Chapter 5

Reconstruction from two calibrated views

This chapter introduces the basic geometry of reconstruction of points in 3-D space from image measurements made from two different (calibrated) camera viewpoints. We then introduce a simple algorithm to recover the 3-D position of such points from their 2-D views. Although the algorithm is primarily conceptual, it illustrates the basic principles that will be used in deriving both more sophisticated algorithms as well as algorithms for the more general case of reconstruction from multiple views which will be discussed in Part III of the book.

The framework for this chapter is rather simple: Assume that we are given two views of a set of N feature points, which we denote formally as $\mathcal{I}$. The unknowns are the position of the points in 3-D space, denoted $\mathcal{Z}$ and the relative pose of the cameras, denoted $\mathcal{G}$, we will in this Chapter first derive constraints between these quantities. These constraints take the general form (one for each feature point)

$$f_i(\mathcal{I}, \mathcal{G}, \mathcal{Z}) = 0, \quad i = 1, \ldots, N \tag{5.1}$$

for some functions f_i. The central problem to be addressed in this Chapter is to start from this system of nonlinear equations and attempt, if possible, to solve it to recover the unknowns $\mathcal{G}$ and $\mathcal{Z}$. We will show that it is indeed possible to recover the unknowns.

The equations in (5.1) are nonlinear and no closed-form solution is known at this time. However, the functions f_i have a very particular structure that allows us to eliminate the parameters $\mathcal{Z}$ that contain information about the 3-D structure of the scene and be left with a constraint on $\mathcal{G}$ and $\mathcal{I}$ alone, which is known as the *epipolar constraint*. As we shall see in Section 5.1, the epipolar constraint has the general form

$$h_j(\mathcal{I}, \mathcal{G}) = 0, \quad j = 1, \ldots, N. \tag{5.2}$$

Interestingly enough, it can be solved for $\mathcal{G}$ *in closed form*, as we shall show in Section 5.2.

The closed-form algorithm, however, is not suitable for use in real images corrupted by noise. In Section 5.3, we discuss how to modify it so as to minimize the effect of noise. When the two views are taken from infinitesimally close vantage points, the basic geometry changes in ways that we describe in Section 5.4.

5.1 The epipolar constraint

We begin by assuming that we are given two images of the same scene taken from two distinct vantage points. Using the tools developed in Chapter 4, we can identify corresponding points in the two images. Under the assumption that the scene is *static* (there are no moving objects) and that the *brightness constancy constraint* is satisfied (Section **??**), corresponding points are images of the same point in space. Therefore, if we call $\mathbf{x}_1, \mathbf{x}_2$ the (homogeneous) coordinates of corresponding points on the image, these two points are related by a precise geometric relationship that we describe in this Section.

5.1.1 Discrete epipolar constraint and the essential matrix

Following the notation of Chapter 3, each camera is represented by an orthonormal reference frame and can therefore be described as a change of coordinates relative to an inertial reference frame. Without loss of generality, we can assume that the inertial frame corresponds to one of the two cameras, while the other is positioned and oriented according to $g = (R, T) \in SE(3)$. Let $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^3$ be the homogeneous coordinates of the projection of the same point p onto the two cameras. If we call $\mathbf{X}_1 \in \mathbb{R}^3$ and $\mathbf{X}_2 \in \mathbb{R}^3$ the 3-D coordinates of the point p relative to the two camera frames, they are related by a rigid body motion

$$\mathbf{X}_2 = R\mathbf{X}_1 + T$$

which can be written in terms of the images $\mathbf{x}_i$'s and the depths λ_i's as

$$\lambda_2 \mathbf{x}_2 = R\lambda_1 \mathbf{x}_1 + T.$$

In order to eliminate the depths λ_i's in the preceding equation, multiply both sides by $\widehat{T}$ to obtain

$$\lambda_2 \widehat{T}\mathbf{x}_2 = \widehat{T}R\lambda_1 \mathbf{x}_1.$$

Since the vector $\widehat{T}\mathbf{x}_2 = T \times \mathbf{x}_2$ is perpendicular to the vector $\mathbf{x}_2$, the inner product $\langle \mathbf{x}_2, \widehat{T}\mathbf{x}_2 \rangle = {\mathbf{x}_2}^T \widehat{T}\mathbf{x}_2$ is zero. This implies that the quantity $\mathbf{x}_2^T \widehat{T} R \lambda_1 \mathbf{x}_1$ is also zero. Since $\lambda_1 > 0$, the two image points $\mathbf{x}_1, \mathbf{x}_2$ satisfy the following constraint:

$$\boxed{\mathbf{x}_2^T \widehat{T} R \mathbf{x}_1 = 0.} \tag{5.3}$$

A geometric explanation for this constraint is immediate from Figure 5.1. The coordinates of the point p in the first camera $\mathbf{X}_1$, its coordinates in the second camera $\mathbf{X}_2$ and the vector connecting the two optical centers form a triangle. Therefore, the three vectors joining these points to the origin lie in the same plane and their triple product, which measures the volume they span is zero. This is true for the coordinates of the points $\mathbf{X}_i,\ i = 1, 2$ as well as for the homogeneous coordinates of their projection $\mathbf{x}_i,\ i = 1, 2$ since $\mathbf{X}_i$ and $\mathbf{x}_i$ (as two vectors) point in the same direction. The constraint (5.3) is just the triple product written in the reference frame of camera 2. The plane determined by the two centers of projection o_1, o_2 and the point p is called *epipolar plane*, the projection of the center of one camera onto the image plane of another camera $\mathbf{e}_i,\ i = 1, 2$ is called *epipole*, and the constraint (5.3) is called *epipolar constraint* or *essential constraint*. The matrix $E \doteq \widehat{T}R \in \mathbb{R}^{3\times 3}$ which captures the relative orientation between the two cameras is called the *essential matrix*.

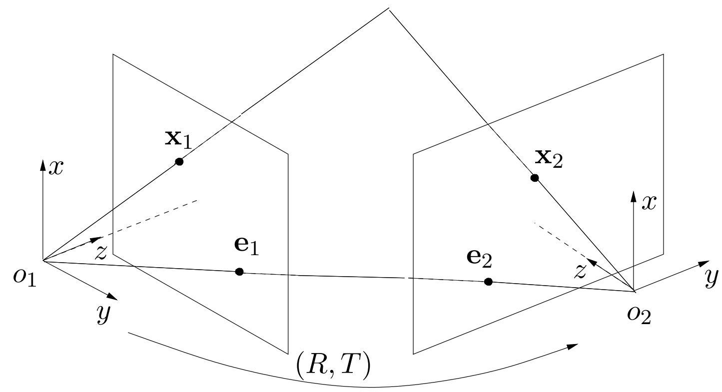

Figure 5.1: Two projections $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^3$ of a 3-D point p from two vantage points. The relative Euclidean transformation between the two vantage points is given by $(R, T) \in SE(3)$. The vectors $\mathbf{x}_1, \mathbf{x}_2$ and T are coplanar, and therefore their triple product (5.3) must be zero.

5.1.2 Elementary properties of the essential matrix

The matrix $E = \widehat{T}R \in \mathbb{R}^{3\times 3}$ in equation (5.3) contains information about the relative position T and orientation $R \in SO(3)$ between the two cameras. Matrices of this form belong to a very particular set of matrices in $\mathbb{R}^{3\times 3}$ which we call the *essential space* and denote by $\mathcal{E}$

$$\mathcal{E} := \left\{ \widehat{T}R \,\middle|\, R \in SO(3), T \in \mathbb{R}^3 \right\} \subset \mathbb{R}^{3\times 3}.$$

Before we study the structure of the space of essential matrices, we introduce a very useful lemma from linear algebra.

Lemma 5.1 (The hat operator). *If $T \in \mathbb{R}^3$, $A \in SL(3)$ and $T' = AT$, then $\widehat{T} = A^T \widehat{T'} A$.*

Proof. Since both $A^T\widehat{(\cdot)}A$ and $\widehat{A^{-1}(\cdot)}$ are linear maps from $\mathbb{R}^3$ to $\mathbb{R}^{3\times 3}$, one may directly verify that these two linear maps agree on the basis $[1,0,0]^T, [0,1,0]^T$ or $[0,0,1]^T$ (using the fact that $A \in SL(3)$ implies that $\det(A) = 1$). □

The following theorem, due to Huang and Faugeras [12], captures the algebraic structure of essential matrices:

Theorem 5.1 (Characterization of the essential matrix). *A non-zero matrix $E \in \mathbb{R}^{3\times 3}$ is an essential matrix* if and only if *E has a singular value decomposition (SVD): $E = U\Sigma V^T$ with*

$$\Sigma = diag\{\sigma, \sigma, 0\}$$

for some $\sigma \in \mathbb{R}_+$ and $U, V \in SO(3)$.

Proof. We first prove the necessity. By definition, for any essential matrix E, there exists (at least one pair) $(R,T), R \in SO(3), T \in \mathbb{R}^3$ such that $\widehat{T}R = E$. For T, there exists a rotation matrix R_0 such that $R_0 T = [0, 0, \|T\|]^T$. Denote this vector $a \in \mathbb{R}^3$. Since $\det(R_0) = 1$, we know $\widehat{T} = R_0^T \widehat{a} R_0$ from Lemma 5.1. Then $EE^T = \widehat{T}RR^T\widehat{T}^T = \widehat{T}\widehat{T}^T = R_0^T \widehat{a}\widehat{a}^T R_0$. It is direct to verify that

$$\widehat{a}\widehat{a}^T = \begin{bmatrix} 0 & -\|T\| & 0 \\ \|T\| & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & \|T\| & 0 \\ -\|T\| & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \|T\|^2 & 0 & 0 \\ 0 & \|T\|^2 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

So the singular values of the essential matrix $E = \widehat{T}R$ are $(\|T\|, \|T\|, 0)$. In general, in the SVD of $E = U\Sigma V^T$, U and V are unitary matrices, that is matrices whose columns are orthonormal, but whose determinants can be ± 1. We still need to prove that $U, V \in SO(3)$ (i.e. have determinant + 1) to establish the theorem. We already have $E = \widehat{T}R = R_0^T \widehat{a} R_0 R$. Let $R_Z(\theta)$ be the matrix which represents a rotation around the Z-axis (or X_3 axis) by an angle of θ radians, i.e. $R_Z(\theta) = e^{\widehat{e}_3\theta}$ with $e_3 = [0, 0, 1]^T \in \mathbb{R}^3$. Then

$$R_Z\left(+\frac{\pi}{2}\right) = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then $\widehat{a} = R_Z(+\frac{\pi}{2})R_Z^T(+\frac{\pi}{2})\widehat{a} = R_Z(+\frac{\pi}{2})\ diag\{\|T\|, \|T\|, 0\}$. Therefore

$$E = \widehat{T}R = R_0^T R_Z\left(+\frac{\pi}{2}\right)\ diag\{\|T\|, \|T\|, 0\} R_0 R.$$

So in the SVD of $E = U\Sigma V^T$, we may choose $U = R_0^T R_Z(+\frac{\pi}{2})$ and $V^T = R_0 R$. By construction, both U and V are rotation matrices.

We now prove the sufficiency. If a given matrix $E \in \mathbb{R}^{3\times 3}$ has a SVD: $E = U\Sigma V^T$ with $U, V \in SO(3)$ and $\Sigma = diag\{\sigma, \sigma, 0\}$. Define $(R_1, T_1) \in SE(3)$ and $(R_2, T_2) \in SE(3)$ to be

$$\begin{cases} (\widehat{T}_1, R_1) &= (UR_Z(+\frac{\pi}{2})\Sigma U^T, UR_Z^T(+\frac{\pi}{2})V^T), \\ (\widehat{T}_2, R_2) &= (UR_Z(-\frac{\pi}{2})\Sigma U^T, UR_Z^T(-\frac{\pi}{2})V^T) \end{cases} \tag{5.4}$$

It is now easy to verify that $\widehat{T}_1 R_1 = \widehat{T}_2 R_2 = E$. Thus, E is an essential matrix. □

Given a rotation matrix $R \in SO(3)$ and a rotation vector $T \in \mathbb{R}^3$, it is immediate to construct an essential matrix $E = \widehat{T}R$. The inverse problem, that is how to reconstruct T and R from a given essential matrix E, is less obvious. Before we show how to solve it in Theorem 5.2, we need the following preliminary lemma.

Lemma 5.2. *Consider an arbitrary non-zero skew symmetric matrix $\widehat{T} \in so(3)$ with $T \in \mathbb{R}^3$. If, for a rotation matrix $R \in SO(3)$, $\widehat{T}R$ is also a skew symmetric matrix, then $R = I$ or $e^{\widehat{u}\pi}$ where $u = \frac{T}{\|T\|}$. Further, $\widehat{T}e^{\widehat{u}\pi} = -\widehat{T}$.*

Proof. Without loss of generality, we assume T is of unit length. Since $\widehat{T}R$ is also a skew symmetric matrix, $(\widehat{T}R)^T = -\widehat{T}R$. This equation gives

$$R\widehat{T}R = \widehat{T}. \tag{5.5}$$

Since R is a rotation matrix, there exists $\omega \in \mathbb{R}^3, \|\omega\| = 1$ and $\theta \in \mathbb{R}$ such that $R = e^{\widehat{\omega}\theta}$. Then, (5.5) is rewritten as $e^{\widehat{\omega}\theta}\widehat{T}e^{\widehat{\omega}\theta} = \widehat{T}$. Applying this equation to ω, we get $e^{\widehat{\omega}\theta}\widehat{T}e^{\widehat{\omega}\theta}\omega = \widehat{T}\omega$. Since $e^{\widehat{\omega}\theta}\omega = \omega$, we obtain $e^{\widehat{\omega}\theta}\widehat{T}\omega = \widehat{T}\omega$. Since ω is the only eigenvector associated to the eigenvalue 1 of the matrix $e^{\widehat{\omega}\theta}$ and $\widehat{T}\omega$ is orthogonal to ω, $\widehat{T}\omega$ has to be zero. Thus, ω is equal to either $\frac{T}{\|T\|}$ or $-\frac{T}{\|T\|}$, i.e. $\omega = \pm u$. R then has the form $e^{\widehat{\omega}\theta}$, which commutes with $\widehat{T}$. Thus from (5.5), we get

$$e^{2\widehat{\omega}\theta}\widehat{T} = \widehat{T}. \tag{5.6}$$

According to *Rodrigues' formula* [21], we have

$$e^{2\widehat{T}\theta} = I + \widehat{\omega}\sin(2\theta) + \widehat{\omega}^2(1 - \cos(2\theta)).$$

(5.6) yields

$$\widehat{\omega}^2 \sin(2\theta) + \widehat{\omega}^3(1 - \cos(2\theta)) = 0.$$

Since $\widehat{\omega}^2$ and $\widehat{\omega}^3$ are linearly independent (Lemma 2.3 in [21]), we have $\sin(2\theta) = 1 - \cos(2\theta) = 0$. That is, θ is equal to $2k\pi$ or $2k\pi + \pi$, $k \in \mathbb{Z}$. Therefore, R is equal to I or $e^{\widehat{\omega}\pi}$. Now if $\omega = u = \frac{T}{\|T\|}$ then, it is direct from the geometric meaning of the rotation $e^{\widehat{\omega}\pi}$ that $e^{\widehat{\omega}\pi}\widehat{T} = -\widehat{T}$. On the other hand if $\omega = -u = -\frac{T}{\|T\|}$ then it follows that $e^{\widehat{\omega}}\widehat{T} = -\widehat{T}$. Thus, in any case the conclusions of the lemma follows. □

The following theorem shows how to extract rotation and translation from an essential matrix as given in closed-form in equation (5.7) at the end of the theorem.

Theorem 5.2 (Pose recovery from the essential matrix). *There exist exactly two relative poses (R, T) with $R \in SO(3)$ and $T \in \mathbb{R}^3$ corresponding to a non-zero essential matrix $E \in \mathcal{E}$.*

Proof. Assume that $(R_1, T_1) \in SE(3)$ and $(R_2, p_2) \in SE(3)$ are both solutions for the equation $\widehat{T}R = E$. Then we have $\widehat{T}_1 R_1 = \widehat{T}_2 R_2$. It yields $\widehat{T}_1 = \widehat{T}_2 R_2 R_1^T$. Since $\widehat{T}_1, \widehat{T}_2$ are both skew symmetric matrices and $R_2 R_1^T$ is a rotation matrix, from the preceding lemma, we have that either $(R_2, T_2) = (R_1, T_1)$ or $(R_2, T_2) = (e^{\widehat{u}_1\pi} R_1, -T_1)$ with $u_1 = T_1/\|T_1\|$. Therefore, given an essential matrix E there are exactly *two* pairs (R, T) such that $\widehat{T}R = E$. Further, if E has the SVD: $E = U\Sigma V^T$ with $U, V \in SO(3)$, the following formulas give the two distinct solutions

$$\boxed{\begin{array}{rl} (\widehat{T}_1, R_1) = & (UR_Z(+\frac{\pi}{2})\Sigma U^T, UR_Z^T(+\frac{\pi}{2})V^T), \\ (\widehat{T}_2, R_2) = & (UR_Z(-\frac{\pi}{2})\Sigma U^T, UR_Z^T(-\frac{\pi}{2})V^T). \end{array}} \tag{5.7}$$

□

5.2 Closed-form reconstruction

In the previous section, we have seen that images of corresponding points are related by the epipolar constraint, which involves the unknown relative pose between the cameras. Therefore, given a number of corresponding points, we could use the epipolar constraints to try to recover camera pose. In this section, we show a simple closed-form solution to this problem. It consists of two steps: First a matrix E is recovered from a number of epipolar constraints, then relative translation and orientation is extracted from E. However, since the matrix E recovered using correspondence data in the epipolar constraint may not be an essential matrix it needs to be projected into the space of essential matrices prior to applying the formula of equation (5.7).

Although the algorithm that we will propose is mainly conceptual and does not work well in the presence of large noise or uncertainty, it is important in its illustration of the geometric structure of the space of essential matrices. We will deal with noise and uncertainty in Section 5.3.

5.2.1 The eight-point linear algorithm

Let $E = \widehat{T}R$ be the essential matrix associated with the epipolar constraint (5.3). When the entries of this 3×3 matrix are denoted as

$$E = \begin{bmatrix} e_1 & e_2 & e_3 \\ e_4 & e_5 & e_6 \\ e_7 & e_8 & e_9 \end{bmatrix} \tag{5.8}$$

and arrayed in a vector, which we define to be the *essential vector* $\mathbf{e} \in \mathbb{R}^9$, we have

$$\mathbf{e} = [e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9]^T \in \mathbb{R}^9.$$

Since the epipolar constraint $\mathbf{x}_2^T E \mathbf{x}_1 = 0$ is linear in the entries of E, we can rewrite it as a linear equation in the entries of e, namely, $\mathbf{a}^T \mathbf{e} = 0$ for some $\mathbf{a} \in \mathbb{R}^9$. The vector $\mathbf{a} \in \mathbb{R}^9$ is a function of the two corresponding image points, $\mathbf{x}_1 = [x_1, y_1, z_1]^T \in \mathbb{R}^3$ and $\mathbf{x}_2 = [x_2, y_2, z_2]^T \in \mathbb{R}^3$ and is given by

$$\mathbf{a} = [x_2x_1, x_2y_1, x_2z_1, y_2x_1, y_2y_1, y_2z_1, z_2x_1, z_2y_1, z_2z_1]^T \in \mathbb{R}^9. \tag{5.9}$$

The epipolar constraint (5.3) can then be rewritten as the inner product of $\mathbf{a}$ and $\mathbf{e}$

$$\mathbf{a}^T \mathbf{e} = 0.$$

Now, given a set of corresponding image points $(\mathbf{x}_1^j, \mathbf{x}_2^j)$, $j = 1, \ldots, n$, define a matrix $A \in \mathbb{R}^{n \times 9}$ associated with these measurements to be

$$A = [\mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^n]^T \tag{5.10}$$

where the j^{th} row $\mathbf{a}^j$ is constructed from each pair $(\mathbf{x}_1^j, \mathbf{x}_2^j)$ using (5.9). In the absence of noise, the essential vector $\mathbf{e}$ has to satisfy

$$A\mathbf{e} = 0. \tag{5.11}$$

This linear equation may now be solved for the vector $\mathbf{e}$. For the solution to be unique (up to scale, ruling out the trivial solution $\mathbf{e} = 0$), the rank of the matrix $A \in \mathbb{R}^{9 \times n}$ needs to be exactly eight. This should be the case given $n \geq 8$ "ideal" corresponding points. In general, however, since correspondences may be noisy there may be no solution to (5.11). In such a case, one may choose e to minimize the function $\|A\mathbf{e}\|^2$, i.e. choose $\mathbf{e}$ is the eigenvector of A^TA corresponding to its smallest eigenvalue. Another condition to be cognizant of is when the rank of A is less than 8, allowing for multiple solutions of equation (5.11). This happens when the feature points are not in "general position", for example when they all lie in a plane. Again in the presence of noise when the feature points are in general position there are multiple small eigenvalues of A^TA, corresponding to the lack of conditioning of the data.

However, even in the absence of noise, for a vector $\mathbf{e}$ to be the solution of our problem, it is not sufficient that it is the null space of A. In fact, it has to satisfy an additional constraint, in that its matrix form E must belong to the space of essential matrices. Enforcing this structure in the determination of the null space of A is difficult. Therefore, as a first cut, we could first estimate the null space of A *ignoring the internal structure of* E, obtaining a matrix possibly not belong to the essential space, and then *orthogonally projecting* the matrix thus obtained onto the essential space. The following theorem says precisely what this projection is.

Theorem 5.3 (Projection onto the essential space). *Given a real matrix $F \in \mathbb{R}^{3\times3}$ with a SVD: $F = U diag\{\lambda_1, \lambda_2, \lambda_3\} V^T$ with $U, V \in SO(3)$, $\lambda_1 \geq \lambda_2 \geq \lambda_3$, then the essential matrix $E \in \mathcal{E}$ which minimizes the error $\|E - F\|_f^2$ is given by $E = U diag\{\sigma, \sigma, 0\} V^T$ with $\sigma = (\lambda_1 + \lambda_2)/2$. The subscript f indicates the Frobenius norm.*

Proof. For any fixed matrix $\Sigma = diag\{\sigma, \sigma, 0\}$, we define a subset $\mathcal{E}_\Sigma$ of the essential space $\mathcal{E}$ to be the set of all essential matrices with SVD of the form $U_1 \Sigma V_1^T$, $U_1, V_1 \in SO(3)$. To simplify the notation, define $\Sigma_\lambda = diag\{\lambda_1, \lambda_2, \lambda_3\}$. We now prove the theorem by two steps:

Step 1: We prove that for a fixed Σ, the essential matrix $E \in \mathcal{E}_\Sigma$ which minimizes the error $\|E - F\|_f^2$ has a solution $E = U\Sigma V^T$ (not necessarily unique). Since $E \in \mathcal{E}_\Sigma$ has the form $E = U_1 \Sigma V_1^T$, we get

$$\|E - F\|_f^2 = \|U_1 \Sigma V_1^T - U\Sigma_\lambda V^T\|_f^2 = \|\Sigma_\lambda - U^T U_1 \Sigma V_1^T V\|_f^2.$$

Define $P = U^T U_1, Q = V^T V_1 \in SO(3)$ which have the forms

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}, \quad Q = \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}. \tag{5.12}$$

Then

$$\|E - F\|_f^2 = \|\Sigma_\lambda - U^T U_1 \Sigma V_1^T V\|_f^2 = tr(\Sigma_\lambda^2) - 2tr(P\Sigma Q^T \Sigma_\lambda) + tr(\Sigma^2).$$

Expanding the second term, using $\Sigma = diag\{\sigma, \sigma, 0\}$ and the notation p_{ij}, q_{ij} for the entries of P, Q, we have

$$tr(P\Sigma Q^T \Sigma_\lambda) = \sigma(\lambda_1(p_{11}q_{11} + p_{12}q_{12}) + \lambda_2(p_{21}q_{21} + p_{22}q_{22})).$$

Since P, Q are rotation matrices, $p_{11}q_{11} + p_{12}q_{12} \leq 1$ and $p_{21}q_{21} + p_{22}q_{22} \leq 1$. Since Σ, Σ_λ are fixed and $\lambda_1, \lambda_2 \geq 0$, the error $\|E - F\|_f^2$ is minimized when $p_{11}q_{11} + p_{12}q_{12} = p_{21}q_{21} + p_{22}q_{22} = 1$. This can be achieved when P, Q are of the general form

$$P = Q = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Obviously $P = Q = I$ is one of the solutions. That implies $U_1 = U, V_1 = V$.

Step 2: From Step 1, we only need to minimize the error function over the matrices of the form $U\Sigma V^T \in \mathcal{E}$ where Σ may vary. The minimization problem is then converted to one of minimizing the error function

$$\|E - F\|_f^2 = (\lambda_1 - \sigma)^2 + (\lambda_2 - \sigma)^2 + (\lambda_3 - 0)^2.$$

Clearly, the σ which minimizes this error function is given by $\sigma = (\lambda_1 + \lambda_2)/2$. □

As we have already pointed out, the epipolar constraint only allows for the recovery of the essential matrix up to a scale (since the epipolar constraint of 5.3 is homogeneous in E, it is not modified by multiplying by any non-zero constant). A typical choice to address this ambiguity is to choose E to have its non-zero singular values to be 1, which corresponds to unit translation, that is, $\|T\| = 1$. Therefore, if E is the essential matrix recovered from image data, we can normalize it at the outset by replacing it with $\frac{E}{\|E\|}$. Note that the sign of the essential matrix is also arbitrary. According to Theorem 5.2, each normalized essential matrix gives two possible poses. So, in general, we can only recover the pose up to four solutions.[1] The overall algorithm, which is due to Longuet-Higgins [14], can then be summarized as follows

[1]Three of them can be eliminated once imposing the positive depth constraint.

Algorithm 5.1 (The eight-point algorithm). *For a given set of image correspondences* $(\mathbf{x}_1^j, \mathbf{x}_2^j)$, $j = 1, \ldots, n$ $(n \geq 8)$, *this algorithm finds* $(R, T) \in SE(3)$ *which solves*

$$\mathbf{x}_2^{jT} \widehat{T} R \mathbf{x}_1^j = 0, \quad j = 1, \ldots, n.$$

1. **Compute a first approximation of the essential matrix**
   Construct the $A \in \mathbb{R}^{n \times 9}$ *from correspondences* $\mathbf{x}_1^j$ *and* $\mathbf{x}_2^j$ *as in (5.9), namely.*

   $$\mathbf{a}^j = [x_2^j x_1^j, x_2^j y_1^j, x_2^j z_1^j, y_2^j x_1^j, y_2^j y_1^j, y_2^j z_1^j, z_2^j x_1^j, z_2^j y_1^j, z_2^j z_1^j]^T \in \mathbb{R}^9.$$

   Find the vector $\mathbf{e} \in \mathbb{R}^9$ *of unit length such that* $\|A\mathbf{e}\|$ *is minimized as follows: compute the SVD* $A = U_A \Sigma_A V_A^T$ *and define* $\mathbf{e}$ *to be the* 9^{th} *column of* V_A. *Rearrange the 9 elements of* $\mathbf{e}$ *into a square* 3×3 *matrix* E *as in (5.8). Note that this matrix will in general* not *be an essential matrix.*

2. **Project onto the essential space**
   Compute the Singular Value Decomposition of the matrix E *recovered from data to be*

   $$E = U diag\{\sigma_1, \sigma_2, \sigma_3\} V^T$$

   where $\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq 0$ *and* $U, V \in SO(3)$. *In general, since* E *may not be an essential matrix,* $\sigma_1 \neq \sigma_2$ *and* $\sigma_3 > 0$. *Compute its projection onto the essential space as* $U \Sigma V^T$, *where* $\Sigma = diag\{1, 1, 0\}$.

3. **Recover displacement from the essential matrix**
   Define the diagonal matrix Σ *to be Extract* R *and* T *from the essential matrix as follows:*

   $$R = U R_Z^T \left(\pm \frac{\pi}{2}\right) V^T, \quad \widehat{T} = U R_Z \left(\pm \frac{\pi}{2}\right) \Sigma U^T.$$

Remark 5.1 (Infinitesimal viewpoint change). *It is often the case in applications that the two views described in this chapter are taken by a moving camera rather than by two static cameras. The derivation of the epipolar constraint and the associated eight-point algorithm does not change, as long as the two vantage points are distinct. In the limit that the two viewpoints come infinitesimally close, the epipolar constraint takes a distinctly different form called the continuous epipolar constraint which we will introduce in Section 5.4.*

Remark 5.2 (Enough parallax). *In the derivation of the epipolar constraint we have implicitly assumed that* $E \neq 0$, *which allowed us to derive the eight-point algorithm where the epipolar matrix is normalized to* $\|E\| = 1$. *Due to the structure of the essential matrix,* $E = 0 \Leftrightarrow T = 0$. *Therefore, the eight-point algorithm requires that* $T \neq 0$. *The translation vector* T *is also known as "parallax". In practice, due to noise, the algorithm will likely return an answer even when there is no parallax. However, in that case the estimated direction of translation will be meaningless. Therefore, one needs to exercise caution to make sure that there is "enough parallax" for the algorithm to operate correctly.*

Remark 5.3 (Positive depth constraint). *Since both* E *and* $-E$ *satisfy the same set of epipolar constraints, they in general give rise to* $2 \times 2 = 4$ *possible solutions to* (R, T). *However, this does on pose a potential problem because there is only one of them guarantees that the depths of the 3-D points being observed by the camera are all* positive. *That is, in general, three out of the four solutions will be physically impossible and hence may be discarded.*

Remark 5.4 (General position requirement). *In order for the above algorithm to work properly, the condition that the given 8 points are in "general position" is very important. It can be easily shown that if these points form certain degenerate configurations, the algorithm will fail. A case of some pratical importance is when all the points happen to lie on the same 2-D plane in $\mathbb{R}^3$. We will discuss the geometry for image features from a plane in a later chapter (in a more general setting of multiple views). Nonetheless, we encourage the reader to solve the Exercise (7) at the end of this chapter that illustrates some basic ideas how to explicitly take into account such coplanar information and modify the essential constraint accordingly.*

5.2.2 Euclidean constraints and structure reconstruction

The eight-point algorithm just described uses as input a set of eight or more point correspondences and returns the relative pose (rotation and translation) between the two cameras up a an arbitrary scale $\gamma \in \mathbb{R}^+$. Relative pose and point correspondences can then be used to retrieve the position of the points in 3-D space by recovering their depths relative to each camera frame.

Consider the basic rigid body equation, where the pose (R, T) has been recovered, with the translation T defined up to the scale γ. The usual rigid body motion, written in terms of the images and the depths, is given by

$$\lambda_2^j \mathbf{x}_2^j = \lambda_1^j R \mathbf{x}_1^j + \gamma T, \quad 1 \leq j \leq n. \tag{5.13}$$

Notice that, since (R, T) are known, the equations given by (5.13) are linear in both the structural scales λ's and the motion scales γ's and could be therefore easily solved. In the presence of noise, these scales can be retrieved by solving a standard linear least-squares estimation problem, for instance in the least-squares sense. Arrange all the unknown scales in (5.13) into an *extended scale vector* $\vec{\lambda}$

$$\vec{\lambda} = [\lambda_1^1, \ldots, \lambda_1^n, \lambda_2^1, \ldots, \lambda_2^n, \gamma]^T \in \mathbb{R}^{2n+1}.$$

Then all the constraints given by (5.13) can be expressed as a single linear equation

$$M\vec{\lambda} = 0 \tag{5.14}$$

where $M \in \mathbb{R}^{3n\times(2n+1)}$ is a matrix depending on (R, T) and $\{(\mathbf{x}_1^j, \mathbf{x}_2^j)\}_{j=1}^n$. For the linear equation (5.14) to have a unique solution, the matrix M needs to have a rank $2n$. In the absence of noise, the matrix M is generically of rank $2n$ if enough points are used[2], and it has a one-dimensional null space. Hence the equation (5.14) determines all the unknown scales up to a single universal scaling. The linear least-squares estimate of $\vec{\lambda}$ is simply the eigenvector of $M^T M$ which corresponds to the smallest eigenvalue. However, the obtained reconstruction may suffer from certain inconsistency that the 3-D locations of a point recovered from different vantage points do not necessarily coincide. In the next section, we address this issue and, at the mean time, address the issue of optimal reconstruction in the presence of noise.

5.3 Optimal reconstruction

The eight-point algorithm to reconstruct camera pose, as described in the previous section, assumes that *exact* point correspondences are given. In the presence of noise in the correspondence, we have

[2]disregarding, of course, configurations of points which are not in "general position".

hinted at possible ways of estimating the essential matrix by solving a least-squares problem. In this section, we describe a more principled approach to minimize the effects of noise in the reconstruction. It requires solving an optimization problem on a differentiable manifold. Rather than develop the machinery to solve such optimization problems in this Section, we will only formulate the problem here and describe an algorithm at a conceptual level. Details of the algorithm as well as more on optimization on manifolds are described in Appendix C.

As we have seen in Chapter 6, a calibrated camera can be described as a plane perpendicular to the Z-axis at a distance of 1 unit from the origin. Therefore, coordinates of image points $\mathbf{x}_1$ and $\mathbf{x}_2$ are of the form $[x, y, 1]^T \in \mathbb{R}^3$. In practice, we cannot measure the actual coordinates but rather their noisy versions

$$\tilde{\mathbf{x}}_1^j = \mathbf{x}_1^j + w_1^j, \quad \tilde{\mathbf{x}}_2^j = \mathbf{x}_2^j + {w_2}^j, \quad j = 1, \ldots, n \tag{5.15}$$

where $\mathbf{x}_1^j$ and $\mathbf{x}_2^j$ are the "ideal" image coordinates and $w_1^j = [{w_1}_1^j,\ {w_1}_2^j,\ 0]^T$ and $w_2^j = [{w_2}_1^j,\ {w_2}_2^j,\ 0]^T$ are localization errors in the correspondence. Notice that it is the (unknown) ideal image coordinates $\mathbf{x}_i^j,\ i = 1, 2$ that satisfy the epipolar constraint $\mathbf{x}_2^{jT}\widehat{T}R\mathbf{x}_1^j = 0$, and *not* the (measured) noisy ones $\tilde{\mathbf{x}}_i^j,\ i = 1, 2$. One could think of the ideal coordinates as a "model" that depends upon the unknown parameters (R, T), and w_i^j as the discrepancy between the model and the measurements: $\tilde{\mathbf{x}}_i^j = \mathbf{x}_i^j(R, T) + w_i^j$. Therefore, one seeks the parameters (R, T) that minimize the discrepancy from the data, i.e. w_i^j. It is clear that, in order to proceed, we need to define a discrepancy criterion.

The simplest choice is to assume that w_i^j are unknown errors and to minimize their norm; for instance, the standard Euclidean two-norm $\|w\|_2 \doteq \sqrt{w^T w}$. Indeed, minimizing the squared norm is equivalent and often results in simpler algorithms: $\hat{R}, \hat{T} = \arg\min_{R,T} \phi(R, T)$ where

$$\phi(R, T) \doteq \sum_{i,j} \|w_i^j\|_2^2 = \sum_{i,j} \|\tilde{\mathbf{x}}_i^j - \mathbf{x}_i^j(R, T)\|_2^2.$$

This corresponds to a *"least-squares" criterion* (LS). Notice that the unknowns are constrained to live in a non-linear space: $R \in SO(3)$ and $T \in \mathbb{S}^2$, the unit sphere, after normalization due to the unrecoverable global scale factor, as discussed in the eight-point algorithm.

Alternatively, one may assume that w_i^j are samples of a stochastic process with a certain, known distribution that depends upon the unknown parameters (R, T), $w_i^j \sim p(w|R, T)$ and maximize the (log)-likelihood function with respect to the parameters: $\hat{R}, \hat{T} = \arg\max_{R,T} \phi(R, T)$, where

$$\phi(R, T) \doteq \sum_{i,j} \log p((\tilde{\mathbf{x}}_i^j - \mathbf{x}_i^j)|R, T).$$

This corresponds to a *"maximum likelihood" criterion* (ML).

If prior information about the unknown parameters (R, T) is available, for instance, their prior density $p(R, T)$, one may combine it with the log-likelihood function to obtain a posterior density $p(R, T|\tilde{\mathbf{x}}_j^j)$ using Bayes' rule, and seek $\hat{R}, \hat{T} = \arg\max_{R,T} \phi(R, T)$, where

$$\phi \doteq p(R, T|\tilde{\mathbf{x}}_i^j\ \forall\ i, j)$$

This choice corresponds to a *"maximum a-posteriori" criterion* (MAP). Although conceptually more "principled", this choice requires defining a probability density on the space of camera pose $SO(3) \times \mathbb{S}^2$ which has a non-trivial Riemannian structure. Not only is this prior not easy to obtain, but also its description is beyond the scope of this book and will therefore not be further explored.

However, it is remarkable that both the ML and LS criteria in the end reduce to optimization problems on the manifold $SO(3) \times \mathbb{S}^2$. The discussion in this section can be applied to ML as well as to LS, however we will choose the LS for simplicity. Now that we have chosen a criterion, the least-squares norm, we can pose the problem of optimal reconstruction by collecting all available constraints as follows: given $\tilde{\mathbf{x}}_i^j, \quad i = 1, 2, \; j = 1, \ldots, n$, find

$$\hat{R}, \hat{T} = \arg\min \sum_{j=1}^{n} \sum_{i=1}^{2} \|w_i^j\|_2^2$$

subject to

$$\begin{cases} \tilde{\mathbf{x}}_i^j = \mathbf{x}_i^j + {w_i}^j, \quad i = 1, 2, j = 1, \ldots, n \\ \mathbf{x}_2^{jT} \widehat{T} R \mathbf{x}_1^j = 0, \quad j = 1, \ldots, n \\ \mathbf{x}_1^{jT} e_3 = 1, \quad j = 1, \ldots, n \\ \mathbf{x}_2^{jT} e_3 = 1, \quad j = 1, \ldots, n \\ R \in SO(3) \\ T \in \mathbb{S}^2 \end{cases} \tag{5.16}$$

Using Lagrange multipliers $\lambda^j, \gamma^j, \eta^j$, we can convert the above minimization problem to an unconstrained minimization problem over $R, T, \mathbf{x}_i^j, \mathbf{x}_2^j, \lambda^j, \gamma^j, \eta^j$ with the constraints of equation (5.16) as

$$\min_{R \in SO(3), T \in \mathbb{S}^2, \mathbf{x}_1^j, \mathbf{x}_2^j} \sum_{j=1}^{n} \|\tilde{\mathbf{x}}_1^j - \mathbf{x}_1^j\|^2 + \|\tilde{\mathbf{x}}_2^j - \mathbf{x}_2^j\|^2 + \lambda^j \mathbf{x}_2^{jT} \widehat{T} R \mathbf{x}_1^j + \gamma^j (\mathbf{x}_1^{jT} e_3 - 1) + \eta^j (\mathbf{x}_2^{jT} e_3 - 1). \tag{5.17}$$

The necessary conditions for the existence of a minimum are:

$$2(\tilde{\mathbf{x}}_1^j - \mathbf{x}_1^j) + \lambda^j R^T \widehat{T}^T \mathbf{x}_2^j + \gamma^j e_3 = 0$$
$$2(\tilde{\mathbf{x}}_2^j - \mathbf{x}_2^j) + \lambda^j \widehat{T} R \mathbf{x}_1^j + \eta^j e_3 = 0$$

Simplifying the necessary conditions, we obtain:

$$\left\{ \begin{array}{ccc} \mathbf{x}_1^j & = & \tilde{\mathbf{x}}_1^j - \frac{1}{2} \lambda^j \widehat{e}_3^T \widehat{e}_3 R^T \widehat{T}^T \mathbf{x}_2^j \\ \mathbf{x}_2^j & = & \tilde{\mathbf{x}}_2^j - \frac{1}{2} \lambda^j \widehat{e}_3^T \widehat{e}_3 \widehat{T} R \mathbf{x}_1^j \\ \mathbf{x}_2^{jT} \widehat{T} R \mathbf{x}_1^j & = & 0 \end{array} \right. \tag{5.18}$$

where λ^j is given by:

$$\lambda^j = \frac{2(\mathbf{x}_2^{jT} \widehat{T} R \tilde{\mathbf{x}}_1^j + \tilde{\mathbf{x}}_2^{jT} \widehat{T} R \mathbf{x}_1^j)}{\mathbf{x}_1^{jT} R^T \widehat{T}^T \widehat{e}_3^T \widehat{e}_3 \widehat{T} R \mathbf{x}_1^j + \mathbf{x}_2^{jT} \widehat{T} R \widehat{e}_3^T \widehat{e}_3 R^T \widehat{T}^T \mathbf{x}_2^j} \tag{5.19}$$

or

$$\lambda^j = \frac{2\mathbf{x}_2^{jT} \widehat{T} R \tilde{\mathbf{x}}_1^j}{\mathbf{x}_1^{jT} R^T \widehat{T}^T \widehat{e}_3^T \widehat{e}_3 \widehat{T} R \mathbf{x}_1^j} = \frac{2\tilde{\mathbf{x}}_2^{jT} \widehat{T} R \mathbf{x}_1^j}{\mathbf{x}_2^{jT} \widehat{T} R \widehat{e}_3^T \widehat{e}_3 R^T \widehat{T}^T \mathbf{x}_2^j}. \tag{5.20}$$

Substituting (5.18) and (5.19) into the least-squares cost function of equation (5.17, we obtain:

$$\phi(R, T, \mathbf{x}_1^j, \mathbf{x}_2^j) \quad = \quad \sum_{j=1}^{n} \frac{(\mathbf{x}_2^{jT} \widehat{T} R \tilde{\mathbf{x}}_1^j + \tilde{\mathbf{x}}_2^{jT} \widehat{T} R \mathbf{x}_1^j)^2}{\|\widehat{e}_3 \widehat{T} R \mathbf{x}_1^j\|^2 + \|\mathbf{x}_2^{jT} \widehat{T} R \widehat{e}_3^T\|^2} \tag{5.21}$$

If one uses instead (5.18) and (5.20), we get:

$$\phi(R, T, \mathbf{x}_1^j, \mathbf{x}_2^j) = \sum_{j=1}^{n} \frac{(\tilde{\mathbf{x}}_2^{jT}\widehat{T}R\mathbf{x}_1^j)^2}{\|\widehat{e}_3\widehat{T}R\mathbf{x}_1^j\|^2} + \frac{(\mathbf{x}_2^{jT}\widehat{T}\tilde{\mathbf{x}}_1^j)^2}{\|\mathbf{x}_2^{jT}\widehat{T}R\widehat{e}_3^T\|^2}. \tag{5.22}$$

Geometrically, both expressions can be interpreted as distances of the image points $\tilde{\mathbf{x}}_1^j$ and $\tilde{\mathbf{x}}_2^j$ from the epipolar lines specified by $\mathbf{x}_1^j, \mathbf{x}_2^j$ and (R, T), as shown in Figure 5.2. We leave the verification of this to the reader as an exercise.

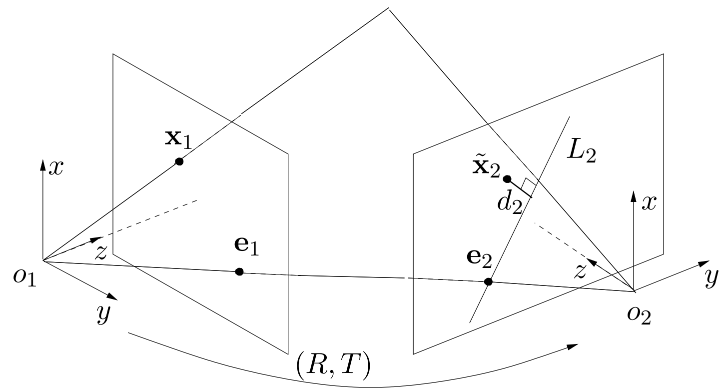

Figure 5.2: Two noisy image points $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^3$. L_2 is the so-called *epipolar line* which is the intersection of the second image plane with the plane formed by the first image point $\mathbf{x}_1$ and the line connecting the two camera centers o_1, o_2. The distance d_2 is the geometric distance between the second image point $\tilde{\mathbf{x}}_2$ and the epipolar line in the second image plane. Symmetrically, one can define a similar geometric distance d_1 in the first image plane.

These expressions for ϕ can finally be minimized with respect to (R, T) as well as $\mathbf{x}_i^j$. In doing so, however, one has to make sure that the optimization proceeds on the space $R \in SO(3)$ and $T \in \mathbb{S}^2$ are enforced. In Appendix C we discuss methods of minimizing function defined on the space $SO(3) \times \mathbb{S}^2$, which can be used to minimize $\phi(R, T, \mathbf{x}_i^j)$ once $\mathbf{x}_i^j$ are known. Since $\mathbf{x}_i^j$ are *not* known, one can set up an alternating minimization scheme where an initial approximation of $\mathbf{x}_i^j$ is used to estimate an approximation of (R, T) which is used, in turn, to update the estimates of $\mathbf{x}_i^j$. It can be shown that each such iteration decreases the cost function, and therefore convergence to a local extremum is guaranteed since the cost function is bounded below by zero.

Algorithm 5.2 (Optimal triangulation).

1. **Initialization**
   Initialize $\mathbf{x}_1^j(R, T)$ *and* $\mathbf{x}_2^j(R, T)$ *as* $\tilde{\mathbf{x}}_1^j$ *and* $\tilde{\mathbf{x}}_2^j$ *respectively.*

2. **Motion estimation**
   Update (R, T) *by minimizing* $\phi(R, T, \mathbf{x}_1^j(R, T), \mathbf{x}_2^j(R, T))$ *given by (5.21) or (5.22).*

3. **Structure triangulation**
   Solve for $\mathbf{x}_1^j(R, T)$ *and* $\mathbf{x}_2^j(R, T)$ *which minimize the objective function* ϕ *with respect to a fixed* (R, T) *computed from the previous step.*

4. **Return** *to Step 2 until the decrement in the value of* ϕ *is below a threshold.*

In step 3 above, for a fixed (R, T), $\mathbf{x}_1^j(R,T)$ and $\mathbf{x}_2^j(R,T)$ can be computed by minimizing the distance $\|\mathbf{x}_1^j - \tilde{\mathbf{x}}_1^j\|^2 + \|\mathbf{x}_2^j - \tilde{\mathbf{x}}_2^j\|^2$ for each pair of image points. Let $t_2^j \in \mathbb{R}^3$ be the normal vector (of unit length) to the (epipolar) plane spanned by $(\mathbf{x}_2^j, T)$. Given such a t_2^j, $\mathbf{x}_1^j$ and $\mathbf{x}_2^j$ are determined by:

$$\mathbf{x}_1^j(t_1^j) = \frac{\widehat{e}_3 t_1^j t_1^{jT} \widehat{e}_3^T \tilde{\mathbf{x}}_1^j + \widehat{t_1^j}^T \widehat{t_1^j} e_3}{e_3^T \widehat{t_1^j}^T \widehat{t_1^j} e_3}, \quad \mathbf{x}_2^j(t_2^j) = \frac{\widehat{e}_3 t_2^j t_2^{jT} \widehat{e}_3^T \tilde{\mathbf{x}}_2^j + \widehat{t_2^j}^T \widehat{t_2^j} e_3}{e_3^T \widehat{t_2^j}^T \widehat{t_2^j} e_3},$$

where $t_1^j = R^T t_2^j \in \mathbb{R}^3$. Then the distance can be explicitly expressed as:

$$\|\mathbf{x}_2^j - \tilde{\mathbf{x}}_2^j\|^2 + \|\mathbf{x}_1^j - \tilde{\mathbf{x}}_1^j\|^2 = \|\tilde{\mathbf{x}}_2^j\|^2 + \frac{t_2^{jT} A^j t_2^j}{t_2^{jT} B^j t_2^j} + \|\tilde{\mathbf{x}}_1^j\|^2 + \frac{t_1^{jT} C^j t_1^j}{t_1^{jT} D^j t_1^j},$$

where $A^j, B^j, C^j, D^j \in \mathbb{R}^{3\times 3}$ are defined by:

$$\begin{array}{rcll} A^j &=& I - (\widehat{e}_3 \tilde{\mathbf{x}}_2^j \tilde{\mathbf{x}}_2^{jT} \widehat{e}_3^T + \widehat{\mathbf{x}_2^j} \widehat{e}_3 + \widehat{e}_3 \widehat{\mathbf{x}_2^j}), & B^j = \widehat{e}_3^T \widehat{e}_3 \\ C^j &=& I - (\widehat{e}_3 \tilde{\mathbf{x}}_1^j \tilde{\mathbf{x}}_1^{jT} \widehat{e}_3^T + \widehat{\mathbf{x}_1^j} \widehat{e}_3 + \widehat{e}_3 \widehat{\mathbf{x}_1^j}), & D^j = \widehat{e}_3^T \widehat{e}_3 \end{array}. \tag{5.23}$$

Then the problem of finding $\mathbf{x}_1^j(R,T)$ and $\mathbf{x}_2^j(R,T)$ becomes one of finding t_2^{j*} which minimizes the function of a sum of two *singular Rayleigh quotients*:

$$\min_{t_2^{jT} T = 0,\, t_2^{jT} t_2^j = 1} V(t_2^j) = \frac{t_2^{jT} A^j t_2^j}{t_2^{jT} B^j t_2^j} + \frac{t_2^{jT} R C^j R^T t_2^j}{t_2^{jT} R D^j R^T t_2^j}. \tag{5.24}$$

This is an optimization problem on the unit circle $\mathbb{S}^1$ in the plane orthogonal to the vector T (therefore, geometrically, motion and structure recovery from n pairs of image correspondences is an optimization problem on the space $SO(3) \times \mathbb{S}^2 \times \mathbb{T}^n$ where $\mathbb{T}^n$ is an n-torus, i.e. an n-fold product of $\mathbb{S}^1$). If $N_1, N_2 \in \mathbb{R}^3$ are vectors such that T, N_1, N_2 form an orthonormal basis of $\mathbb{R}^3$, then $t_2^j = \cos(\theta) N_1 + \sin(\theta) N_2$ with $\theta \in \mathbb{R}$. We only need to find θ^* which minimizes the function $V(t_2^j(\theta))$. From the geometric interpretation of the optimal solution, we also know that the global minimum θ^* should lie between two values: θ_1 and θ_2 such that $t_2^j(\theta_1)$ and $t_2^j(\theta_2)$ correspond to normal vectors of the two planes spanned by $(\tilde{\mathbf{x}}_2^j, T)$ and $(R\tilde{\mathbf{x}}_1^j, T)$ respectively (if $\tilde{\mathbf{x}}_1^j, \tilde{\mathbf{x}}_2^j$ are already triangulated, these two planes coincide). The problem now becomes a simple bounded minimization problem for a scalar function and can be efficiently solved using standard optimization routines (such as "fmin" in Matlab or the Newton's algorithm).

5.4 Continuous case

As we pointed out in Section 5.1, the limit case where the two viewpoints are infinitesimally close requires extra attention. From the practical standpoint, this case is relevant to the analysis of a video stream where the camera motion is slow relative to the sampling frequency. In this section, we follow the steps of the previous section by giving a parallel derivation of the geometry of points in space as seen from a moving camera, and deriving a conceptual algorithm for reconstructing camera motion and scene structure. In light of the fact that the camera motion is slow relative to the sampling frequency we will treat the motion of the camera as continuous. While the derivations proceed in parallel it is an important point of caution to the reader that there are some subtle differences.

5.4.1 Continuous epipolar constraint and the continuous essential matrix

Let us assume that camera motion is described by a smooth (i.e. continuously differentiable) trajectory $g(t) = (R(t), T(t)) \in SE(3)$ with body velocities $(\omega(t), v(t)) \in se(3)$ as defined in Chapter 2 For a point $p \in \mathbb{R}^3$, its coordinates as a function of time $\mathbf{X}(t)$ satisfy

$$\dot{\mathbf{X}}(t) = \widehat{\omega}(t)\mathbf{X}(t) + v(t). \tag{5.25}$$

From now on, for convenience, we will drop the time-dependency from the notation. The image of the point p taken by the camera is the vector $\mathbf{x}$ which satisfies $\lambda\mathbf{x} = \mathbf{X}$. Denote the velocity of the image point $\mathbf{x}$ by $\mathbf{u} = \dot{\mathbf{x}} \in \mathbb{R}^3$. $\mathbf{u}$ is also called *image motion field*, which under the brightness constancy assumption discussed in Chapter 3 can be approximated by the *optical flow*.

Consider now the inner product of the vectors in (5.25) with the vector $(v \times \mathbf{x})$

$$\dot{\mathbf{X}}^T(v \times \mathbf{x}) = (\widehat{\omega}\mathbf{X} + v)^T(v \times \mathbf{x}) = \mathbf{X}^T\widehat{\omega}^T\widehat{v}\mathbf{x}. \tag{5.26}$$

Further note that

$$\dot{\mathbf{X}} = \dot{\lambda}\mathbf{x} + \lambda\dot{\mathbf{x}} \quad \text{and} \quad \mathbf{x}^T(v \times \mathbf{x}) = 0.$$

Using this in equation (5.26), we have

$$\lambda\dot{\mathbf{x}}^T\widehat{v}\mathbf{x} - \lambda\mathbf{x}^T\widehat{\omega}^T\widehat{v}\mathbf{x} = 0.$$

When $\lambda \neq 0$, we obtain a constraint that plays the role of the epipolar constraint for the case of continuous-time images, in the sense that it does not depend upon the position of the points in space, but only on their projection and the motion parameters:

$$\mathbf{u}^T\widehat{v}\mathbf{x} + \mathbf{x}^T\widehat{\omega}\widehat{v}\mathbf{x} = 0. \tag{5.27}$$

Before proceeding, we state a lemma that will become useful in the remainder of this section.

Lemma 5.3. *Consider the matrices $M_1, M_2 \in \mathbb{R}^{3\times3}$. Then $\mathbf{x}^TM_1\mathbf{x} = \mathbf{x}^TM_2\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^3$ if and only if $M_1 - M_2$ is a skew symmetric matrix, i.e. $M_1 - M_2 \in so(3)$.*

We leave the proof of this lemma as an exercise. Following the lemma, for any skew symmetric matrix $M \in \mathbb{R}^{3\times3}$, $\mathbf{x}^TM\mathbf{x} = 0$. Since $\frac{1}{2}(\widehat{\omega}\widehat{v} - \widehat{v}\widehat{\omega})$ is a skew symmetric matrix, $\mathbf{x}^T\frac{1}{2}(\widehat{\omega}\widehat{v} - \widehat{v}\widehat{\omega})\mathbf{x} = 0$. If we define the *symmetric epipolar component* to be the following matrix

$$s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega}) \quad \in \mathbb{R}^{3\times3}.$$

then we have

$$\mathbf{u}^T\widehat{v}\mathbf{x} + \mathbf{x}^Ts\mathbf{x} = 0. \tag{5.28}$$

This equation suggests some redundancy in the continuous epipolar constraint (5.27). Indeed, the matrix $\widehat{\omega}\widehat{v}$ can only be recovered up to its symmetric component $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$[3]. This structure is substantially different from the discrete case, and it cannot be derived as a first-order approximation of the essential matrix $\widehat{T}R$. In fact a naive discretization would lead to a matrix of

[3] This redundancy is the very reason why different forms of the continuous epipolar constraint exist in the literature [38, 23, 33, 18, 3], and, accordingly, various approaches have been proposed to recover ω and v (see [26]).

the form $\widehat{v}\widehat{\omega}$, whereas in the true continuous case we have to deal with only its symmetric component $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$! The set of interest in this case is the space of 6×3 matrices of the form

$$\mathcal{E}' = \left\{ \begin{bmatrix} \widehat{v} \\ \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega}) \end{bmatrix} \middle| \omega, v \in \mathbb{R}^3 \right\} \quad \subset \mathbb{R}^{6\times 3}$$

which we call the *continuous essential space*. A matrix in this space is called a *continuous essential matrix*. Note that the continuous epipolar constraint (5.28) is homogeneous in the linear velocity v. Thus v may be recovered only up to a constant scale. Consequently, in motion recovery, we will concern ourselves with matrices belonging to the *normalized continuous essential space* with v normalized to be 1:

$$\mathcal{E}'_1 = \left\{ \begin{bmatrix} \widehat{v} \\ \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega}) \end{bmatrix} \middle| \omega \in \mathbb{R}^3, v \in \mathbb{S}^2 \right\} \quad \subset \mathbb{R}^{6\times 3}.$$

5.4.2 Properties of continuous essential matrices

The skew-symmetric part of a continuous essential matrix simply corresponds to the velocity v. The characterization of the (normalized) essential matrix only focuses on its part $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$. We call the space of all the matrices of this form the *symmetric epipolar space*

$$\mathcal{S} = \left\{ \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega}) \middle| \omega \in \mathbb{R}^3, v \in \mathbb{S}^2 \right\} \subset \mathbb{R}^{3\times 3}.$$

A matrix in this space is called a *symmetric epipolar component*. The motion estimation problem is now reduced to the one of *recovering the velocity* (ω, v) *with* $\omega \in \mathbb{R}^3$ *and* $v \in \mathbb{S}^2$ *from a given symmetric epipolar component* s.

The characterization of symmetric epipolar components depends on a characterization of matrices of the form $\widehat{\omega}\widehat{v} \in \mathbb{R}^{3\times 3}$, which is given in the following lemma. We define the matrix $R_Y(\theta)$ to be the rotation around the Y-axis by an angle $\theta \in \mathbb{R}$, i.e. $R_Y(\theta) = e^{\widehat{e}_2\theta}$ with $e_2 = [0, 1, 0]^T \in \mathbb{R}^3$.

Lemma 5.4. *A matrix* $Q \in \mathbb{R}^{3\times 3}$ *has the form* $Q = \widehat{\omega}\widehat{v}$ *with* $\omega \in \mathbb{R}^3$, $v \in \mathbb{S}^2$ *if and only if*

$$Q = -V R_Y(\theta) diag\{\lambda, \lambda\cos(\theta), 0\} V^T \tag{5.29}$$

for some rotation matrix $V \in SO(3)$. *the positive scalar* $\lambda = \|\omega\|$ *and* $\cos(\theta) = \omega^T v/\lambda$.

Proof. We first prove the necessity. The proof follows from the geometric meaning of $\widehat{\omega}\widehat{v}$ for any vector $q \in \mathbb{R}^3$,

$$\widehat{\omega}\widehat{v}q = \omega \times (v \times q).$$

Let $b \in \mathbb{S}^2$ be the unit vector perpendicular to both ω and v. Then, $b = \frac{v\times\omega}{\|v\times\omega\|}$ (if $v \times \omega = 0$, b is not uniquely defined. In this case, ω, v are parallel and the rest of the proof follows if one picks any vector b orthogonal to v and ω). Then $\omega = \lambda \exp(\widehat{b}\theta)v$ (according this definition, θ is the angle between ω and v, and $0 \le \theta \le \pi$). It is direct to check that if the matrix V is defined to be

$$V = (e^{\widehat{b}\frac{\pi}{2}}v, b, v),$$

then Q has the given form (5.29).

We now prove the sufficiency. Given a matrix Q which can be decomposed into the form (5.29), define the orthogonal matrix $U = -VR_Y(\theta) \in O(3)$.[4] Let the two skew symmetric matrices $\widehat{\omega}$ and $\widehat{v}$ given by the formulae

$$\widehat{\omega} = UR_Z(\pm\frac{\pi}{2})\Sigma_\lambda U^T, \quad \widehat{v} = VR_Z(\pm\frac{\pi}{2})\Sigma_1 V^T \tag{5.30}$$

where $\Sigma_\lambda = diag\{\lambda, \lambda, 0\}$ and $\Sigma_1 = diag\{1, 1, 0\}$. Then

$$\begin{aligned} \widehat{\omega}\widehat{v} &= UR_Z(\pm\frac{\pi}{2})\Sigma_\lambda U^T VR_Z(\pm\frac{\pi}{2})\Sigma_1 V^T \\ &= UR_Z(\pm\frac{\pi}{2})\Sigma_\lambda(-R_Y^T(\theta))R_Z(\pm\frac{\pi}{2})\Sigma_1 V^T \\ &= Udiag\{\lambda, \lambda\cos(\theta), 0\}V^T \\ &= Q. \end{aligned} \tag{5.31}$$

Since ω and v have to be, respectively, the left and the right zero eigenvectors of Q, the reconstruction given in (5.30) is unique. □

The following theorem reveals the structure of the symmetric epipolar component.

Theorem 5.4 (Characterization of the symmetric epipolar component). *A real symmetric matrix $s \in \mathbb{R}^{3\times 3}$ is a symmetric epipolar component if and only if s can be diagonalized as $s = V\Sigma V^T$ with $V \in SO(3)$ and*

$$\Sigma = diag\{\sigma_1, \sigma_2, \sigma_3\}$$

with $\sigma_1 \geq 0, \sigma_3 \leq 0$ and $\sigma_2 = \sigma_1 + \sigma_3$.

Proof. We first prove the necessity. Suppose s is a symmetric epipolar component, there exist $\omega \in \mathbb{R}^3, v \in \mathbb{S}^2$ such that $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$. Since s is a symmetric matrix, it is diagonalizable, all its eigenvalues are real and all the eigenvectors are orthogonal to each other. It then suffices to check that its eigenvalues satisfy the given conditions.

Let the unit vector b and the rotation matrix V be the same as in the proof of Lemma 5.4, so are θ and γ. Then according to the lemma, we have

$$\widehat{\omega}\widehat{v} = -VR_Y(\theta)diag\{\lambda, \lambda\cos(\theta), 0\}V^T.$$

Since $(\widehat{\omega}\widehat{v})^T = \widehat{v}\widehat{\omega}$, it yields

$$s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega}) = \frac{1}{2}V\left(-R_Y(\theta)diag\{\lambda, \lambda\cos(\theta), 0\} - diag\{\lambda, \lambda\cos(\theta), 0\}R_Y^T(\theta)\right)V^T.$$

Define the matrix $D(\lambda, \theta) \in \mathbb{R}^{3\times 3}$ to be

$$\begin{aligned} D(\lambda, \theta) &= -R_Y(\theta)diag\{\lambda, \lambda\cos(\theta), 0\} - diag\{\lambda, \lambda\cos(\theta), 0\}R_Y^T(\theta) \\ &= \lambda \begin{bmatrix} -2\cos(\theta) & 0 & \sin(\theta) \\ 0 & -2\cos(\theta) & 0 \\ \sin(\theta) & 0 & 0 \end{bmatrix}. \end{aligned}$$

[4] $O(3)$ represents the space of all orthogonal matrices (of determinant ± 1.)

Directly calculating its eigenvalues and eigenvectors, we obtain that

$$D(\lambda,\theta) = R_Y\left(\frac{\theta-\pi}{2}\right) diag\left\{\lambda(1-\cos(\theta)), -2\lambda\cos(\theta), \lambda(-1-\cos(\theta))\right\} R_Y^T\left(\frac{\theta-\pi}{2}\right). \quad (5.32)$$

Thus $s = \frac{1}{2}VD(\lambda,\theta)V^T$ has eigenvalues

$$\left\{\frac{1}{2}\lambda(1-\cos(\theta)), \quad -\lambda\cos(\theta), \quad \frac{1}{2}\lambda(-1-\cos(\theta))\right\}, \quad (5.33)$$

which satisfy the given conditions.

We now prove the sufficiency. Given $s = V_1 diag\{\sigma_1,\sigma_2,\sigma_3\}V_1^T$ with $\sigma_1 \geq 0, \sigma_3 \leq 0$ and $\sigma_2 = \sigma_1 + \sigma_3$ and $V_1^T \in SO(3)$, these three eigenvalues uniquely determine $\lambda, \theta \in \mathbb{R}$ such that the σ_i's have the form given in (5.33)

$$\begin{cases} \lambda = \sigma_1 - \sigma_3, & \lambda \geq 0 \\ \theta = \arccos(-\sigma_2/\lambda), & \theta \in [0,\pi] \end{cases}$$

Define a matrix $V \in SO(3)$ to be $V = V_1 R_Y^T\left(\frac{\theta}{2}-\frac{\pi}{2}\right)$. Then $s = \frac{1}{2}VD(\lambda,\theta)V^T$. According to Lemma 5.4, there exist vectors $v \in \mathbb{S}^2$ and $\omega \in \mathbb{R}^3$ such that

$$\widehat{\omega}\widehat{v} = -VR_Y(\theta) diag\{\lambda, \lambda\cos(\theta), 0\}V^T.$$

Therefore, $\frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega}) = \frac{1}{2}VD(\lambda,\theta)V^T = s$. □

Figure 5.3 gives a geometric interpretation of the three eigenvectors of the symmetric epipolar component s for the case when both ω, v are of unit length. Theorem 5.4 was given as an exercise

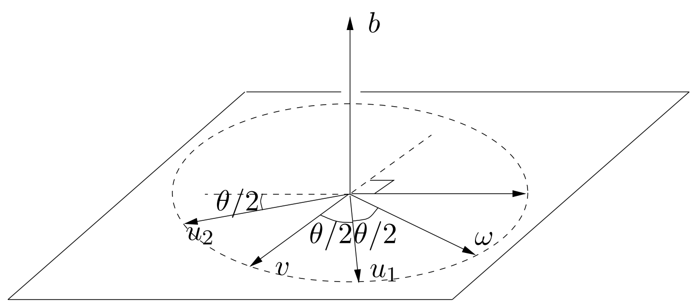

Figure 5.3: Vectors u_1, u_2, b are the three eigenvectors of a symmetric epipolar component $\frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$. In particular, b is the normal vector to the plane spanned by ω and v, and u_1, u_2 are both in this plane. u_1 is the average of ω and v. u_2 is orthogonal to both b and u_1.

problem in Kanatani [13] but it has never been really exploited in the literature for designing algorithms. The constructive proof given above is important in this regard, since it gives an explicit decomposition of the symmetric epipolar component s, which will be studied in more detail next.

Following the proof of Theorem 5.4, if we already know the eigenvector decomposition of a symmetric epipolar component s, we certainly can find at least one solution (ω, v) such that $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$. We now discuss uniqueness, i.e. how many solutions exist for $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$.

Theorem 5.5 (Velocity recovery from the symmetric epipolar component). *There exist exactly four 3-D velocities (ω, v) with $\omega \in \mathbb{R}^3$ and $v \in \mathbb{S}^2$ corresponding to a non-zero $s \in \mathcal{S}$.*

Proof. Suppose (ω_1, v_1) and (ω_2, v_2) are both solutions for $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$. Then we have

$$\widehat{v}_1\widehat{\omega}_1 + \widehat{\omega}_1\widehat{v}_1 = \widehat{v}_2\widehat{\omega}_2 + \widehat{\omega}_2\widehat{v}_2. \tag{5.34}$$

From Lemma 5.4, we may write

$$\left\{ \begin{array}{rcl} \widehat{\omega}_1\widehat{v}_1 & = & -V_1R_Y(\theta_1)diag\{\lambda_1, \lambda_1\cos(\theta_1), 0\}V_1^T \\ \widehat{\omega}_2\widehat{v}_2 & = & -V_2R_Y(\theta_2)diag\{\lambda_2, \lambda_2\cos(\theta_2), 0\}V_2^T. \end{array} \right. \tag{5.35}$$

Let $W = V_1^T V_2 \in SO(3)$, then from (5.34)

$$D(\lambda_1, \theta_1) = WD(\lambda_2, \theta_2)W^T. \tag{5.36}$$

Since both sides of (5.36) have the same eigenvalues, according to (5.32), we have

$$\lambda_1 = \lambda_2, \quad \theta_2 = \theta_1.$$

We can then denote both θ_1 and θ_2 by θ. It is immediate to check that the only possible rotation matrix W which satisfies (5.36) is given by $I_{3\times3}$ or

$$\begin{bmatrix} -\cos(\theta) & 0 & \sin(\theta) \\ 0 & -1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \quad or \quad \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & -1 & 0 \\ -\sin(\theta) & 0 & -\cos(\theta) \end{bmatrix}.$$

From the geometric meaning of V_1 and V_2, all the cases give either $\widehat{\omega}_1\widehat{v}_1 = \widehat{\omega}_2\widehat{v}_2$ or $\widehat{\omega}_1\widehat{v}_1 = \widehat{v}_2\widehat{\omega}_2$. Thus, according to the proof of Lemma 5.4, if (ω, v) is one solution and $\widehat{\omega}\widehat{v} = Udiag\{\lambda, \lambda\cos(\theta), 0\}V^T$, then all the solutions are given by

$$\left\{ \begin{array}{rcl} \widehat{\omega} & = & UR_Z(\pm\frac{\pi}{2})\Sigma_\lambda U^T, \quad \widehat{v} = VR_Z(\pm\frac{\pi}{2})\Sigma_1 V^T; \\ \widehat{\omega} & = & VR_Z(\pm\frac{\pi}{2})\Sigma_\lambda V^T, \quad \widehat{v} = UR_Z(\pm\frac{\pi}{2})\Sigma_1 U^T \end{array} \right. \tag{5.37}$$

where $\Sigma_\lambda = diag\{\lambda, \lambda, 0\}$ and $\Sigma_1 = diag\{1, 1, 0\}$. □

Given a non-zero continuous essential matrix $E \in \mathcal{E}'$, according to (5.37), its symmetric component gives four possible solutions for the 3-D velocity (ω, v). However, in general only one of them has the same linear velocity v as the skew symmetric part of E. Hence, compared to the discrete case where there are two 3-D motions (R, T) associated to an essential matrix, the velocity (ω, v) corresponding to a continuous essential matrix is unique. This is because, in the continuous case, the so-called *twisted-pair ambiguity* which occurs in discrete case and is caused by a 180° rotation of the camera around the translation direction (see Maybank [18]), is now avoided.

5.4.3 The eight-point linear algorithm

Based on the preceding study of the continuous essential matrix, this section describes an algorithm to recover the 3-D velocity of the camera from a set of (possibly noisy) optical flows.

Let $E = \begin{bmatrix} \widehat{v} \\ s \end{bmatrix} \in \mathcal{E}_1'$ with $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$ be the essential matrix associated with the continuous epipolar constraint (5.28). Since the sub-matrix $\widehat{v}$ is skew symmetric and s is symmetric, they have the following form

$$\widehat{v} = \begin{bmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{bmatrix}, \quad s = \begin{bmatrix} s_1 & s_2 & s_3 \\ s_2 & s_4 & s_5 \\ s_3 & s_5 & s_6 \end{bmatrix}. \tag{5.38}$$

Define the (continuous) *essential vector* $\mathbf{e} \in \mathbb{R}^9$ to be

$$\mathbf{e} = [v_1, v_2, v_3, s_1, s_2, s_3, s_4, s_5, s_6]^T. \tag{5.39}$$

Define a vector $\mathbf{a} \in \mathbb{R}^9$ associated to optical flow $(\mathbf{x}, \mathbf{u})$ with $\mathbf{x} = [x, y, z]^T \in \mathbb{R}^3, \mathbf{u} = [u_1, u_2, u_3]^T \in \mathbb{R}^3$ to be[5]

$$\mathbf{a} = [u_3 y - u_2 z, u_1 z - u_3 x, u_2 x - u_1 y, x^2, 2xy, 2xz, y^2, 2yz, z^2]^T. \tag{5.40}$$

The continuous epipolar constraint (5.28) can be then rewritten as

$$\mathbf{a}^T \mathbf{e} = 0.$$

Given a set of (possibly noisy) optical flow vectors $(\mathbf{x}^j, \mathbf{u}^j)$, $j = 1, \ldots, n$ generated by the same motion, define a matrix $A \in \mathbb{R}^{n \times 9}$ associated to these measurements to be

$$A = [\mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^n]^T \tag{5.41}$$

where $\mathbf{a}^j$ are defined for each pair $(\mathbf{x}^j, \mathbf{u}^j)$ using (5.40). In the absence of noise, the essential vector $\mathbf{e}$ has to satisfy

$$A\mathbf{e} = 0. \tag{5.42}$$

In order for this equation to have a unique solution for $\mathbf{e}$, the rank of the matrix A has to be eight. Thus, *for this algorithm, the optical flow vectors of at least eight points are needed to recover the 3-D velocity, i.e.* $n \geq 8$, although the minimum number of optical flows needed is actually 5 (see Maybank [18]).

When the measurements are noisy, there might be no solution of $\mathbf{e}$ for $A\mathbf{e} = 0$. As in the discrete case, one may approximate the solution my minimizing the error function $\|A\mathbf{e}\|^2$.

Since the continuous essential vector $\mathbf{e}$ is recovered from noisy measurements, the symmetric part s of E directly recovered from $\mathbf{e}$ is not necessarily a symmetric epipolar component. Thus one can not directly use the previously derived results for symmetric epipolar components to recover the 3-D velocity. Similarly to what we have done for the discrete case, we can first estimate the symmetric matrix s, and then project it onto the space of symmetric epipolar components.

Theorem 5.6 (Projection to the symmetric epipolar space). *If a real symmetric matrix* $F \in \mathbb{R}^{3\times 3}$ *is diagonalized as* $F = V diag\{\lambda_1, \lambda_2, \lambda_3\} V^T$ *with* $V \in SO(3)$, $\lambda_1 \geq 0, \lambda_3 \leq 0$ *and* $\lambda_1 \geq \lambda_2 \geq \lambda_3$, *then the symmetric epipolar component* $E \in \mathcal{S}$ *which minimizes the error* $\|E - F\|_f^2$ *is given by* $E = V diag\{\sigma_1, \sigma_2, \sigma_2\} V^T$ *with*

$$\sigma_1 = \frac{2\lambda_1 + \lambda_2 - \lambda_3}{3}, \quad \sigma_2 = \frac{\lambda_1 + 2\lambda_2 + \lambda_3}{3}, \quad \sigma_3 = \frac{2\lambda_3 + \lambda_2 - \lambda_1}{3}. \tag{5.43}$$

Proof. Define $\mathcal{S}_\Sigma$ to be the subspace of $\mathcal{S}$ whose elements have the same eigenvalues $\Sigma = diag\{\sigma_1, \sigma_2, \sigma_3\}$. Thus every matrix $E \in \mathcal{S}_\Sigma$ has the form $E = V_1 \Sigma V_1^T$ for some $V_1 \in SO(3)$. To simplify the notation, define $\Sigma_\lambda = diag\{\lambda_1, \lambda_2, \lambda_3\}$. We now prove this theorem by two steps.

Step 1: We prove that the matrix $E \in \mathcal{S}_\Sigma$ which minimizes the error $\|E - F\|_f^2$ is given by $E = V \Sigma V^T$. Since $E \in \mathcal{S}_\Sigma$ has the form $E = V_1 \Sigma V_1^T$, we get

$$\|E - F\|_f^2 = \|V_1 \Sigma V_1^T - V \Sigma_\lambda V^T\|_f^2 = \|\Sigma_\lambda - V^T V_1 \Sigma V_1^T V\|_f^2.$$

[5]For perspective projection, $z = 1$ and $u_3 = 0$ thus the expression for $\mathbf{a}$ can be simplified.

Define $W = V^T V_1 \in SO(3)$ and W has the form

$$W = \begin{bmatrix} w_1 & w_2 & w_3 \\ w_4 & w_5 & w_6 \\ w_7 & w_8 & w_9 \end{bmatrix}. \tag{5.44}$$

Then

$$\begin{aligned} \|E - F\|_f^2 &= \|\Sigma_\lambda - W\Sigma W^T\|_f^2 \\ &= tr(\Sigma_\lambda^2) - 2tr(W\Sigma W^T\Sigma_\lambda) + tr(\Sigma^2). \end{aligned} \tag{5.45}$$

Substituting (5.44) into the second term, and using the fact that $\sigma_2 = \sigma_1 + \sigma_3$ and W is a rotation matrix, we get

$$\begin{aligned} tr(W\Sigma W^T\Sigma_\lambda) &= \sigma_1(\lambda_1(1 - w_3^2) + \lambda_2(1 - w_6^2) + \lambda_3(1 - w_9^2)) \\ &+ \sigma_3(\lambda_1(1 - w_1^2) + \lambda_2(1 - w_4^2) + \lambda_3(1 - w_7^2)). \end{aligned} \tag{5.46}$$

Minimizing $\|E - F\|_f^2$ is equivalent to maximizing $tr(W\Sigma W^T\Sigma_\lambda)$. From (5.46), $tr(W\Sigma W^T\Sigma_\lambda)$ is maximized if and only if $w_3 = w_6 = 0$, $w_9^2 = 1$, $w_4 = w_7 = 0$ and $w_1^2 = 1$. Since W is a rotation matrix, we also have $w_2 = w_8 = 0$ and $w_5^2 = 1$. All possible W give a unique matrix in $\mathcal{S}_\Sigma$ which minimizes $\|E - F\|_f^2$: $E = V\Sigma V^T$.

Step 2: From step one, we only need to minimize the error function over the matrices which have the form $V\Sigma V^T \in \mathcal{S}$. The optimization problem is then converted to one of minimizing the error function

$$\|E - F\|_f^2 = (\lambda_1 - \sigma_1)^2 + (\lambda_2 - \sigma_2)^2 + (\lambda_3 - \sigma_3)^2$$

subject to the constraint

$$\sigma_2 = \sigma_1 + \sigma_3.$$

The formula (5.43) for $\sigma_1, \sigma_2, \sigma_3$ are directly obtained from solving this minimization problem. □

Remark 5.1. *For symmetric matrices which do not satisfy conditions $\lambda_1 \geq 0$ or $\lambda_3 \leq 0$, one may simply choose $\lambda_1' = max(\lambda_1, 0)$ or $\lambda_3' = min(\lambda_3, 0)$.*

Finally, we can outline an eigenvalue decomposition-based algorithm for estimating 3-D velocity from optical flow, which serves as a continuous counterpart of the eight-point algorithm given in Section 5.1.

Algorithm 5.3 (The continuous eight-point algorithm). *For a given set of images and optical flow vectors $(\mathbf{x}^j, \mathbf{u}^j)$, $j = 1, \ldots, n$, this algorithm finds $(\omega, v) \in SE(3)$ which solves*

$$\mathbf{u}^{jT}\widehat{v}\mathbf{x}^j + \mathbf{x}^{jT}\widehat{\omega}\widehat{v}\mathbf{x}^j = 0, \quad j = 1, \ldots, n.$$

1. **Estimate essential vector**
   Define a matrix $A \in \mathbb{R}^{n\times 9}$ whose j^{th} row is constructed from $\mathbf{x}^j$ and $\mathbf{u}^j$ as in (5.40). Use the SVD to find the vector $\mathbf{e} \in \mathbb{R}^9$ such that $A\mathbf{e} = 0$: $A = U_A\Sigma_A V_A^T$ and $\mathbf{e} = V(:, 9)$. Recover the vector $v_0 \in \mathbb{S}^2$ from the first three entries of $\mathbf{e}$ and a symmetric matrix $s \in \mathbb{R}^{3\times 3}$ from the remaining six entries as in (5.39).[6]

[6] In order to guarantee v_0 to be of unit length, one needs to "re-normalize" $\mathbf{e}$, i.e. to multiply $\mathbf{e}$ with a scalar such that the 3-D vector determined by the first three entries of $\mathbf{e}$ is of unit length.

2. **Recover the symmetric epipolar component**
   Find the eigenvalue decomposition of the symmetric matrix s

$$s = V_1 diag\{\lambda_1, \lambda_2, \lambda_3\} V_1^T$$

with $\lambda_1 \geq \lambda_2 \geq \lambda_3$. *Project the symmetric matrix* s *onto the symmetric epipolar space* $\mathcal{S}$. *We then have the new* $s = V_1 diag\{\sigma_1, \sigma_2, \sigma_3\} V_1^T$ *with*

$$\sigma_1 = \frac{2\lambda_1 + \lambda_2 - \lambda_3}{3}, \quad \sigma_2 = \frac{\lambda_1 + 2\lambda_2 + \lambda_3}{3}, \quad \sigma_3 = \frac{2\lambda_3 + \lambda_2 - \lambda_1}{3};$$

3. **Recover velocity from the symmetric epipolar component**
   Define

$$\begin{cases} \lambda &= \sigma_1 - \sigma_3, \quad \lambda \geq 0, \\ \theta &= \arccos(-\sigma_2/\lambda), \quad \theta \in [0, \pi]. \end{cases}$$

Let $V = V_1 R_Y^T\left(\frac{\theta}{2} - \frac{\pi}{2}\right) \in SO(3)$ *and* $U = -V R_Y(\theta) \in O(3)$. *Then the four possible 3-D velocities corresponding to the matrix* s *are given by*

$$\begin{cases} \widehat{\omega} &= U R_Z(\pm\frac{\pi}{2}) \Sigma_\lambda U^T, \quad \widehat{v} = V R_Z(\pm\frac{\pi}{2}) \Sigma_1 V^T \\ \widehat{\omega} &= V R_Z(\pm\frac{\pi}{2}) \Sigma_\lambda V^T, \quad \widehat{v} = U R_Z(\pm\frac{\pi}{2}) \Sigma_1 U^T \end{cases}$$

where $\Sigma_\lambda = diag\{\lambda, \lambda, 0\}$ *and* $\Sigma_1 = diag\{1, 1, 0\}$;

4. **Recover velocity from the continuous essential matrix**
   From the four velocities recovered from the matrix s *in step 3, choose the pair* (ω^*, v^*) *which satisfies*

$$v^{*T} v_0 = \max_i v_i^T v_0.$$

Then the estimated 3-D velocity (ω, v) *with* $\omega \in \mathbb{R}^3$ *and* $v \in \mathbb{S}^2$ *is given by*

$$\omega = \omega^*, \quad v = v_0.$$

Remark 5.2. *Since both* $E, -E \in \mathcal{E}_1'$ *satisfy the same set of continuous epipolar constraints, both* $(\omega, \pm v)$ *are possible solutions for the given set of optical flows. However, as in the discrete case, one can get rid of the ambiguous solution by enforcing the "positive depth constraint".*

In situations where the motion of the camera is partially constrained, the above linear algorithm can be further simplified. The following example illustrates how it can be done.

Example 5.1 (Kinematic model of an aircraft). *This example shows how to utilize the so called nonholonomic constraints (see Murray, Li and Sastry [21]) to simplify the proposed linear motion estimation algorithm in the continuous case. Let* $g(t) \in SE(3)$ *represent the position and orientation of an aircraft relative to the spatial frame, the inputs* $\omega_1, \omega_2, \omega_3 \in \mathbb{R}$ *stand for the rates of the rotation about the axes of the aircraft and* $v_1 \in \mathbb{R}$ *the velocity of the aircraft. Using the standard homogeneous representation for* g *(see Murray, Li and Sastry [21]), the kinematic equations of the aircraft motion are given by*

$$\dot{g} = \begin{bmatrix} 0 & -\omega_3 & \omega_2 & v_1 \\ \omega_3 & 0 & -\omega_1 & 0 \\ -\omega_2 & \omega_1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} g$$

where ω_1 stands for pitch rate, ω_2 for roll rate, ω_3 for yaw rate and v_1 the velocity of the aircraft. Then the 3-D velocity (ω, v) in the continuous epipolar constraint (5.28) has the form $\omega = [\omega_1, \omega_2, \omega_3]^T, v = [v_1, 0, 0]^T$. For the algorithm given above, we here have extra constraints on the symmetric matrix $s = \frac{1}{2}(\widehat{\omega}\widehat{v} + \widehat{v}\widehat{\omega})$: $s_1 = s_5 = 0$ and $s_4 = s_6$. Then there are only four different essential parameters left to determine and we can re-define the essential parameter vector $\mathbf{e} \in \mathbb{R}^4$ to be $\mathbf{e} = [v_1, s_2, s_3, s_4]^T$. Then the measurement vector $\mathbf{a} \in \mathbb{R}^4$ is to be $\mathbf{a} = [u_3 y - u_2 z, 2xy, 2xz, y^2 + z^2]^T$. The continuous epipolar constraint can then be rewritten as

$$\mathbf{a}^T \mathbf{e} = 0.$$

If we define the matrix A from $\mathbf{a}$ as in (5.41), the matrix $A^T A$ is a 4×4 matrix rather than a 9×9 one. For estimating the velocity (ω, v), the dimensions of the problem is then reduced from 9 to 4. In this special case, the minimum number of optical flow measurements needed to guarantee a unique solution of $\mathbf{e}$ is reduced to 3 instead of 8. Furthermore, the symmetric matrix s recovered from $\mathbf{e}$ is automatically in the space $\mathcal{S}$ and the remaining steps of the algorithm can thus be dramatically simplified. From this simplified algorithm, the angular velocity $\omega = [\omega_1, \omega_2, \omega_3]^T$ can be fully recovered from the images. The velocity information can then be used for controlling the aircraft.

As in the discrete case, the linear algorithm proposed above is not optimal since it does not enforce the structure of the parameter space during the minimization. Therefore, the recovered velocity does not necessarily minimize the originally chosen error function $\|A\mathbf{e}(\omega, v)\|^2$ on the space $\mathcal{E}_1'$. Again like in the discrete case, we have to assume that translation is not zero. If the motion is purely rotational, then one can prove that there are infinitely many solutions to the epipolar constraint related equations. We leave this as an exercise to the reader.

5.4.4 Euclidean constraints and structure reconstruction

As in the discrete case, the purpose of exploiting Euclidean constraints is to reconstruct the scales of the motion and structure. From the above linear algorithm, we know we can only recover the linear velocity v up to an arbitrary scale. Without loss of generality, we may assume the velocity of the camera motion is $(\omega, \eta v)$ with $\|v\| = 1$ and $\eta \in \mathbb{R}$. By now, only the scale η is unknown to us. Substituting $\mathbf{X}(t) = \lambda(t)\mathbf{x}(t)$ into the equation

$$\dot{\mathbf{X}}(t) = \widehat{\omega}\mathbf{X}(t) + \eta v(t),$$

we obtain for the image $\mathbf{x}^j$ of each point $p^j \in \mathbb{E}^3, j = 1, \ldots, n$,

$$\dot{\lambda}^j \mathbf{x}^j + \lambda^j \dot{\mathbf{x}}^j = \hat{\omega}(\lambda^j \mathbf{x}^j) + \eta v \quad \Leftrightarrow \quad \dot{\lambda}^j \mathbf{x}^j + \lambda^j(\dot{\mathbf{x}}^j - \hat{\omega}\mathbf{x}^j) - \eta v = 0. \tag{5.47}$$

As one may expect, in the continuous case, the scale information is then encoded in $\lambda, \dot{\lambda}$ for the location of the 3-D point, and $\eta \in \mathbb{R}^+$ for the linear velocity v. Knowing $\mathbf{x}, \dot{\mathbf{x}}, \omega$ and v, these constraints are all linear in $\lambda^j, \dot{\lambda}^j, 1 \le j \le n$ and η. Also, if $\mathbf{x}^j, 1 \le j \le n$ are linearly independent of v, i.e. the feature points do not line up with the direction of translation, it can be shown that these linear constraints are not degenerate hence the unknown scales are determined up to a universal scale. We may then arrange all the unknown scales into a single vector $\vec{\lambda}$

$$\vec{\lambda} = [\lambda^1, \ldots, \lambda^n, \dot{\lambda}^1, \ldots, \dot{\lambda}^n, \eta]^T \in \mathbb{R}^{2n+1}.$$

For n optical flows, $\vec{\lambda}$ is a $2n + 1$ dimensional vector. (5.47) gives $3n$ (scalar) linear equations. The problem of solving $\vec{\lambda}$ from (5.47) is usually over-determined. It is easy to check that in the absence of noise the set of equations given by (5.47) uniquely determines $\vec{\lambda}$ if the configuration is non-critical. We can therefore write all the equations in a matrix form

$$M\vec{\lambda} = 0$$

with $M \in \mathbb{R}^{3n\times(2n+1)}$ a matrix depending on ω, v and $\{(\mathbf{x}^j, \dot{\mathbf{x}}^j)\}_{j=1}^n$. Then, in the presence of noise, the linear least-squares estimate of $\vec{\lambda}$ is simply the eigenvector of $M^T M$ corresponding to the smallest eigenvalue.

Notice that the rate of scales $\{\dot{\lambda}^j\}_{j=1}^n$ are also estimated. Suppose we have done the above recovery for a time interval, say (t_0, t_f), then we have the estimate $\vec{\lambda}(t)$ as a function of time t. But $\vec{\lambda}(t)$ at each time t is only determined up to an arbitrary scale. Hence $\rho(t)\vec{\lambda}(t)$ is also a valid estimation for any positive function $\rho(t)$ defined on (t_0, t_f). However, since $\rho(t)$ is multiplied to both $\lambda(t)$ and $\dot{\lambda}(t)$, their ratio

$$r(t) = \dot{\lambda}(t)/\lambda(t)$$

is independent of the choice of $\rho(t)$. Notice $\frac{d}{dt}(\ln \lambda) = \dot{\lambda}/\lambda$. Let the logarithm of the structural scale λ to be $y = \ln \lambda$. Then a time-consistent estimation $\lambda(t)$ needs to satisfy the following ordinary differential equation, which we call the *dynamic scale ODE*

$$\dot{y}(t) = r(t).$$

Given $y(t_0) = y_0 = \ln(\lambda(t_0))$, solve this ODE and obtain $y(t)$ for $t \in [t_0, t_f]$. Then we can recover a consistent scale $\lambda(t)$ given by

$$\lambda(t) = \exp(y(t)).$$

Hence (structure and motion) scales estimated at different time instances now are all relative to the same scale at time t_0. Therefore, in the continuous case, we are also able to recover all the scales as functions of time up to a universal scale. The reader must be aware that the above scheme is only *conceptual*. In practice, the ratio function $r(t)$ would never be available for all time instances in $[t_0, t_f]$.

Comment 5.1 (Universal scale ambiguity). *In both the discrete and continuous cases, in principle, the proposed schemes can reconstruct both the Euclidean structure and motion up to a universal scale.*

5.5 Summary

The seminal work of Longuet-Higgins [14] on the characterization of the so called *epipolar constraint*, has enabled the decoupling of the structure and motion problems and led to the development of numerous linear and nonlinear algorithms for motion estimation from two views, see [6, 13, 18, 34] for overviews.

5.6 Exercises

1. **Linear equation**
   Solve $x \in \mathbb{R}^n$ from the following equation

$$Ax = b$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. In terms of conditions on the matrix A and vector b, describe: When does a solution exist or not exist, and when is the solution unique or not unique? In case the solution is not unique, describe the whole solution set.

2. **Properties of skew symmetric matrix**
   (a) Prove Lemma 5.1.
   (b) Prove Lemma 5.3.

3. **A rank condition for epipolar constraint**
   Show that
   $$\mathbf{x}_2^T \widehat{T} R \mathbf{x}_1 = 0$$
   if and only if
   $$\text{rank}\,[\widehat{\mathbf{x}_2} R \mathbf{x}_1 \; \widehat{\mathbf{x}_2} T] \leq 1.$$

4. **Rotational motion**
   Assume that camera undergoes pure rotational motion, i.e. it rotates around its center. Let $R \in SO(3)$ be the rotation of the camera and $\omega \in so(3)$ be the angular velocity. Show that, in this case, we have:
   (a) Discrete case: $\mathbf{x}_2^T \widehat{T} R \mathbf{x}_1 \equiv 0, \quad \forall T \in \mathbb{R}^3$;
   (b) Continuous case: $\mathbf{x}^T \widehat{\omega} \widehat{v} \mathbf{x} + \mathbf{u}^T \widehat{v} \mathbf{x} \equiv 0, \quad \forall v \in \mathbb{R}^3$.

5. **Projection to $O(3)$**
   Given an arbitrary 3×3 matrix $M \in \mathbb{R}^{3 \times 3}$ with positive singular values, find the orthogonal matrix $R \in O(3)$ such that the error $\|R - M\|_f^2$ is minimized. Is the solution unique? Note: Here we allow $\det(R) = \pm 1$.

6. **Geometric distance to epipolar line**
   Given two image points $\mathbf{x}_1, \tilde{\mathbf{x}}_2$ with respect to camera frames with their relative motion (R, T), show that the geometric distance d_2 defined in Figure 5.2 is given by the formula
   $$d_2^2 = \frac{(\tilde{\mathbf{x}}_2^T \widehat{T} R \mathbf{x}_1)^2}{\|\widehat{e}_3 \widehat{T} R \mathbf{x}_1\|^2}$$
   where $e_3 = [0, 0, 1]^T \in \mathbb{R}^3$.

7. **Planar scene and homography**
   Suppose we have a set of n fixed coplanar points $\{p^j\}_{j=1}^n \subset \mathbf{P}$, where $\mathbf{P}$ denotes a plane. Figure 5.4 depicts the geometry of the camera frame relative to the plane. Without loss of generality, we further assume that the focal length of the camera is 1. That is, if $\mathbf{x}$ is the image of a point $p \in \mathbf{P}$ with 3-D coordinates $\mathbf{X} = [X_1, X_2, X_3]^T$, then
   $$\mathbf{x} = \mathbf{X}/X_3 = [X_1/X_3, X_2/X_3, 1]^T \in \mathbb{R}^3.$$
   Follow the following steps to establish the so-called *homography* between two images of the plane $\mathbf{P}$:
   (a) Verify the following simple identity
   $$\widehat{\mathbf{x}} \mathbf{X} = 0. \tag{5.48}$$

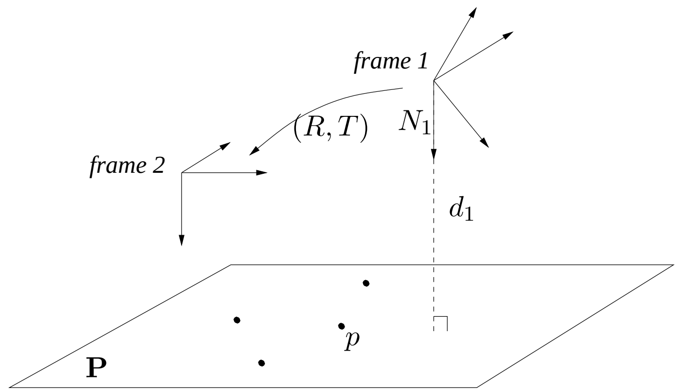

Figure 5.4: Geometry of camera frames 1 and 2 relative to a plane **P**.

(b) Suppose that the $(R, T) \in SE(3)$ is the rigid body transformation from frame 1 to 2. Then the coordinates $\mathbf{X}_1, \mathbf{X}_2$ of a fixed point $p \in \mathbf{P}$ relative to the two camera frames are related by

$$\mathbf{X}_2 = \left(R + \frac{1}{d_1} T N_1^T\right) \mathbf{X}_1 \tag{5.49}$$

where d_1 is the perpendicular distance of camera frame 1 to the plane $\mathbf{P}$ and $N_1 \in \mathbb{S}^2$ is the unit surface normal of $\mathbf{P}$ relative to camera frame 1. In the equation, the matrix $H = (R + \frac{1}{d_1} T N_1^T)$ is the so-called *homography matrix* in the computer vision literature. It represents the transformation from $\mathbf{X}_1$ to $\mathbf{X}_2$.

(c) Use the above identities to show that: Given the two images $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^3$ of a point $p \in \mathbf{P}$ with respect to camera frames 1 and 2 respectively, they satisfy the constraint

$$\widehat{\mathbf{x}_2}(R + \frac{1}{d_1} T N_1^T)\mathbf{x}_1 = 0. \tag{5.50}$$

(d) Prove that in order to solve uniquely (up to a scale) H from the above equation, one needs the (two) images of at least 4 points in $\mathbf{P}$ in general position.

8. **Singular values of the homography matrix**
Prove that any matrix of the form $H = R + uv^T$ with $R \in SO(3)$ and $u, v \in \mathbb{R}^3$ must have a singular value 1. (Hint: prove that the matrix $uv^T + vu^T + vu^T uv^T$ has a zero eigenvalue by constructing the corresponding eigenvector.)

9. **The symmetric component of the outer product of two vectors**
Suppose $u, v \in \mathbb{R}^3$, and $\|u\|^2 = \|v\|^2 = \alpha$. If $u \neq v$, the matrix $D = uv^T + vu^T \in \mathbb{R}^{3\times 3}$ has eigenvalues $\{\lambda_1, 0, \lambda_3\}$, where $\lambda_1 > 0$, and $\lambda_3 < 0$. If $u = \pm v$, the matrix D has eigenvalues $\{\pm 2\alpha, 0, 0\}$.

10. **The continuous homography matrix**
The continuous version of the homography matrix H introduced above is $H' = \widehat{\omega} + \frac{1}{d_1} v N_1^T$ where $\omega, v \in \mathbb{R}^3$ are the angular and linear velocities of the camera respectively. Suppose that you are given a matrix M which is also known to be of the form $M = H' + \lambda I$ for some $\lambda \in \mathbb{R}$. But you are not told the actual value of λ. Prove that you can uniquely recover H' from M and show how.

11. **Implementation of the SVD-based pose estimation algorithm**
Implement a version of the three step pose estimation algorithm for two views. Your MATLAB codes are responsible for:

    - Initialization: Generate a set of N (≥ 8) 3-D points; generate a rigid body motion (R, T) between two camera frames and project (the coordinates of) the points (relative to the camera frame) onto the image plane correctly. Here you may assume the focal length is 1. This step will give you corresponding images as input to the algorithm.
    - Motion Recovery: using the corresponding images and the algorithm to compute the motion $(\tilde{R}, \tilde{T})$ and compare it to the ground truth (R, T).

    After you get the correct answer from the above steps, here are a few suggestions for you to play with the algorithm (or improve it):

    - A more realistic way to generate these 3-D points is to make sure that they are all indeed "in front of" the image plane before and after the camera moves. If a camera has a field of view, how to make sure that all the feature points are shared in the view before and after.
    - Systematically add some noises to the projected images and see how the algorithm responds. Try different camera motions and different layouts of the points in 3-D.
    - Finally, to fail the algorithm, take all the 3-D points from some plane in front of the camera. Run the codes and see what do you get (especially with some noises on the images).

12. **Implementation of the eigenvalue-decomposition based velocity estimation algorithm**
Implement a version of the four step velocity estimation algorithm for optical flows. Your MATLAB codes are responsible for:

    - Initialization: Choose a set of N (≥ 8) 3-D points and a rigid body velocity (ω, v). Correctly obtain the image $\mathbf{x}$ and compute the image velocity $\mathbf{u} = \dot{\mathbf{x}}$ – you need to figure out how to compute $\mathbf{u}$ from (ω, v) and $\mathbf{X}$. Here you may assume the focal length is 1. This step will give you images and their velocities as input to the algorithm.
    - Motion Recovery: Use the algorithm to compute the motion $(\tilde{\omega}, \tilde{v})$ and compare it to the ground truth (ω, v).