First Step: Solve for Translation

(1)

- This is trivial, because we can pick a simple origin.
  - World origin is arbitrary.
  - Example: We can assume first point is at origin.
    - Rotation then doesn't effect that point.
    - All its motion is translation.
  - Better to pick center of mass as origin.
    - Average of all points.
    - This also averages all noise.

Specifically, we can never tell where the world points were to begin with. Adding one to every x coordinate in P and then subtracting 1 in every tx is undetectable.

So, wlog we can assume that \( \text{sum}(P(k,:)) = 0 \) for \( k \) from 1 to 3, i.e., \( \text{sum}(x1 \ldots xn) = 0 \), \( \text{sum}(y1 \ldots yn) = 0 \), \( \text{sum}(z1 \ldots zn) = 0 \).

Rotation doesn't move the origin, which is now the center of mass. Neither does scaled orthographic projection. So, this only moves from translation.

Explicitly, we assume \( \text{sum}(p) = (0,0,0)^\top \). Then:

\[
\text{sum}(sR(p)) = sR(\text{sum}(p)) = sR((0,0,0)^\top) = (0,0,0)^\top. 
\]

\( ^\top \) means transpose.

More explicitly, suppose \( \text{sum}(p) = (0,0,0,n)^\top \). Then,

\[
\text{sum}(R^\top P) = R^\top(\text{sum}(P)) = R^\top((0,0,0,n)^\top) = (0,0,0,n)^\top. 
\]

\( P = ((s_{11},s_{12},s_{13})*P \). \( I = s \) part of matrix + \( t \) part of matrix.

Even more explicitly. Consider the first row of the image matrix \( I \). Average together all the entries in this row. This gives us:

\[
\text{sum}(s11,s12,s13)*N(x_i,y_i,z_i) + ex/n \\
= (s11,s12,s13)*N(x_i,y_i,z_i) + n \cdot tx \\
= (s11,s12,s13)*(0,0,0) + tx = tx.
\]

So we've solved for \( tx \). If we subtract \( tx \) from every element in the first row of \( I \), we remove the effects of translation.
First Step: Solve for Translation
(3)
\[
\begin{bmatrix}
\vec{u} \\
\vec{v} \\
\end{bmatrix}
= \begin{bmatrix}
\vec{u}' \\
\vec{v}'
\end{bmatrix}
\]
As if by magic, there's no translation.

Solve for S
- SVD is made to do this.
  \[
  \hat{\mathbf{I}} = \mathbf{U} \mathbf{D} \mathbf{V}^T
  \]
  
  D is diagonal with non-increasing values.
  
  U and V have orthonormal rows.
  
  Ignoring values that get set to 0, we have \( \mathbf{U}(:,1:3) \) for S, and
  
  \( \mathbf{D}(1:3,1:3) \mathbf{V}(1:3,:) \) for P.

Noise
- \( \hat{\mathbf{I}} \) has full rank.
- Best solution is to estimate \( \mathbf{I} \) that's as near to \( \hat{\mathbf{I}} \) as possible, with estimate of \( \mathbf{I} \) having rank 3.
- Our current method does this.

Rank Theorem
\[
\hat{\mathbf{I}} \text{ has rank 3.}
\]
This means there are 3 vectors such that every row of \( \hat{\mathbf{I}} \) is a linear combination of these vectors.
These vectors are the rows of P.

Linear Ambiguity (as before)
\[
\hat{\mathbf{I}} = \mathbf{U}(:,1:3) \ast \mathbf{D}(1:3,1:3) \ast \mathbf{V}(1:3,:)
\]
\[
\hat{\mathbf{I}} = (\mathbf{U}(:,1:3) \ast \mathbf{A}) \ast (\mathbf{inv}(\mathbf{A}) \ast \mathbf{D}(1:3,1:3) \ast \mathbf{V}(1:3,:))
\]

Weak Perspective Motion
Row 2k and 2k+1 of S should be orthogonal.
All rows should be unit vectors.
(Push all scale into P).
Choose A so \((\mathbf{U}(:,1:3) \ast \mathbf{A})\) satisfies these conditions.
Related problems we won’t cover

• Missing data.
• Points with different, known noise.
• Multiple moving objects.

Final Messages

• Structure-from-motion for points can be reduced to linear algebra.
• Epipolar constraint reemerges.
• SVD important.
• Rank Theorem says the images a scene produces aren’t complicated (also important for recognition).