtain E by input correspondences $(\vec{x}_1^i, \vec{x}_2^i)$, we thus may solve (4.71) for the vector E^s.

Considering the solution of a homogeneous linear system

$$A\vec{x} \;=\; \vec{0}\,, \quad A \in \mathbb{R}^{m\times n} \text{ for } M > n,\ \vec{x} \in \mathbb{R}^n,\ \vec{0} \in \mathbb{R}^m \quad (4.72)$$

the solution is trivial (i.e. $\vec{x} = \vec{0}$) if $\text{rank}(A) = n$. In order to have a unique but non-trivial solution – meaning, up to a scalar factor, since the solution of $A\vec{x} = \vec{0}$ have a subspace structure – it must hold that $\text{rank}(A) = n - 1$.

Applying these theoretical considerations at (4.71), we see that we must have

$$\text{rank}(M) \;=\; 8 \quad \text{and} \quad k \geq 8 \quad (4.73)$$

Since in reality correspondences will be not perfectly arranged as e.g. due to noise, we need to consider how we might deal with deviations from these exact requirements.

4.3 The Algorithm of Longuet and Higgins

Let us consider again

$$ME^s \;=\; \vec{0}\,,\; M \in \mathbb{R}^{k\times 9}\,,\; E^s \in \mathbb{R}^9 \tag{4.74}$$

for computing the essential matrix E.

In general, the number k of measurements that we have at hand will be relatively large. Thus, the k equations in (4.74) will be contradictore to each other, i.e. the system will be overdetermined.

A possible approach to deal with this issue is to minimise the overall error in (4.74), i.e.

$$\|ME^s - \vec{0}\|^2 \;=\; \|ME^s\|^2 \tag{4.75}$$

We tackle this *least squares (LS) problem* as follows. Minimising (4.75) means to look for the non-trivial minimiser of

$$\langle M\vec{x}, M\vec{x}\rangle \;=\; (M\vec{x})^\top (M\vec{x}) \;=\; \vec{x}^\top M^\top M\vec{x} \tag{4.76}$$

The matrix $M^\top M \in \mathbb{R}^{9\times 9}$ is in our problem setting of full rank and it is symmetric. This means, it has a complete set of orthonormal eigenvectors $\{\vec{v}_I\}$ spanning the $\mathbb{R}^9$.
This means, we can express $\vec{x}$ via the eigenvectors $\vec{v}_i$

$$\vec{x} \;=\; \alpha_1\vec{v}_1 + \ldots + \alpha_9\vec{v}_9 \tag{4.77}$$

Then with corresponding eigenvalues $\lambda_i > 0$ it holds

$$\begin{aligned}
\vec{x}^\top M^\top M\vec{x} &= (\alpha_1\vec{v}_1 + \ldots + \alpha_9\vec{v}_9)^\top M^\top M\,(\alpha_1\vec{v}_1 + \ldots + \alpha_9\vec{v}_9) \qquad (4.78)\\
&= (\alpha_1\vec{v}_1 + \ldots + \alpha_9\vec{v}_9)^\top (\alpha_1\lambda_1\vec{v}_1 + \ldots + \alpha_9\lambda_9\vec{v}_9)\\
&\overset{\vec{v}_i \text{ orthonormal}}{=} \lambda_1\alpha_1^2 + \ldots + \lambda_9\alpha_9^2\\
&> \min_i \lambda_i \left(\alpha_1^2 + \ldots + \alpha_9^2\right)
\end{aligned}$$

Assuming that E^s is of unit length it holds

$$\alpha_1^2 + \ldots + \alpha_9^2 \;=\; 1 \tag{4.79}$$

Therefore, the minimiser we are looking for is given by the eigenvector $\vec{v}$ of the smallest eigenvalue $\lambda := \min_i \lambda_i$ of $M^\top M$.

In practice, this is done by computing the SVD of M:

$$M \;=\; U\Sigma V^\top \tag{4.80}$$

Then

$$
\begin{aligned}
M^\top M &= \left(U\Sigma V^\top\right)^\top U\Sigma V^\top &(4.81)\\
&= V\Sigma^\top \underbrace{U^\top U}_{=I} \Sigma V^\top \\
&= V \underbrace{\Sigma^\top \Sigma}_{=\text{diag}(\sigma_1^2,\ldots,\sigma_9^2)\in\mathbb{R}^{9\times 9}} V^\top
\end{aligned}
$$

where the matrix V contains the eigenvectors of $M^\top M$, which can easily be seen by comparing (4.81) with the principal axis transform of $M^\top M$, i.e. $M^\top M = WDW^\top$ wher W contains the orthonormal eigenvectors and D is a diagonal matrix with the eigenvalues.

As by $\sigma_i \geq \sigma_{i+1}$ we are interested in $\lambda = \sigma_9^2$, and the sought eigenvector is identical to the 9th column of V computed by the SVD of M.

Theorem 4.3.1 *Given $F \in \mathbb{R}^{3\times 3}$ with a SVD equal to $F = U\text{diag}(\lambda_1, \lambda_2, \lambda_3)V^\top$, with $U, V \in SO(3)$ and $\lambda_1 \geq \lambda_2 \geq \lambda_3$, then the essential matrix E that minimises $\|E - f\|_F^2$ is given by $E = U\text{diag}(\sigma, \sigma, 0)V^\top$ with $\sigma = (\lambda_1 + \lambda_2)/2$.*

Proof. By the Eckart-Young-Mirsky Theorem it holds

$$\|E - f\|_F^2 = (\lambda_1 - \sigma)^2 + (\lambda_2 - \sigma)^2 + (\lambda_3 - 0)^2 \tag{4.82}$$

The minimiser is given by

$$\frac{\partial}{\partial\sigma}\|E - f\|_F^2 \overset{!}{=} 0 \tag{4.83}$$

We make this condition precise computing the derivative using (4.82):

$$
\begin{aligned}
2(\lambda_1 - \sigma)(-1) + 2(\lambda_2 - \sigma)(-1) &= 0\\
\Leftrightarrow \quad \lambda_1 + \lambda_2 - 2\sigma &= 0\\
\Leftrightarrow \quad \sigma &= \frac{\lambda_1 + \lambda_2}{2} &(4.84)
\end{aligned}
$$

■

The above operation is often called *projection onto the essential space.*

As we can observe by the epipolar constraint

$$\langle \vec{x}_2, E\vec{x}_1 \rangle = 0 \tag{4.85}$$

the essential matirx E is defined only up to a nonzero scalar factor.

One possible way to deal with this problem is to normalise E, setting $\|E\| = 1$. Alternatively, instead of computing σ as recommended in Theorem 4.3.1 one may set that diagonal matrix to $\mathrm{diag}(1, 1, 0)$.

Note, that by $E = \hat{T}R$ where $R \in SO(3)$, a normalisation $\|E\| = 1$ is equivalent to assuming a unit translation $\|T\| = 1$.

Let us now comment on the use of the SVD in determining E. We have used for both E and F in the above theorem aSVD of structure $U\mathrm{diag}(.)V^\top$ with U, V in $SO(3)$. However, a standard SVD algorithm will use orthogonal matrices U and V with $\det = \pm 1$, but not U, V in $SO(3)$. On the other hand, $\|E\| = 1$ holds for both $\pm E$.

There is actually no problem: It is easy to see that one of the matrices $\pm E$ has has a SVD with U, V in $SO(3)$. We just do not know in advance so that both $\pm E$ need to be checked.

We now describe and comment on the *Algorithm of Longuet and Higgins* from 1981.

Algorithm 4.3.1 *For a given set of corespondences* $(\vec{x}_1^i, \vec{x}_2^i)$, $i = 1, \ldots, k$ *and* $k \geq 8$, *this algorithm computes* $(R, t) \in SE(3)$ *satisfying for all pairs* $(\vec{x}_1^i, \vec{x}_2^i)$ *the epipolar constraint in a LS sense.*

Step 1: Compute first approximation of E

Compute the SVD of M *as* $M = U\Sigma V^\top$ *and find* F^s *by taking the 9th column of* V, *then normalise* $\|F^s\| = 1$.

Step 2: Projection onto essential space

Compute SVD of F *and determine* E *as by Theorem 4.3.1.*

Step 3: Compute four solutions (R, T)

This is done by using the determined U *and* V, *and* $\pm E$ *in (4.43).*

The question remains, how one can distinguish the true solution from the other three.

The answer is to consider the *positive depth constraint.* The equation

$$\lambda_2 \vec{x}_2 = \lambda_1 R \vec{x}_1 + T \tag{4.86}$$

will yield for three out of four solutions at least one $\lambda_i < 0$. The corresponding pairs (R, T) can be discarded.

4.4 Structure Reconstruction

The Longuet-Higgins (LH) algorithm returns the relative pose (R, T) between two cameras up to an arbitrary scale $\gamma \in \mathbb{R}_+$. This plus the requirement $\gamma > 0$ can easily be seen by considering that we identified the correct sign of E, and by the normalisation step $\|F^s\| = 1$.

Employing $\|E\| = 1$, this implies a translation of unit length, since $E = \hat{T}R$ and $R \in SO(3)$ does not change the norm of a vector: $\|R\vec{x}\| = \|\vec{x}\|$.

Let us consider the basic rigid-body transform

$$\lambda_2^i \vec{x}_2^i \;=\; \lambda_1^i R \vec{x}_1^i + \gamma T\,, \quad i = 1, \ldots, k \tag{4.87}$$

Since (R, T) are known, the k equations in (4.87) are linear in λ_1^i, λ_2^i and the scale γ.

Let us refine the set of constituting equations (4.87).

For each point $\vec{x}_j$, λ_j is the *depth* of the 3-D object. Since $\lambda_1 \vec{x}_1$ and $\lambda_2 \vec{x}_2$ denote the same 3-D point, one value λ is redundant. Thus, we eliminate λ_2 from (4.87) by multiplying with $\hat{x}_2^i$:

$$\lambda_1^i \hat{x}_2^i R \vec{x}_1^i + \gamma \hat{x}_2^i T \;=\; \vec{0}\,, \quad i = 1, \ldots, k \tag{4.88}$$

This is equivalent to solving for all indices i

$$M^i \tilde{\lambda}^i \;:=\; \left[\hat{x}_2^i R \vec{x}_1^i,\, \hat{x}_2^i T\right] \begin{pmatrix} \lambda_1^i \\ \gamma \end{pmatrix} \;=\; \vec{0} \tag{4.89}$$

where we have in detail

$$M^i \;=\; [\; \underbrace{\hat{x}_2^i R \vec{x}_1^i}_{=:\vec{m}_1^i \in \mathbb{R}^3},\;\; \underbrace{\hat{x}_2^i T}_{=:\vec{m}_2^i \in \mathbb{R}^3} \;] \in \mathbb{R}^{3\times 2} \tag{4.90}$$

and

$$\tilde{\lambda}^i \;=\; \begin{pmatrix} \lambda_1^i \\ \gamma \end{pmatrix} \in \mathbb{R}^2 \tag{4.91}$$

In order to have up to a scaling a unique solution of (4.89), M^i needs to have a rank of 1.

Concerning the latter requirement, let us note that, as

$$\vec{x}_2^i \;\|\; R\vec{x}_1^i \tag{4.92}$$

because it holds $\vec{x}_2^i \sim R\vec{x}_1^i$ up to a translation T by (4.87), we have

$$\hat{x}_2^i R \vec{x}_1^i \;=\; \vec{0} \tag{4.93}$$

so that $\vec{m}_1^i = \vec{0}$. Then M^i is of rank 1 if and only if $\vec{m}_2^i \neq \vec{0}$. The case

$$\vec{m}_2^i \;=\; \hat{x}_2^i T \;=\; \vec{0} \tag{4.94}$$

happens when the point p_i (with image coordinates $\vec{x}_1^i$, $\vec{x}_2^i$) is on the base line connecting $\vec{o}_1$ and $\vec{o}_2$.

Let us note that all the n equations (4.89) share the same γ.

We summarise our equations by defining

$$\vec{\lambda} \;:=\; \left(\lambda_1^1, \lambda_1^2, \ldots, \lambda_1^k, \gamma\right)^\top \tag{4.95}$$

and a matrix

$$M \;:=\; \begin{pmatrix} \hat{x}_2^1 R \vec{x}_1^1 & 0 & & \cdots & 0 & \hat{x}_2^1 T \\ 0 & \hat{x}_2^2 R \vec{x}_1^2 & 0 & & 0 & \hat{x}_2^2 T \\ & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & & & & 0 & \vdots \\ 0 & & \cdots & 0 & \hat{x}_2^k R \vec{x}_1^k & \hat{x}_2^k T \end{pmatrix} \in \mathbb{R}^{(3n)\times(n+1)} \tag{4.96}$$

Then the equation

$$M\vec{\lambda} \;:=\; \vec{0} \tag{4.97}$$

determines all unknown depths up to a single universal scale.

This *scale ambiguity* is natural, as one cannot distinguish a setting with a reference translation T from another one, where ones has a combination of a translation T' twice as large as T and a scene twice as large but two times further away.

The LS solution $\vec{\lambda}$ of (4.97) is the eigenvector of $M^\top M$ corresponding to the smallest eigenvalue, compare our previous proceedings.

4.5 Optimal Pose Recovery

The LH algorithm as well as the developed theory assumes that *exact* point correspondences, image coordinates etc. are given.

There are some problems:

- We performed a projection of the computed matrix F onto essential space. This is just an approximation.
- The estimated pose may not be close to the true solution.
- For image coordinates $\vec{x}_1$, $\vec{x}_2$ taken from a discrete pixel grid, the corresponding lines do not meet in some point p.

We model these issues as follows. The image coordinates $\vec{x}_1$, $\vec{x}_2$ that are of the format $(x, y, 1)^\top$ are *idealisations* that are given in practice via *perturbed* coordinates

$$\tilde{x}_1^i \; = \; \vec{x}_1^i + \delta\vec{x}_1^i \, , \quad \tilde{x}_2^i \; = \; \vec{x}_2^i + \delta\vec{x}_2^i \tag{4.98}$$

where

$$\delta\vec{x}_j^i \; = \; \left(\delta\vec{x}_{j,1}^i, \delta\vec{x}_{j,2}^i\right)^\top \tag{4.99}$$

The measured coordinates $(\tilde{x}_1^i, \tilde{x}_2^i)$ do in general not satisfy the epipolar constraints, i.e.

$$\langle \tilde{x}_2^i, \, \hat{T}R\tilde{x}_1^i \rangle \; \neq \; 0 \tag{4.100}$$

The aim is now to minimise the difference between the *ideal model* $\vec{x}$ and the data $\tilde{x}$ in terms of parameters $(\vec{x}, T, R)$.

There is an artistic freedom in the means how to measure that difference and how to realise the minimisation.

Let us simply assume that $\delta\vec{x}_j^i$ are unknown errors and that we aim to minimise them in a LS sense, via minimising the *objective function*

$$\phi(\vec{x}, T, R) \; = \; \sum_{i=1}^{k} \left(\|\delta v_1^i\|^2 + \|\delta v_2^i\|^2\right) \; := \; \sum_{i=1}^{k} \left(\|\tilde{x}_1^i - \vec{x}_1^i\|^2 + \|\tilde{x}_2^i - \vec{x}_2^i\|^2\right) \tag{4.101}$$

Since $(\vec{x}_1^i, \vec{x}_2^i)$ are the recovered 3-D points p_i projected back onto the image planes, minimising (4.101) means to minimise the *reprojection error*.

The above problem formulation is not completely concise, since the sought coordinates $(\vec{x}_1^i, \vec{x}_2^i)$ shall satisfy the epipolar constraints

$$\langle \vec{x}_2^i, \, \hat{T}R\vec{x}_1^i \rangle \; = \; 0 \tag{4.102}$$

Therefore, the complete minimisation task is a *constrained optimisation problem*:

$$\begin{cases} \text{min! } \phi(\vec{x}, T, R) = \sum_{i=1}^{k} \sum_{j=1}^{2} \|\tilde{x}_j^i - \vec{x}_j^i\|^2 \\ \text{subject, for } i = 1, \ldots, k, \text{ to} \\ \langle \vec{x}_2^i, \, \hat{T}R\vec{x}_1^i \rangle = 0 \\ \langle \vec{x}_1^i, e_3 \rangle = 1 \\ \langle \vec{x}_2^i, e_3 \rangle = 1 \end{cases} \tag{4.103}$$

where $e_3 = (0, 0, 1)^\top$.

If (R, T) are considered to be estimated in an optimal way, we can find a pair $(\vec{x}_1^*, \vec{x}_2^*)$ satisfying $\langle \vec{x}_2^*, \hat{T}R\vec{x}_1^* \rangle = 0$ and minimising the reprojection error

$$\phi(\vec{x}) = \|\tilde{x}_1 - \vec{x}_1\|^2 + \|\tilde{x}_2 - \vec{x}_2\|^2 \tag{4.104}$$

This is called the *triangulation problem*.

The value of ϕ depends in the latter only on the position of the epipolar plane, which may rotate around the baseline $(\vec{o}_1, \vec{o}_2)$. To parameterise the position of the epipolar plane P, let $(\vec{e}_2, \vec{n}_1, \vec{n}_2)$ (with $\vec{e}_2$ being the epipole) be an orthonormal basis in the second camera frame. Then P is determined by its normal vector $\vec{l}_2$, which is in turn determined by the angle θ between $\vec{l}_2$ and $\vec{n}_1$. The minimisation can be carried out with respect to θ.

Chapter 5

Optimisation for Pose and Structure

The objective of this chapter is to discuss in some detail how to approach the optimisation of the reprojection error under epipolar constraints.

Corresponding techniques can be applied in various ways in computer vision.

5.1 Lagrange Multiplier

We consider the optimisation problem subject to equality constraints

$$\begin{cases} x^* = \min f(x) \\ \text{subject to } h(x) = 0 \end{cases} \tag{5.1}$$

where, in general, $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ and

$$h : \mathbb{R}^n \to \mathbb{R}^m, \quad x \mapsto (h_1, \ldots, h_m)^\top(x) \tag{5.2}$$

For each constraint $h_j(x) = 0$ to be effective at the minimiser x^*, one often assumes that the gradients

$$\nabla h_1(x^*), \ldots, \nabla h_m(x^*) \in \mathbb{R}^n \tag{5.3}$$

are linearly independent. Such constraints are called *regular*.

The basic idea that we will follow now is to build a new objective function that includes the constraints in such a way, that the minimisation of the new function solves the problem (5.1).

To this end, let us define the *Lagrangian function* $L : \mathbb{R}^{n+m} \to \mathbb{R}$ as

$$L(x, \lambda) \;:=\; f(x) + \lambda^\top h(x) \tag{5.4}$$

where we have introduced the vector of *Lagrange multipliers*

$$\lambda := (\lambda_1, \ldots, \lambda_m)^\top \in \mathbb{R}^m \tag{5.5}$$

The *necessary conditions* for a minimum of the function L are that the first-order derivatives are zero, i.e.

$$\frac{\partial L}{\partial x_i} = 0 \quad \text{and} \quad \frac{\partial L}{\partial \lambda_j} = 0 \tag{5.6}$$

Especially, carrying out the differentiation w.r.t. the Lagrange multipliers, we obtain

$$\frac{\partial L}{\partial \lambda_j} = \frac{\partial}{\partial \lambda_j}\left(f + \lambda^\top h\right) = h_j = 0 \tag{5.7}$$

i.e. the minimiser (x^*, λ^*) of L satisfies necessarily the equality constraints of (5.1).

For *sufficient conditions* to have a minimum (x^*, λ^*), the *Hessian matrix* $\mathcal{H}(L)$ that corresponds to taking second-order derivatives should be positive semi-definite, i.e.

$$v^\top \mathcal{H}(L) v \geq 0 \quad \forall v \neq 0 \tag{5.8}$$

Since

$$\frac{\partial^2 L}{\partial \lambda_k \partial \lambda_l} = 0 \quad \forall k, l \tag{5.9}$$

it suffices to consider the parts corresponding to *two cases*: *(i)* second-order derivatives of x-variables and *(ii)* a mixture of first-order derivatives in x and λ.

Let us take care of the *first case*. Writing

$$\frac{\partial^2 L}{\partial x^2} := \left(\frac{\partial^2 L(x^*, \lambda^*)}{\partial x_i \partial x_k}\right)_{i,k=1,\ldots,n} \tag{5.10}$$

we obtain by (5.8) the condition

$$v^\top \frac{\partial^2 L}{\partial x^2} v \geq 0 \quad \forall v \in \mathbb{R}^n,\ v \neq 0 \tag{5.11}$$

We turn to the *second case*. Carrying out differentiation w.r.t. λ_j we obtain by (5.7) $h_j(x) = 0$. Another differentiation w.r.t. the i-th component of x yields

$$\frac{\partial^2 L(x^*, \lambda^*)}{\partial x_i \partial \lambda_j} = \frac{\partial}{\partial x_i} h_j(x) \tag{5.12}$$

Taking into account $x \in \mathbb{R}^n$ we will consider the *gradient*, in detail

$$\nabla h_j(x), \quad \nabla := \left(\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\right)^\top \tag{5.13}$$

While the constraints $h(x) = 0$ are enforced by the necessary conditions, our problem formulation bears no implication on $\nabla h_j(x)$ for any j.

As we remember that we originally intended to look for a minimum just w.r.t. the variables x, we solve this problem by requiring

$$\langle \nabla h_j(x), v \rangle \ = \ 0 \quad \text{for } j = 1, \ldots, m \tag{5.14}$$

for the vectors v.

Thus, we summarise the sufficient conditions as

$$v^\top \frac{\partial^2 L}{\partial x^2} v \geq 0 \quad \forall v \ : \ \underbrace{v^\top}_{\in \mathbb{R}^{1\times n}} \underbrace{[\nabla h_1, \ldots, \nabla h_m]}_{\in \mathbb{R}^{n\times m}} = \underbrace{0}_{\in \mathbb{R}^{1\times m}} \tag{5.15}$$

The *Theorem of Lagrange* combines these developments, saying that the necessary and sufficient conditions (5.6) and (5.15) give the solution of (5.1).

A *possible algorithmical approach* to obtain candidates for local minima of problem (5.1) is to solve the necessary optimality conditions (5.6). They give a system of $n + m$ equations in the $n + m$ unknowns (x, λ). However, the system will only be regular if the constraints $h(x) = 0$ are regular: In a local minimum it holds ideally $\nabla f(x) = 0$ so that the ∇h_j need to be linearly independent in order to obtain a unique solution.

5.2 Optimisation Subject to Epipolar Constraints

We recall the constrained optimisation problem (4.103) of minimising the reprojection error under epipolar constraints:

$$\begin{cases} \text{min! } \phi(\vec{x}, T, R) = \sum_{i=1}^{k} \sum_{j=1}^{2} \|\vec{x}_k^i - \tilde{x}_k^i\|^2 \\ \text{subject, for } i = 1, \ldots, k, \text{ to} \\ \langle \vec{x}_2^i, \hat{T} R \vec{x}_1^i \rangle = 0 \\ \langle \vec{x}_1^i, e_3 \rangle = 1 \\ \langle \vec{x}_2^i, e_3 \rangle = 1 \end{cases} \tag{5.16}$$

where $e_3 = (0, 0, 1)^\top$ and where $\tilde{x}_k^i$ are measured data.

Using the technique of Lagrange multipliers, we write down the associated Lagrangian function

$$\begin{aligned} & L(\vec{x}, R, T, \lambda, \gamma, \eta) \qquad\qquad (5.17) \\ := \ & \sum_{j=1}^{n} \left[\|\vec{x}_1^j - \tilde{x}_1^j\|^2 + \|\vec{x}_2^j - \tilde{x}_2^j\|^2 + \lambda^j \langle \vec{x}_2^j, \hat{T} R x_1^j \rangle + \gamma^j \left(\langle \vec{x}_1^j, e_3 \rangle - 1 \right) + \eta^j \left(\langle \vec{x}_2^j, e_3 \rangle - 1 \right) \right] \end{aligned}$$

The necessary condition for having a minimiser is $\nabla L = 0$, where in our case (5.17) the gradient is taken w.r.t. the variables in $\vec{x}_1^j$, $\vec{x}_2^j$, λ^j, γ^j and η^j.

As discussed previously, setting derivatives w.r.t. the Lagrange multipliers λ^j, γ^j and η^j to zero gives back the epipolar constraints.

Differentiation of (5.17) w.r.t. $\vec{x}_1^j$ (in the sense of individual entries $\vec{x}_{1,k}^j$) gives

$$\frac{\partial}{\partial \vec{x}_{1,k}^j} L(\vec{x}, R, T, \lambda, \gamma, \eta) \;=\; 2\left(\vec{x}_1^j - \tilde{x}_1^j\right) + \sum_{l=1}^{n} \lambda^j \vec{x}_{2,k}^j \left(\hat{T}R\right)_{l,k} + \gamma^j e_{3,k} \qquad (5.18)$$

Combing the expressions for the k entries in (5.18) and setting the derivatives to zero gives

$$2\left(\vec{x}_1^j - \tilde{x}_1^j\right) + \lambda^j \left(\hat{T}R\right)^\top \vec{x}_2^j + \gamma^j e_3 \;=\; \vec{0} \qquad (5.19)$$

Analogously we obtain by differentiation w.r.t. $\vec{x}_2^j$

$$2\left(\vec{x}_2^j - \tilde{x}_2^j\right) + \lambda^j \hat{T}R\vec{x}_1^j + \eta^j e_3 \;=\; \vec{0} \qquad (5.20)$$

Let us focus on (5.19). Multiplication with $\hat{e}_3$ gives

$$2\hat{e}_3\left(\vec{x}_1^j - \tilde{x}_1^j\right) + \lambda^j \hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j \;=\; \vec{0} \qquad (5.21)$$

or equivalently

$$\hat{e}_3 \vec{x}_1^j \;=\; \hat{e}_3 \tilde{x}_1^j - \frac{1}{2}\lambda^j \hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j \qquad (5.22)$$

Another multiplication by $\hat{e}_3^\top$, where

$$\hat{e}_3^\top \hat{e}_3 \;=\; \mathrm{diag}(1,1,0) \qquad (5.23)$$

gives

$$\vec{x}_1^j \;=\; \tilde{x}_1^j - \frac{1}{2}\lambda^j \hat{e}_3^\top \hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j \qquad (5.24)$$

Note that for the third component of the x-vectors, there is a 1 in homogeneous coordinates. We can write down (5.24) in the current, simple form since it is not important to distinguish in the equality $0 = 0$ or $1 = 1$ – as by multiplication with $\hat{e}_3^\top \hat{e}_3$ – in the third entry.

Analogously to the above derivation of (5.24) we obtain from (5.20) the equation

$$\vec{x}_2^j \;=\; \tilde{x}_2^j - \frac{1}{2}\lambda^j \hat{e}_3^\top \hat{e}_3 \hat{T}R\vec{x}_1^j \qquad (5.25)$$

In a next step we solve for the Lagrange multipliers λ^j.

Starting from (5.24) we want to obtain scalar values λ^j from this equality of vectors. To this end we again make use of the epipolar constraint we impose on our ideal solution:

$$
\begin{aligned}
\lambda^j \hat{e}_3^\top \hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j &= 2\left(\tilde{x}_1^j - \vec{x}_1^j\right) \\
\Leftrightarrow \quad \lambda^j \left(\vec{x}_2^j\right)^\top \hat{T} R \hat{e}_3^\top \hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j &= 2\left(\vec{x}_2^j\right)^\top \hat{T} R \tilde{x}_1^j - 2\underbrace{\left(\vec{x}_2^j\right)^\top \hat{T} R \vec{x}_1^j}_{=0} \\
\Leftrightarrow \quad \lambda^j \langle \hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j, \hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j \rangle &= 2\left(\vec{x}_2^j\right)^\top \hat{T} R \tilde{x}_1^j \\
\Leftrightarrow \quad \lambda^j &= \frac{2\left(\vec{x}_2^j\right)^\top \hat{T} R \tilde{x}_1^j}{\|\hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j\|} \qquad (5.26)
\end{aligned}
$$

Analogously we compute by (5.25):

$$
\begin{aligned}
\lambda^j \hat{e}_3^\top \hat{e}_3 \hat{T} R \vec{x}_1^j &= 2\left(\tilde{x}_2^j - \vec{x}_2^j\right) \\
\Leftrightarrow \quad \lambda^j \left(\vec{x}_1^j\right)^\top R^\top \hat{T}^\top \hat{e}_3^\top \hat{e}_3 &= 2\left(\tilde{x}_2^j\right)^\top - 2\left(\vec{x}_2^j\right)^\top \\
\Leftrightarrow \quad \lambda^j \left(\vec{x}_1^j\right)^\top R^\top \hat{T}^\top \hat{e}_3^\top \hat{e}_3 \hat{T} R \vec{x}_1^j &= 2\left(\tilde{x}_2^j\right)^\top \hat{T} R \vec{x}_1^j - 2\underbrace{\left(\vec{x}_2^j\right)^\top \hat{T} R \vec{x}_1^j}_{=0} \\
\Leftrightarrow \quad \lambda^j &= \frac{2\left(\tilde{x}_2^j\right)^\top \hat{T} R \vec{x}_1^j}{\|\hat{e}_3 \hat{T} R \vec{x}_1^j\|} \qquad (5.27)
\end{aligned}
$$

We may now plug the expressions we found into our objective function from (5.16).

Let us focus on the contribution $\|\vec{x}_1^j - \tilde{x}_1^j\|^2$ in ϕ:

$$
\begin{aligned}
&\|\vec{x}_1^j - \tilde{x}_1^j\|^2 \\
&\overset{(5.24)}{=} \quad \|\frac{1}{2}\lambda^j \hat{e}_3^\top \hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j\|^2 \\
&\overset{(5.26)}{=} \quad \|\underbrace{\frac{\left(\vec{x}_2^j\right)^\top \hat{T} R \tilde{x}_1^j}{\|\hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j\|}}_{=:\alpha} \hat{e}_3^\top \hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j\|^2 \\
&= \quad \langle \alpha \underbrace{\hat{e}_3^\top \hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j}_{=:\vec{a}}, \alpha \hat{e}_3^\top \hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j, \rangle \qquad (5.28)
\end{aligned}
$$

Since α is a scalar number , we may withdraw it from the latter inner product:

$$
\langle \alpha \vec{a}, \alpha \vec{a} \rangle = \alpha^2 \langle \vec{a}, \vec{a} \rangle \qquad (5.29)
$$

Let us consider in some more detail the expression $\langle \vec{a}, \vec{a} \rangle$:

$$\begin{aligned}
\langle \vec{a}, \vec{a} \rangle &= \langle \hat{e}_3^\top \hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j, \hat{e}_3^\top \hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j \rangle \\
&= \left(\vec{x}_2^j\right)^\top \hat{T} R \underbrace{\hat{e}_3^\top \hat{e}_3}_{=\mathrm{diag}(1,1,0)} \underbrace{\hat{e}_3^\top \hat{e}_3}_{=\mathrm{diag}(1,1,0)} R^\top \hat{T}^\top \vec{x}_2^j \\
&= \left(\vec{x}_2^j\right)^\top \hat{T} R \hat{e}_3^\top \hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j \\
&= \langle \hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j, \hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j \rangle \\
&= \| \hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j \|^2 \qquad (5.30)
\end{aligned}$$

With our definitions, we thus obtain

$$\|\vec{x}_1^j - \tilde{x}_1^j\|^2 = \alpha^2 \langle \vec{a}, \vec{a} \rangle = \frac{\left(\left(\vec{x}_2^j\right)^\top \hat{T} R \tilde{x}_1^j\right)^2}{\|\hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j\|^4} \|\hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j\|^2 = \frac{\left(\left(\vec{x}_2^j\right)^\top \hat{T} R \tilde{x}_1^j\right)^2}{\|\hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j\|^2} \qquad (5.31)$$

Analogously, we obtain

$$\|\vec{x}_2 j - \tilde{x}_2^j\|^2 = \frac{\left(\left(\tilde{x}_2^j\right)^\top \hat{T} R \vec{x}_1^j\right)^2}{\|\hat{e}_3 \hat{T} R \vec{x}_1^j\|^2} \qquad (5.32)$$

so that in total

$$\phi(\vec{x}, R, T) = \sum_{j=1}^n \left[\frac{\left(\left(\vec{x}_2^j\right)^\top \hat{T} R \tilde{x}_1^j\right)^2}{\|\hat{e}_3 R^\top \hat{T}^\top \vec{x}_2^j\|^2} + \frac{\left(\left(\tilde{x}_2^j\right)^\top \hat{T} R \vec{x}_1^j\right)^2}{\|\hat{e}_3 \hat{T} R \vec{x}_1^j\|^2} \right] \qquad (5.33)$$

Let us stress, that the latter expression incorporates necessary optimality conditions, andcan finally be minimised.

One may simplify the derived objective function by substituting $\tilde{x}$ for unknowns $\vec{x}$. This leads to

$$\phi(R, T) = \sum_{j=1}^n \left[\frac{\left(\left(\tilde{x}_2^j\right)^\top \hat{T} R \tilde{x}_1^j\right)^2}{\|\hat{e}_3 R^\top \hat{T}^\top \tilde{x}_2^j\|^2} + \frac{\left(\left(\tilde{x}_2^j\right)^\top \hat{T} R \tilde{x}_1^j\right)^2}{\|\hat{e}_3 \hat{T} R \tilde{x}_1^j\|^2} \right] \qquad (5.34)$$

which is to be minimised only w.r.t. the pose (R, T).

As it turns out, one may interprete the terms in (5.33) as distances from the image points $\tilde{x}_k^j$ to corresponding epipolar lines.

The whole developments lead to an algorithm that we may sketch as follows:

Algorithm 5.2.1 (for Optimising Pose and Structure) *We describe a simple algorithm for alternating minimisation:*

Step 1 Initialisation:

- *Initialise* $\vec{x}_1$ *and* $\vec{x}_2$ *as* $\tilde{x}_1$ *and* $\tilde{x}_2$*, respectively.*
- *Initialise* (R, T) *by the result of the Longuet-Higgings-Algorithm.*

Step 2 Pose Estimation: *For* $\vec{x}_1$ *and* $\vec{x}_2$ *computed from the previous step, update* (R, T) *by minimising the reprojection error* $\phi(\vec{x}, R, T)$*.*

Step 3 Structure Triangulation: *For each image pair* $(\tilde{x}_1, \tilde{x}_2)$ *and* (R, T) *computed in Step 2, solve for* $(\vec{x}_1, \vec{x}_2)$ *that minimise the reprojection error* $\phi(\vec{x}) = \|\vec{x}_1 - \tilde{x}_1\|^2 + \|\vec{x}_2 - \tilde{x}_2\|^2$.

Step 4 Return to Step 2 until the decrement in the value of ϕ *is below a threshold.*

The algorihtm above is conceptually identical to the so-called *bundle adjustment*: This denotes the direct optimisation of the reprojection error w.r.t. all the unknowns for several input images.

Chapter 6

Planar Homography

We now address the situation that the input data for the Longuet-Higgins-Algorithm is not good in the sense, that the solution is not unique.

As it happens, this situation may occur if all measured points lie on certain 2-D surfaces called *critical surfaces*. As 2-D planes that fall into this category are quite common in man-made environments, we now discuss the correspondong *planar* setting.

6.1 What is a Planar Homography?

Let us consider two images $\vec{x}_1$, $\vec{x}_2$ of points p on a 2-D plane P in 3-D space. We now investigate the availbale relations between such pairs.

Writing the corresponding coordinate transform we have

$$\vec{X}_2 \;=\; R\vec{X}_1 + T \tag{6.1}$$

with $\vec{X}_1$, $\vec{X}_2$ being the coordinates of p.

The two images $\vec{x}_1$, $\vec{x}_2$ of p must satisfy the epipolar constraint

$$\vec{x}_2^\top E\vec{x}_1 \;=\; \vec{x}_2^\top \hat{T}R\vec{x}_1 \;=\; 0 \tag{6.2}$$

If points are given on a common plane P, this will induce another constraint, beyond the standard relations (6.1), (6.2).

Let $\vec{N} := (n_1, n_2, n_3)^\top$ be the unit normal vector of the plane P, and let d be the distance from the plane P to the optical center of the first camera.

Since $\vec{X}_1$ denotes the coordinates of some point on P, and as $\vec{N}$ and $\vec{X}_1$ are both given in terms of the first camera frame, it holds

$$\langle \vec{N}, \vec{X}_1 \rangle \;=\; n_1X_1 + n_2Y_1 + n_3Z_1 \;=\; d \tag{6.3}$$

i.e.

$$\frac{1}{d}\langle \vec{N}, \vec{X}_1 \rangle \;=\; 1 \quad \forall\, \vec{X}_1 \in P \tag{6.4}$$

Making use of the latter in (6.1) gives

$$\vec{X}_2 \;=\; R\vec{X}_1 + T \cdot 1 \;=\; R\vec{X}_1 + T \cdot \frac{1}{d}\langle \vec{N}, \vec{X}_1 \rangle \;=\; \left(R + \frac{1}{d} T \vec{N}^\top \right) \vec{X}_1 \tag{6.5}$$

The matrix

$$H \;:=\; R + \frac{1}{d} T \vec{N}^\top \quad \in \mathbb{R}^{3\times 3} \tag{6.6}$$

is called the *planar homography matrix*.

Let us stress that H depends on the pose (R, T) as well as on the parameters d and $\vec{N}$ describing the plane P.

From

$$\lambda_1 \vec{x}_1 = \vec{X}_1, \quad \lambda_2 \vec{x}_2 = \vec{X}_2, \quad \vec{X}_2 = H\vec{X}_1 \tag{6.7}$$

we have

$$\lambda_2 \vec{x}_2 = H \lambda_1 \vec{x}_1 \quad \Rightarrow \quad \vec{x}_2 \sim H\vec{x}_1 \tag{6.8}$$

which highlights that there is an inherent scale ambiguity. This arises due to the open scaling parameter in $\frac{1}{d}T$.

Often the equation

$$\vec{x}_2 \sim H\vec{x}_1 \tag{6.9}$$

is called *planar homography mapping* induced by the plane P.

H introduces a mapping between points in the first and the second image in the sense described below.

Proposition 6.1.1 *Let p be a point on the plane P. Consider the points $\vec{x}_1$ in the first image and $\vec{x}_2$ in the second image that correspond to p. The $\vec{x}_2$ is uniquely determined as $\vec{x}_2 \sim H\vec{x}_1$: For any other point $\vec{x}_2'$ on the same epipolar line $\vec{l}_2 \sim E\vec{x}_1$, the ray $\lambda_1 \vec{x}_1$ at a point p′ out of the plane.*

Proposition 6.1.2 *If $\vec{x}_1$ is the image of some point p′ not on P, then $\vec{x}_2 \sim H\vec{x}_1$ is only a point that is on the same epipolar line $\vec{l}_2 \sim E\vec{x}_1$ as its actual corresponding image $\vec{x}_2'$. That implies $\vec{l}_2^\top \vec{x}_2 = \vec{l}_2^\top \vec{x}_2' = 0$.*

We combine these assertions as:

Theorem 6.1.1 *Given a homography H induced by plane P in 3-D between two images. For any pair of corresponding images $(\vec{x}_1, \vec{x}_2)$ of a 3-D point p that is not necessarily on P, the associated epipolar lines are*

$$\vec{l}_2 \sim \hat{x}_2 H \vec{x}_1 \tag{6.10}$$

and

$$\vec{l}_1 \sim H^\top \vec{l}_2 \tag{6.11}$$

Proof. We will only show (6.10). If p is not on P, the equation (6.10) is true by Proposition 6.1.2.

For points on the plane P, $\vec{x}_2 = H\vec{x}_1$ implies $\hat{x}_2 H \vec{x}_1 = 0$. This means, the equation (6.10) is still true.

The above properties of H allow to compute epipolar lines without knowing the essential matrix E. However, for a given image point we cannot find a corresponding point, but only a corresponding epipolar line.

As a remark, H can be computed from a small number of correspondences, and in turn knowing it helps in establishing correspondences.

Chapter 7

Estimating the Planar Homography Matrix

In order to eliminate the unknown scale in

$$\vec{x}_2 \sim H\vec{x}_1 \tag{7.1}$$

we multiply both sides by $\hat{x}_2$ to obtain

$$\hat{x}_2 H\vec{x}_! \;=\; 0 \tag{7.2}$$

This equation is called *planar epipolar constraint*, or *planar homography constraint*.

Let us point at an issue we observe by the latter constraint. Since for any vector $u \in \mathbb{R}^3$ we have that

$$\hat{u}\vec{x}_2 \;=\; \vec{u} \times \vec{x}_2 \tag{7.3}$$

is orthogonal to $\vec{x}_2$. By (7.1) there is also

$$\hat{u}H\vec{x}_1 \;=\; \vec{u} \times (H\vec{x}_1) \;\sim\; \vec{u} \times \vec{x}_2 \tag{7.4}$$

so that

$$\hat{u}H\vec{x}_1 \perp \vec{x}_2 \tag{7.5}$$

This implies

$$\vec{x}_2^\top \hat{u}H\vec{x}_1 \;=\; 0 \quad \forall u \in \mathbb{R}^3 \tag{7.6}$$

That means, $\vec{x}_2^\top E\vec{x}_1 = 0$ for a family of matrices $E = \hat{u}H$ besides the essential matrix $E = \hat{T}R$. This explains why the Longuet-Higgins-Algorithm does not apply to feature points from a planar scene.

7.1 Homography from Rotation

The homography relation $\vec{x}_2 \sim H\vec{x}_1$ also shows up when the camera is rotating only, i.e. for $\vec{X}_2 = R\vec{X}_1$. In this case the homography matrix becomes

$$H = R \tag{7.7}$$

since $T = \vec{0}$. Consequently, the constraint

$$\hat{x}_2 R\vec{x}_1 = 0 \tag{7.8}$$

arises.

One can interprete this example as follows. By the underlying relation

$$H = R + \frac{1}{d}T\vec{N}^\top \tag{7.9}$$

it must hold

$$\frac{1}{d}T\vec{N}^\top \to 0 \tag{7.10}$$

in order to obtain $H = R$. This is achieved by

$$\lim_{d\to\infty}\left[R + \frac{1}{d}T\vec{N}^\top\right] = R \tag{7.11}$$

Since d denotes the distance of the plane P to the origin of the first camera frame, one may interprete it as a special planar scene case where all points lie on a plane infinitely far away. Thus, without translation, information about the depth of the sceneis completely lost in the images.

7.2 Computation of the Homography Matrix

We now briefly sketch how to compute the matrix H. The procedure is analogously to the Longuet-Higgins-Algorithm.

Proposition 7.2.1 *A set of four point correspondences suffices to compute H up to a scale factor.*

The proof relies on the same construction as in the Longuet-Higgins-Algorithm that needs eight points. We here need only four since by $\vec{x}_2 \sim H\vec{x}_1$ each point correspondence gives two constraints.

Since there is a scalar factor λ involved, one has to take care of it using the structure of H:

Lemma 7.2.1 *[Normalisation of the Planar Homography] For a matrix $H_L := \lambda(R + \frac{1}{d}T\vec{N}^\top)$, we have*

$$|\lambda| = \sigma_2(H_L) \tag{7.12}$$

where σ_2 is the second largest singular value of H_L.

Let us give the key idea of the proof.

Setting $\vec{u} := \frac{1}{d}R^\top T \in \mathbb{R}^3$, then

$$\begin{aligned} H_L^\top H_L &= \left(\lambda(R + \frac{1}{d}T\vec{N}^\top)\right)^\top \left(\lambda(R + \frac{1}{d}T\vec{N}^\top)\right) \\ &= \lambda^2(R^\top + \frac{1}{d}NT^\top)(R + \frac{1}{d}T\vec{N}^\top) \\ &= \lambda^2\left(I + \vec{u}\vec{N}^\top + \vec{N}\vec{u}^\top + \|\vec{u}\|^2\vec{N}\vec{N}^\top\right) \end{aligned} \tag{7.13}$$

where we have used

$$\frac{1}{d}NT^\top\frac{1}{d}T\vec{N}^\top = \vec{N}\underbrace{\frac{1}{d}T^\top R}_{=\vec{u}^\top}\underbrace{\frac{1}{d}R^\top T}_{=\vec{u}}\vec{N}^\top = \|\vec{u}\|^2\vec{N}\vec{N}^\top \tag{7.14}$$

Let us observe that the vector

$$\vec{u} \times \vec{N} = \hat{u}\vec{N} \tag{7.15}$$

is by the properties of the cross product orthogonal to both $\vec{u}$ and $\vec{N}$. Thus from (7.13) we obtain

$$\begin{aligned} &H_L^\top H_L\left(\hat{u}\vec{N}\right) \\ &= \lambda^2\left(I + \vec{u}\vec{N}^\top + \vec{N}\vec{u}^\top + \|\vec{u}\|^2\vec{N}\vec{N}^\top\right)\hat{u}\vec{N} \\ &= \lambda^2\left(\hat{u}\vec{N}\right) + \vec{u}\underbrace{\vec{N}^\top\left(\hat{u}\vec{N}\right)}_{=0} + \vec{N}\underbrace{\vec{u}^\top\left(\hat{u}\vec{N}\right)}_{=0} + \|\vec{u}\|^2\vec{N}\underbrace{\vec{N}^\top\left(\hat{u}\vec{N}\right)}_{=0} \end{aligned} \tag{7.16}$$

This means, as

$$H_L^\top H_L\left(\hat{u}\vec{N}\right) = \lambda^2\left(\hat{u}\vec{N}\right) \tag{7.17}$$

the vector $\hat{u}\vec{N}$ is an eigenvector of $H_L^\top H_L$ and $|\lambda|$ is a singular value of H_L. It remains to show that it is the second largest singular value, which we skip here.

By the above lemma, we set

$$H := H_L/\sigma_2(H_L) \tag{7.18}$$

This recovers H up to the form

$$H = \pm\left(R + \frac{1}{d}T\vec{N}^{\top}\right) \tag{7.19}$$

To get the correct sign, we need to impose the positive depth constraint.

Similarly as with the essential matrix E, one can decompose the matrix H for estimation of the parameters R, $\frac{1}{d}T$ and $\vec{N}$.

Chapter 8

Reconstruction from Two Uncalibrated Views

We recall that the projection of a point with coordinates $\vec{X}$ onto the image plane has homogeneous coordinates $\vec{x}'$ satisfying

$$\lambda\vec{x}' = K\Pi g\vec{X} \tag{8.1}$$

where $\Pi = [I, \vec{0}] \in \mathbb{R}^{3\times4}$ and where $g \in SE(3)$ is the pose of the camera w.r.t. the world frame. The invertible calibration matrix

$$K = \begin{pmatrix} fs_x & fs_\theta & o_x \\ 0 & fs_y & o_y \\ 0 & 0 & 1 \end{pmatrix} \tag{8.2}$$

describes the intrinsic properties of the camera, such as the position of the optical center (o_x, o_y), the pixel size (s_x, s_y), the skew factor s_θ and the focal length f.

In what follows, we denote pixel coordinates with $\vec{x}'$ and metric coordinates by $\vec{x}$. These satisfy the relation

$$\vec{x}' = K\vec{x} \quad \Leftrightarrow \quad \vec{x} = K^{-1}\vec{x}' \tag{8.3}$$

Hence the knowledge of K is crucial for recovering the true 3-D Euclidean structure of a scene.

Unfortunately, the matrix K is in general not known which is the situation of *uncalibrated views*.

8.1 Geometric Interpretation of Uncalibrated Views

In standard Euclidean space, the canonical inner product is given as

$$\langle\vec{u}, \vec{v}\rangle = \vec{u}^\top\vec{v} \tag{8.4}$$

To understand the geometry associated with an uncalibrated camera, consider an invertible linear mapping ψ represented by our (unknown) matrix K that transforms spatial coordinates as

$$\psi \;:\; \mathbb{R}^3 \to \mathbb{R}^3, \quad \vec{X} \mapsto \vec{X}' = K\vec{X} \tag{8.5}$$

This mapping induces a transformation of the inner product:

$$\langle K^{-1}\vec{u}, K^{-1}\vec{v}\rangle \;=\; \vec{u}^\top K^{-\top} K^{-1}\vec{v} \;=\; \langle \vec{u}, \vec{v}\rangle_S \quad \forall \vec{u}, \vec{v} \in \mathbb{R}^3 \tag{8.6}$$

where the matrix

$$S \;:=\; K^{-\top}K^{-1} \tag{8.7}$$

is symmetric and positive definite.

Therefore, if one wants to write the inner product between two vectors, but only their pixel coordinates $\vec{u}, \vec{v}$ are available, one has to weigh the inner product as in (8.6).

The matrix S is called the *metric* of the space. The distortion of the 3-D Euclidean space induced by S alters both the length of vectors and the angles between them.

We have learned that rigid-body motions must preserve distances and angles, but these are now expresed via the metric S. Thus, how does a rigid-body motion look like in distorted space?

The Euclidean coordinates $\vec{X}$ of a moving point p at time t are given by

$$\vec{X} \;=\; R\vec{X}_0 + T \tag{8.8}$$

where $\vec{X}_0$ are the initial coordinates of p. This coordinate transform is then given in the uncalibrated camera coordinates $\vec{X}'$ by

$$K\vec{X} \;=\; KR\vec{X}_0 + KT \quad \Leftrightarrow \quad \vec{X}' \;=\; KRK^{-1}\vec{X}_0' + T' \tag{8.9}$$

where $\vec{X}' = K\vec{X}$ and $T' = KT$. Therefore, the mapping transforming $\vec{X}_0'$ to $\vec{X}'$ can be written in homogeneous coordinates as

$$G' \;=\; \left\{ g' := \begin{pmatrix} KRK^{-1} & T' \\ \vec{0}^\top & 1 \end{pmatrix} \;:\; T' \in \mathbb{R}^3, R \in SO(3) \right\} \tag{8.10}$$

These transformations form a matrix group called the *conjugate* of the Euclidean group $G = SE(3)$.

Applying the conjugate of G to the image formation model (8.1), we compute:

$$\begin{aligned}
\lambda\vec{x}' &= K\Pi g\vec{X} \qquad \text{(remember } K \text{ is unknown)} \\
&= K\Pi\left(R\vec{X}+T\right) \qquad \text{(again in homogeneous coordinates)} \\
&= K\left(R\vec{X}+T\right) \qquad \text{(after projection!)} \\
&= KR\vec{X}+KT \\
&= KRK^{-1}K\vec{X}+KT \\
&= KRK^{-1}\vec{X}'+T' \qquad \text{(with } \vec{X}' \in \mathbb{R}^3\text{)} \qquad (8.11)
\end{aligned}$$

Taking into account by (8.10) that

$$g'\begin{pmatrix} \vec{X}' \\ 1 \end{pmatrix} = \begin{pmatrix} KRK^{-1}\vec{X}'+T' \\ 1 \end{pmatrix} \qquad (8.12)$$

we obtain from (8.11)

$$\lambda\vec{x}' = \Pi g'\vec{X}' \qquad (8.13)$$

with $\vec{X}'$ in homogeneous coordinates and no explicit calibration matrix.

The relation (8.13) is similar to the calibrated case, but it relates uncalibrated quantities from the distorted space $\vec{X}'$, $\vec{x}'$ (here uncalibrated by interpretation) via the conjugate motion g'.

We also observe that an uncalibrated camera moving in a calibrated space

$$\lambda\vec{x}' = K\Pi g\vec{X} \qquad (8.14)$$

(K is the unknown calibration!) is equivalent to a calibrated camera moving in a distorted space, i.e.

$$\lambda\vec{x}' = \Pi g'\vec{X}' \qquad (8.15)$$

Thereby, we recall that

- $\vec{X}$ are Euclidean coordinates w.r.t. an undistorted space ('calibrated space');
- $\vec{X}'$ are uncalibrated camera coordinates;
- K is the unknown calibration.

Let us summarise these thoughts.

Proposition 8.1.1 *An uncalibrated camera with calibration matrix K viewing points in a calibrated Euclidean world moving with (R,T) is equivalent to a calibrated camera viewing points in a distorted space governed by an inner product $\langle\vec{u},\vec{v}\rangle_S$ with $S = K^{-\top}K^{-1}$ moving with (KRK^{-1}, KT).*

As a consequence, all algorithms described for the calibrated case can be transferred to the uncalibrated case by rewriting everything in terms of the new inner product $\langle \cdot, \cdot \rangle_S$. Only in the case $S = I$ the reconstruction corresponds up to a scalar to the true Euclidean structure.

Chapter 9

The Fundamental Matrix

We now study epipolar geometry for uncalibrated cameras. In particular, we will derive the epipolar constraint in terms of uncalibrated image coordinates, and we will see how the structure of the essential matrix is modified by the calibration matrix.

For simplicity of notation, we will assume $K = K_1 = K_2$, i.e. the same camera has captured both images.

9.1 Uncalibrated Epipolar Geometry

The epipolar constraint $\langle \vec{x}_2, \hat{T}R\vec{x}_1 \rangle = 0$ expresses, that the three vectors $\vec{x}_1$, $\vec{x}_2$ and T are coplanar.

In the uncalibrated space the corresponding three vectors are obtained by making use of the transform $\vec{x} = K^{-1}\vec{x}'$.

Substituting vectors in the epipolar constraint $\langle \vec{x}_2, \hat{T}R\vec{x}_1 \rangle = 0$ we compute:

$$\vec{x}_2^\top \hat{T}R\vec{x}_1 = 0 \quad \Leftrightarrow \quad \left(K^{-1}\vec{x}_2'\right)^\top \hat{T}R\left(K^{-1}\vec{x}_1\right) = 0 \tag{9.1}$$

so that we obtain the uncalibrated version of the epipolar constraint as

$$(\vec{x}_2')^\top K^{-\top}\hat{T}RK^{-1}\vec{x}_1 \;=\; 0 \tag{9.2}$$

The matrix

$$F \;:=\; K^{-\top}\hat{T}RK^{-1} \in \mathbb{R}^{3\times 3} \tag{9.3}$$

appearing in (9.2) is called the *fundamental matrix*.

As an alternative derivation, one may follow the same procedure as we proposed for the calibrated case, i.e. by elimination of the unknown depth scales λ_1, λ_2 from the rigid-body motion

$$\lambda_2\vec{x}_2 \;=\; R\lambda_1\vec{x}_1 + T \tag{9.4}$$

where $\lambda\vec{x} = \vec{X}$. Multiplication of (9.4) with the calibration matrix K gives

$$\lambda_2 K\vec{x}_2 = KR\lambda_1\vec{x}_1 + KT \quad \Leftrightarrow \quad \lambda_2\vec{x}'_2 = KRK^{-1}\lambda_1\vec{x}'_1 + T' \tag{9.5}$$

where $\vec{x}' = K\vec{x}$ and $T' = KT$. In order to eliminate the unknown depth, we multiply both sides with

$$T' \times \vec{x}'_2 = \hat{T}'\vec{x}'_2 \begin{cases} \perp \vec{x}'_2 \\ \perp T' \end{cases} \tag{9.6}$$

(of course taking the transpose) yielding

$$\lambda_1(\vec{x}'_2)^\top(\hat{T}')^\top KRK^{-1}\vec{x}'_1 = 0 \tag{9.7}$$

Using $(\hat{T}')^\top = -\hat{T}'$ and dividing by $-\lambda_1$ gives

$$(\vec{x}'_2)^\top\hat{T}'KRK^{-1}\vec{x}'_1 = 0 \tag{9.8}$$

This is an alternative form of the epipolar constraint, expressed using T'.

Let us consider the relation between (9.2) and (9.8). To this end we make use of

Lemma 9.1.1 *For a vector $T \in \mathbb{R}^3$ and a matrix $K \in \mathbb{R}^{3\times 3}$, if $det(K) = +1$ and $T' = KT$, then $\hat{T} = K^\top\hat{T}'K$.*

The *proof* makes use of the issue that for $det(K) = +1$, the mappings $K^\top\hat{\cdot}K$ and $\widehat{K^{-1}(\cdot)}$ act identically upon the canonical basis vectors of $\mathbb{R}^3$.

In our context, the lemma states that

$$K^{-\top}\hat{T}K^{-1} = \widehat{KT} \tag{9.9}$$

if $det(K) = +1$. Under the same condition we have

$$F = K^{-\top}\hat{T}RK^{-1} = \underbrace{K^{-\top}\hat{T}K^{-1}}_{=\hat{T}'}KRK^{-1} \tag{9.10}$$

Thus, for $det(K) = +1$ we have

$$F = K^{-\top}\hat{T}RK^{-1} = \hat{T}'KRK^{-1} \tag{9.11}$$

For $det(K) \neq 1$, one may simply scale all the matrices by a factor. We will have

$$K^{-\top}\hat{T}RK^{-1} \sim \hat{T}'KRK^{-1} \tag{9.12}$$

so that we can just assume $det(K) = +1$.

9.2 Properties of the Fundamental Matrix

The fundamental matrix F maps a point $\vec{x}_1'$ in the first view to a vector

$$\vec{l}_2 \ := \ F\vec{x}_1' \in \mathbb{R}^3 \tag{9.13}$$

in the second view, so that

$$(\vec{x}_2')^\top F\vec{x}_1' \ = \ (\vec{x}_2')^\top \vec{l}_2 \ = \ 0 \tag{9.14}$$

This means, the vector $\vec{l}_2$ defines implicitly a line in the second image plane as the collection of image points $\{\vec{x}_2'\}$ that satisfy

$$\vec{l}_2^\top \vec{x}_2' \ = \ 0 \tag{9.15}$$

Similarly, we may interprete the equation

$$\vec{l}_1 \ := \ F^\top \vec{x}_2' \in \mathbb{R}^3 \tag{9.16}$$

(obtained by taking the transpose in $(\vec{x}_2')^\top F\vec{x}_1'$) as F transferring a point in the second image to a line in the first.

These lines are again the *epipolar lines*. We can conclude:

Lemma 9.2.1 *Two image points $\vec{x}_1'$, $\vec{x}_2'$ correspond to a single point in space if and only if $\vec{x}_1'$ is on the epipolar line $\vec{l}_1 = F^\top \vec{x}_2'$, or equivalently, $\vec{x}_2'$ is on the epipolar line $\vec{l}_2 = F\vec{x}_1'$.*

This lemma is very useful in establishing correspondences. In fact, knowing the fundamental matrix allows to restrict the search for corresponding points to the epipolar lines only, rather than on the entire image.

Since the fundamental matrix F is the product of a skew symmetric matrix $\hat{T}'$ of rank 2 (for $T' \neq \vec{0}$) and a matrix $KRK^{-1} \in \mathbb{R}^{3\times 3}$ of rank 3, it must have rank 2. Hence, F can be characterised in terms of the SVD $F = U\Sigma V^\top$ with

$$\Sigma \ = \ \text{diag}(\sigma_1, \sigma_2, 0) \tag{9.17}$$

with some $\sigma_i \in \mathbb{R}_+$. In contrast to the essential matrix where $\Sigma = \text{diag}(\sigma, \sigma, 0)$ holds, we have only $\sigma_1 \geq \sigma_2$.

In the calibrated case the essential matrix provides enough information to recover pose and structure. In the uncalibrated case it is not possible to unravel R and T from F.

Let us consider the uncalibrated epipolar constraint:

$$(\vec{x}_2')^\top \hat{T}' KRK^{-1} \vec{x}_1' \;=\; 0 \tag{9.18}$$

Because $\hat{T}'(T'\vec{v}^\top) = (\hat{T}'T')\vec{v}^\top = 0$ for an arbitrary vector $\vec{v} \in \mathbb{R}^3$, we can manipulate (9.18) to have equivalently

$$(\vec{x}_2')^\top \hat{T}' \left(KRK^{-1} + T'\vec{v}^\top\right) \vec{x}_1' \;=\; 0 \tag{9.19}$$

Thus, if we wish to extract the relative camera pose (KRK^{-1}, KT) all we can do is to obtain instead

$$\Pi \;:=\; \left[KRK^{-1} + T'\vec{v}^\top, v_4 T'\right] \tag{9.20}$$

for some $v \in \mathbb{R}^3$ and $v_4 \in \mathbb{R}$. However, as one can show there is a canonical choice for fixing the open parameters in (9.20).

Index