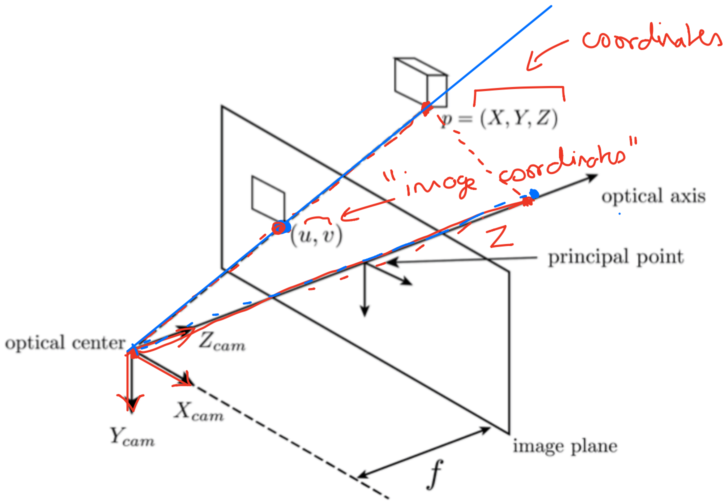
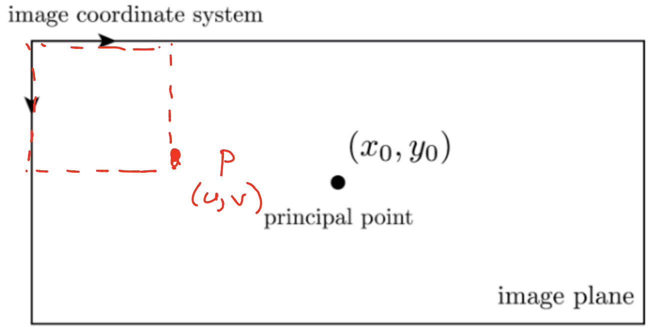
Two-View Geometry

"An Invitation to 3D Vision" chapter 5

1. Pinhole Model Review
2. Two-View reconstruction: Problem Setup.
3. Geometry of two views: epipolar constraint
4. Essential Matrix: Characterization
5. Estimating the essential matrix from point correspondences
6. Recovering pose from an essential matrix
7. Two view reconstruction algorithm:
    - A. Find point correspondences
    - B. Compute essential matrix
    - C. Find candidate poses
    - D. Disambiguate using positive depth constraint
    - E. Triangulate to get locations of visible points

Pinhole Model





"perspective projection",

$x = (u, v, 1)$ ← homogeneous image coordinates of p.

$\bar{X} = (X, Y, Z)$ ← coordinates of p in camera frame

$$x = \frac{1}{Z} K \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \xrightarrow{Z = \lambda} \boxed{\lambda x = K \bar{X}}$$

3x3 matrix "camera matrix" ← focal lengths, skew

When K is known, it is no loss of generality to let $K = I$. K^{-1} exists: $\lambda x = \bar{X}$

Using 1 view, can we recover the location of p given its image coordinates?

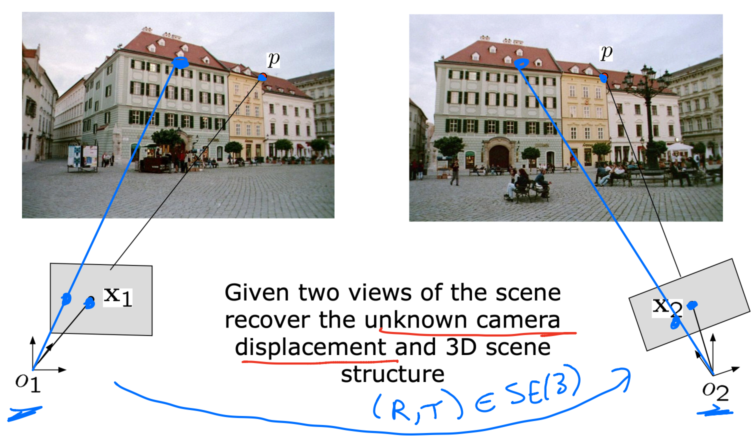
$\lambda x = \bar{X}$

Two-view reconstruction problem setup:

1. Given corresponding points ✓
2. Find camera motion ←
3. Find scene structure (coordinates of each visible point)

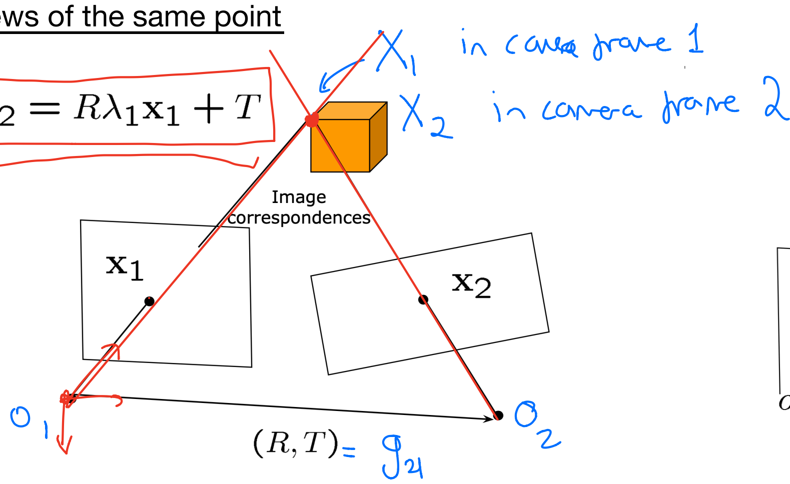
↑ location of P.

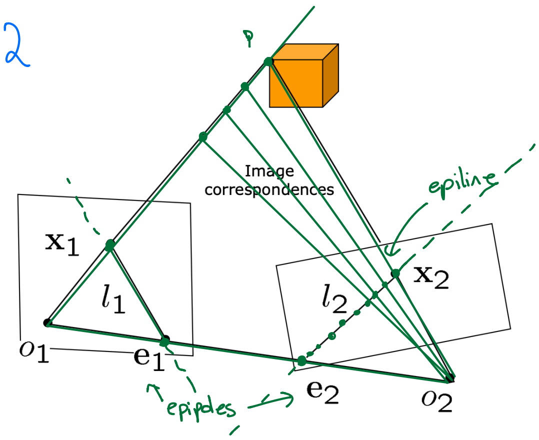
Aux: how many points do we need?



Two views of the same point

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





Aside: If we know (R, T) (i.e. camera motion is known) then we can find location of point using *Triangulation.*

$$\left.\begin{aligned} X_2 &= RX_1 + T \\ \lambda_1 x_1 &= X_1, \quad \lambda_2 x_2 = X_2 \end{aligned}\right\}$$

$$\left.\begin{aligned} &\lambda_2 x_2 = R\lambda_1 x_1 + T \\ &\text{is this enough to find } X_1 ? \end{aligned}\right\}$$

$$\lambda_1 Rx_1 - \lambda_2 x_2 = -T$$

$$\underbrace{\begin{bmatrix} Rx_1 & -x_2 \end{bmatrix}}_{3\times 2} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = -T$$

$$Ax = b$$

In general, what can we say about the relationship between corresponding points?

1. Epipoles
2. Epipolar plane
3. Epilines

$$\lambda_2 x_2 = R\lambda_1 x_1 + T$$

$$\underbrace{x_2^T \hat{T} \lambda_2 x_2}_{0} = x_2^T \hat{T} R \lambda_1 x_1 + \underbrace{\cancel{x_2^T \hat{T} T}}_{0}$$

$$\cancel{\lambda_2 x_2^T (\underbrace{\hat{T} x_2}_{T \times x_2})}$$

$$\boxed{x_2^T \hat{T} R x_1 = 0}$$

Epipolar constraint.

$$\hat{T} R := E$$

"essential matrix" for this system.

Theorem 5.1 (Epipolar constraint). *Consider two images $\boldsymbol{x}_1, \boldsymbol{x}_2$ of the same point p from two camera positions with relative pose (R, T), where $R \in SO(3)$ is the relative orientation and $T \in \mathbb{R}^3$ is the relative position. Then $\boldsymbol{x}_1, \boldsymbol{x}_2$ satisfy*

$$\langle \boldsymbol{x}_2, T \times R\boldsymbol{x}_1 \rangle = 0, \quad \text{or} \quad \boxed{\boldsymbol{x}_2^T \widehat{T} R \boldsymbol{x}_1 = 0.} \tag{5.2}$$

The matrix

$$E \doteq \widehat{T} R \quad \in \mathbb{R}^{3\times 3}$$

in the epipolar constraint equation (5.2) is called the *essential matrix.*

$$E = \hat{T} R$$

Essential Matrices

1. Encode the relative transform (R, T) between two frames.
2. Provide a constraint that corresponding points *must* satisfy.

The matrix $E = \widehat{T}R \in \mathbb{R}^{3\times 3}$ in equation (5.2) 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 special set of matrices in $\mathbb{R}^{3\times 3}$ called the *essential space* and denoted by $\mathcal{E}$:

$\hat{T} \in so(3)$

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

Aside: Singular Value Decomposition (for square matrices)

Theorem: Let $A \in \mathbb{R}^{n\times n}$. Then we can write

$\leftarrow U^TU = I, \det(U) = \pm 1$ $\quad A = U\Sigma V^T$

$\Sigma = \text{diag}(\sigma_1 \cdots \sigma_n)$

$\sigma_i \geq 0$

where U, V are orthogonal $n \times n$ matrices and Σ is a diagonal matrix with non-negative entries. The diagonal entries of Σ are called the *singular values* of A and equal the square roots of the eigenvalues of A^TA.

σ_i

Any essential matrix can be written as $E = \hat{T}R$.

But what if I'm given $E \in \mathbb{R}^{3\times 3}$. How can I tell if it is an essential matrix?

Theorem 5.5 (Characterization of the essential matrix). *A nonzero 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*

$$\Sigma = \text{diag}\{\sigma, \sigma, 0\} \quad \leftarrow \quad (1) \quad \boxed{\sigma_1 = \sigma_2 = \sigma, \quad \sigma_3 = 0}$$

for some $\sigma \in \mathbb{R}_+$ and $U, V \in SO(3)$. $\leftarrow$ (2) $U, V \in SO(3)$: $\det(U) = \det(V) = 1$.

Estimating the essential matrix from point correspondences

$(x_1^{(1)}, x_2^{(1)}), \ (x_1^{(2)}, x_2^{(2)}) \ \cdots \ (x_1^{(n)}, x_2^{(n)})$

→ Can I compute E for this system?

for $i = 1 \ldots n$:

$$\left\{ \begin{array}{l} x_2^{(i)T} cE \, x_1^{(i)} = 0 \\ \vdots \\ x_2^{(n)T} cE \, x_1^{(n)} = 0 \end{array} \right\} \quad \begin{array}{l} E \in \bar{\mathcal{E}} \leftarrow \text{essential space.} \\ E = \hat{T}R \end{array} \quad - \quad \begin{array}{l} \tilde{E} = \hat{T}R \\ c\tilde{E} = \underbrace{(\widehat{cT})}\, R \end{array}$$

① If $\hat{E}$ is a solution to this system, then so is $c\tilde{E}$. for any scalar c.

We can only solve for E up to a single scale factor.
So, we restrict ourselves to finding an E = \hat T R such that ||T|| = 1

We try to find $E = \hat{T}R$ where $\|T\| = 1$.

Normalized Essential Space

$$\bar{\mathcal{E}} = \{ \hat{T}R \mid R \in SO(3), T \in \mathbb{R}^3, \|T\| = 1 \}$$

Solving for E using a linear algorithm

$x_2^T E x_1$

$= \sum_{i,j} E_{ij} (x_2)_i (x_1)_j$

$= [(x_2)_1 (x_1)_1 \; \cdots \; (x_2)_3 (x_1)_3] \begin{bmatrix} E_{11} \\ \vdots \\ E_{33} \end{bmatrix}$

$v^T E u = \sum_{i,j} \underbrace{v_i u_j} E_{ij}$

Let $E = \widehat{T}R$ be the essential matrix associated with the epipolar constraint (5.2). The entries of this 3×3 matrix are denoted by

$$E = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \in \mathbb{R}^{3\times 3} \tag{5.10}$$

← unknown to us.

and stacked into a vector $E^s \in \mathbb{R}^9$, which is typically referred to as the *stacked version* of the matrix E (Appendix A.1.3):

$$\underline{E^s \doteq [e_{11}, e_{21}, e_{31}, e_{12}, e_{22}, e_{32}, e_{13}, e_{23}, e_{33}]^T} \quad \in \mathbb{R}^9.$$

known

Or, more specifically, if $\boldsymbol{x}_1 = [x_1, y_1, z_1]^T \in \mathbb{R}^3$ and $\boldsymbol{x}_2 = [x_2, y_2, z_2]^T \in \mathbb{R}^3$, then

Known

$$\boldsymbol{a} = [x_1x_2, x_1y_2, x_1z_2, y_1x_2, y_1y_2, y_1z_2, z_1x_2, z_1y_2, z_1z_2]^T \quad \in \mathbb{R}^9. \tag{5.12}$$

Since the epipolar constraint $\boldsymbol{x}_2^T E \boldsymbol{x}_1 = 0$ is linear in the entries of E, using the above notation we can rewrite it as the inner product of $\boldsymbol{a}$ and E^s:

$$\boxed{\boldsymbol{a}^T E^s = 0.}$$ } one equation for every pair correspondence.

Now, we can stack the vectors ***a*** *that we get from each corresponding point into a matrix* ***X****. Then the set of epipolar constraints from each of the corresponding pairs can be written together as*

In the absence of noise, the vector E^s satisfies

$$\chi E^s = 0. \tag{5.14}$$

$$\underbrace{\begin{bmatrix} - a_1^T - \\ \vdots \\ - a_n^T - \end{bmatrix}}_{9 \text{ cols}} \begin{bmatrix} e_1 \\ \vdots \\ e_{33} \end{bmatrix} = 0$$

linear equation $Ax = 0$ for E.

$$\underbrace{A}_{\text{Known}} \underbrace{x'}_{\text{unknown}} = 0$$

x nonzero: → A has x in its nullspace ($Ax = 0$)

→ A cannot be full rank.

when is x unique up to scale?

⟶ A has a one-dimensional nullspace.

How many unknowns? How many constraints? How many equations do we need?

→ want $\overset{x}{\underline{X}}$ to have rank = 8.

(1) 9 unknowns (2) 1 constraint for the scale degree of freedom. $\|T\| = 1$.

(3) 8 points are enough.

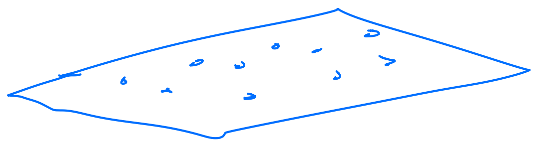
To get a unique solution for Es above (up to scale) we require that the rank of **X** be exactly 8. Then, Es is the unique nonzero direction in the null space of **X**.

$$\begin{bmatrix} a_1^T \\ \vdots \\ a_8^T \end{bmatrix} \leftarrow \text{rank } 8.$$

So, 8 points in "general position" are enough to solve for Es.

If the rank of **X** falls below 8, then the points we are using are not in "general position". They are in some degenerate configuration. For example, this may happen when all points lie on some simple surface, like a plane.

This is called the Eight-Point Algorithm



for estimating E

$$E_s = [e_1 \cdots e_9]$$

$$\tilde{E} = \begin{bmatrix} e_1 & & \\ e_2 & \ddots & \\ e_3 & & e_9 \end{bmatrix}$$

Is $\tilde{E}$ an essential matrix?

(1) find the "closest" essential matrix E to $\tilde{E}$.

"Project" $\tilde{E}$ onto $\overline{\mathcal{E}}$

↑ normalized essential space.

Once we have the vector Es, we can unstack it into a 3x3 matrix. However, this matrix may not be an essential matrix! So, we need to "project" it onto the *Essential Space*, which we described above.

$\sigma = \frac{\sigma_1+\sigma_2}{2}$

onto $\mathcal{E}$

Theorem 2a (Project to Essential Manifold)
If the SVD of a matrix $F \in \mathcal{R}^{3\times3}$ is given by $F = Udiag(\sigma_1, \sigma_2, \sigma_3)V^T$ then the essential matrix E which minimizes the Frobenius distance $\|E - F\|_f^2$ is given by $E = Udiag(\sigma, \sigma, 0)V^T$ with $\sigma = \frac{\sigma_1+\sigma_2}{2}$

To project onto the *normalized* essential space, we can set sigma = 1.

onto $\bar{\mathcal{E}}$

$\sigma = 1$

$$E = U \begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & 0 \end{bmatrix} V^T$$

How do we know U, V have det = 1.

both E and −E satisfy our equation

one of these will have det(U) = det(V) = 1.

Pose Reconstruction from the essential matrix

Can we solve for (R, T) given an essential matrix E?

Recall that we found a normalized E, so we can only solve for the translation up to a scale factor. In particular, we found E such that ||T|| = 1.

Let such a *normalized* E be given. Then the singular values of E are {1, 1, 0} as we stated above. We have the following theorem

differ by a screw about Z.

Theorem 1a (Pose Recovery)
There are two relative poses (R, T) with $T \in \mathcal{R}^3$ and $R \in SO(3)$ corresponding to a non-zero matrix essential matrix.

$$E = U\Sigma V^T$$

g_{21}

g'_{21}

$$\begin{aligned}(\hat{T}_1, R_1) &= (\overset{\hat{T}}{\boxed{UR_Z(+\tfrac{\pi}{2})\Sigma U^T}}, \overset{R}{\boxed{UR_Z^T(+\tfrac{\pi}{2})V^T}})\\ (\hat{T}_2, R_2) &= (\boxed{UR_Z(-\tfrac{\pi}{2})\Sigma U^T}, \boxed{UR_Z^T(-\tfrac{\pi}{2})V^T})\end{aligned}$$

different solutions?

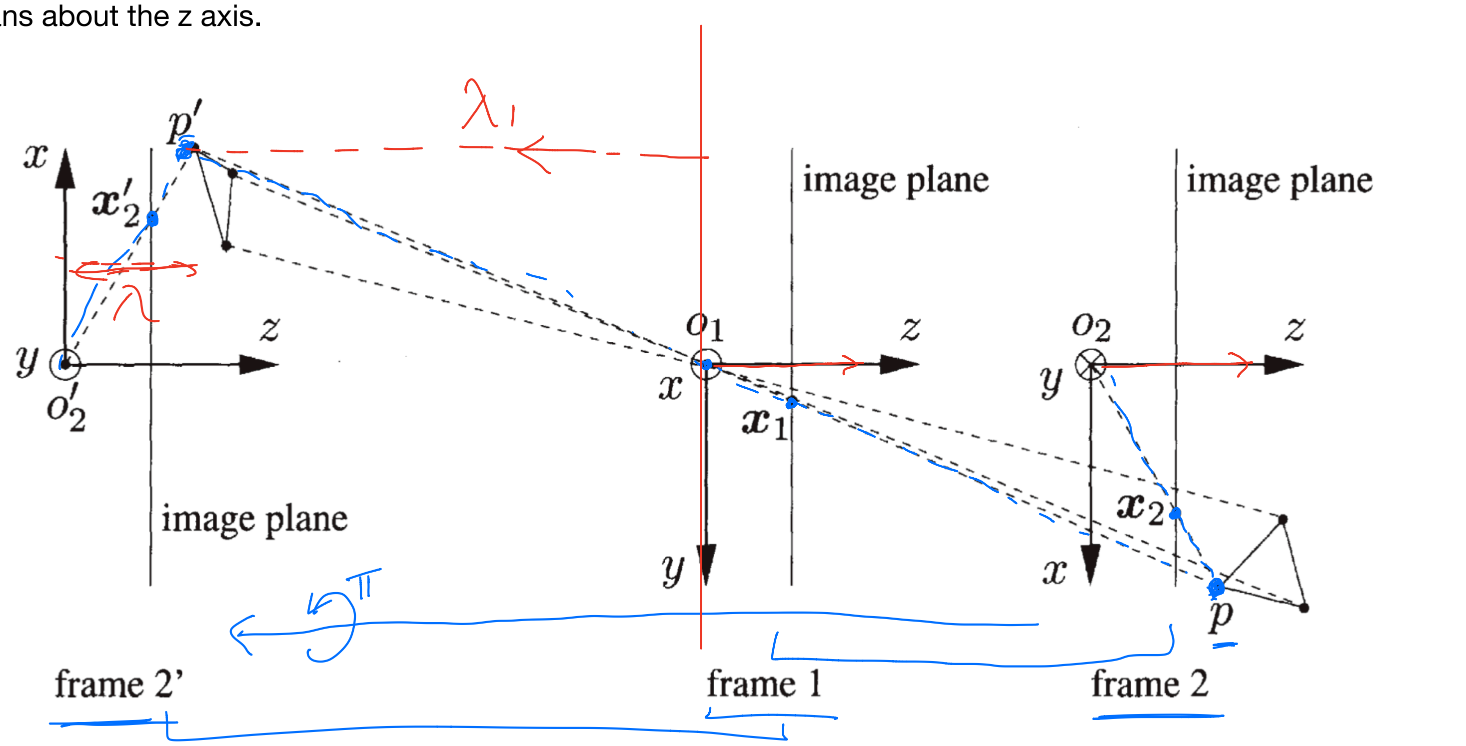
$$\Sigma = diag([1, 1, 0]) \qquad R_z(+\tfrac{\pi}{2}) = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$(1, 2)$

$(1, 2')$

Twisted pair ambiguity

The two solutions we found above correspond to two different possible pairs of camera frames (1, 2) and (1, 2'). The ambiguity between the two poses is one of a screw motion: a translation about the z axis, and a rotation by pi radians about the z axis.



But of course, the true point cannot be *behind* either camera. So, only one of these solutions is physically possible: the one that assigns positive depth to both frames!

To check which is correct, use triangulation with both poses to find the depths in both frames. Pick the one that assigns positive depths to both views.

given $\underbrace{R, T}, x_1, x_2 \longrightarrow \lambda_1, \lambda_2$

↑ both should be positive.

Sign ambiguity

The final ambiguity we must deal with is the sign of E. Note that even with the constraint that ||T|| = 1, both E and -E are possible solutions.

1. There are two possible solutions for (R, T) from E, as shown above.
2. Likewise, two possible solutions from -E.

Only one of these 4 solutions is physically possible.

→ all other 3 will give negative depths for some (point, frame) pair.

Full 8-Point Linear Algorithm

Algorithm 5.1 (The eight-point algorithm).

For a given set of image correspondences $(\boldsymbol{x}_1^j, \boldsymbol{x}_2^j)$, $j = 1, 2, \ldots, n$ $(n \geq 8)$, this algorithm recovers $(R, T) \in SE(3)$, which satisfy

$$\boxed{\boldsymbol{x}_2^{jT} \widehat{T} R \boldsymbol{x}_1^j = 0, \quad j = 1, 2, \ldots, n.}$$

1. **Compute a first approximation of the essential matrix**
Construct $\chi = [\boldsymbol{a}^1, \boldsymbol{a}^2, \ldots, \boldsymbol{a}^n]^T \in \mathbb{R}^{n \times 9}$ from correspondences $\boldsymbol{x}_1^j$ and $\boldsymbol{x}_2^j$ as in (5.12), namely,

$$\boldsymbol{a}^j = \boldsymbol{x}_1^j \otimes \boldsymbol{x}_2^j \quad \in \mathbb{R}^9.$$

$$\begin{bmatrix} a_1^T \\ \vdots \\ a_n^T \end{bmatrix} \begin{bmatrix} e_1 \\ \vdots \\ e_9 \end{bmatrix} = 0$$

Find the vector $E^s \in \mathbb{R}^9$ of unit length such that $\|\chi E^s\|$ is minimized as follows: compute the SVD of $\chi = U_\chi \Sigma_\chi V_\chi^T$ and define E^s to be the ninth column of V_χ. Unstack the nine elements of E^s into a square 3×3 matrix E as in (5.10). Note that this matrix will in general *not* be in the essential space.

2. **Project onto the essential space**
Compute the singular value decomposition of the matrix E recovered from data to be

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

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 $\sigma_3 \neq 0$. But its projection onto the normalized essential space is $U \Sigma V^T$, where $\Sigma = \mathrm{diag}\{1, 1, 0\}$.

project onto $\mathcal{E}$

$\|T\| = 1$

3. **Recover the displacement from the essential matrix**
We now need only U and V to extract R and T from the essential matrix as

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

where $R_Z^T \left(\pm \frac{\pi}{2}\right) \doteq \begin{bmatrix} 0 & \pm 1 & 0 \\ \mp 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$

By considering both E and -E, we will get 4 solutions for the pair (R, T). We can eliminate 3 of them by imposing the positive depth constraint.

Only get $\|T\| = 1$

true $\tilde{T} = \gamma T$.

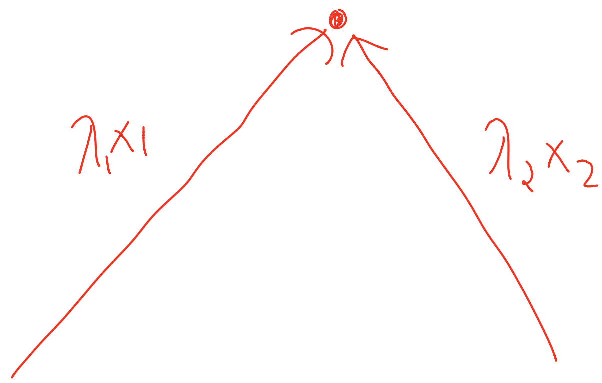
Recovering Structure

Finally, we can use our computed solution for (R, T) to triangulate and find the location of the visible points.

$$\lambda_2 x_2 = R \lambda_1 x_1 + \gamma T$$

redundant.

$\gamma = 1$



multiply both sides by $\hat{x}_2$.

$$\lambda_2 \overset{0}{\cancel{\hat{x}_2 x_2}} = \lambda_1 \hat{x}_2 R x_1 + \gamma \hat{x}_2 T$$

$$\lambda_1 \hat{x}_2 R x_1 + \gamma \hat{x}_2 T = 0$$

$$\begin{bmatrix} \hat{x}_2 R x_1 & \hat{x}_2 T \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \gamma \end{bmatrix} = 0$$

only up to scale.

$\rightarrow$ always some universal scale factor.

$C(\lambda_1 \lambda_2 \lambda_3)$