Interlude: Mathematical Tools

Goals:

- Refresh and acquire new knowledge
- Apply mathematical tools to Computer Vision
- Terminology
- Connection Computer Vision-Projective Geometry

We will learn:

- SVD - crucial to solving many linear system of equations in many vision problem.
- homogeneous coordinate – a key component of 3D vision, 3D graphics, robotics. Projective geometry in many cases simpler than Euclidean geometry.

Singular Value Decomposition

- Any *m* x *n* matrix A can be written as the product of three matrices:

$$A = UDV^T$$

$$= \underbrace{\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \dots & \mathbf{u}_m \end{bmatrix}}_{\text{mxm}} \underbrace{\begin{bmatrix} D_0 & 0 \\ 0 & 0 \end{bmatrix}}_{\text{mxn}} \underbrace{\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n \end{bmatrix}^T}_{\text{nxn}}$$

Vector u_i is eigenvector of AA^T

Vector v_i is eigenvector of A^TA

$$D_0 = \mathrm{diag}(\sigma_1, \sigma_2, \dots \sigma_r)$$

$$\sigma_1 \geq \sigma_2 \geq \dots \sigma_r > 0$$

σ_i^2 are the nonvanishing eigenvalues of A^TA or AA^T

σ_i are the nonvanishing singular values of A

Singular Value Decomposition

- Singular values tell you how close a nxn square matrix A is to be singular
  - A is nonsingular iff $D=D_0$, i.e., all its singular values are non-zero.
  - The ratio $$C = \frac{\sigma_1}{\sigma_n}$$ called condition number, measures the degree of singularity of A. If C is large, the matrix is said to be ill-conditioned.
- The number of nonzero σ_i equals the rank of A.
- The pseudoinverse of A, A^+, (irrespective of A being singular or not) is defined as

$$A^+ = V \overset{nxm}{\begin{bmatrix} D_0^{-1} & 0 \\ 0 & 0 \end{bmatrix}} U^T$$

If A is nonsingular, then $D_0^{-1}=D^{-1}$ and $A^+=A^{-1}$.

Solution of linear equations Ax=b (b $\neq$ 0)

- Consider the following linear system

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 8 \\ 10x_1 + 18x_2 + 12x_3 &= 78 \\ 20x_1 + 22x_2 + 40x_3 &= 144 \end{aligned}$$

This problem can be expressed as Ax = b. Write down A & b.

- More equations than unknown (m>n)? find a solution where Ax is approximately equal to b, i.e. we minimize $||Ax-b||^2$

  - Least Square Solution

Solution of linear equations Ax=b (b $\neq$ 0)

- Given a system of mxn linear equations $A\boldsymbol{x} = \boldsymbol{b}$, of rank n (maximal column rank), the least square solution is $\boldsymbol{x} = A^+ \boldsymbol{b}$.

$$A^+ = V \begin{bmatrix} D_0^{-1} & 0 \\ 0 & 0 \end{bmatrix} U^T$$

- If the columns of A are not linearly independent, we have a deficient-rank system (column rank < n)
  - Solution is not unique. $\boldsymbol{x} = A^+ \boldsymbol{b}$ is one of the solutions that **also** minimize || $\boldsymbol{x}$ ||.
  - This solution has certain nice numerical properties. (Optional: Hartley & Zisserman, p559-560)

Solution of linear equations Ax=b (b $\neq$ 0)

- The linear LS problem can also be solved by a method involving the so-called normal equations

$$A^T A\boldsymbol{x} = A^T\boldsymbol{b}$$

- If A is of full column rank (rank n), A^TA is invertible.
  - solution may be found by $\boldsymbol{x} = (A^TA)^{-1}A^T\boldsymbol{b}$.
- This implies $A^+ = (A^TA)^{-1}A^T$ in the case of A having maximal rank.
- Good idea to compute the pseudoinverse of A through SVD, rather than through $(A^TA)^{-1}A^T$ – in case A is near an rank-deficient system.
- *See an example of solving a LS system using SVD, normal equation, QR decomposition on the course website (MATLAB code provided).*

Solution of linear equations (b = 0)

- What happens when b=0? Find x that minimizes $||Ax||$.
  - Known as a homogeneous system where mxn A has rank n-1.
  - An example: find the best straight line through a few points
- The obvious solution of x=0 is not of interest.
- If x is a solution,so is kx for any scalar k. A reasonable constraint would be to seek a solution for which || x ||=1.
- Solution: Choose x to be the eigenvector associated with the smallest eigenvalue of A^TA – in the ideal case, this would be the only zero eigenvalue of A^TA . (For derivation and application, see the chapter on motion).

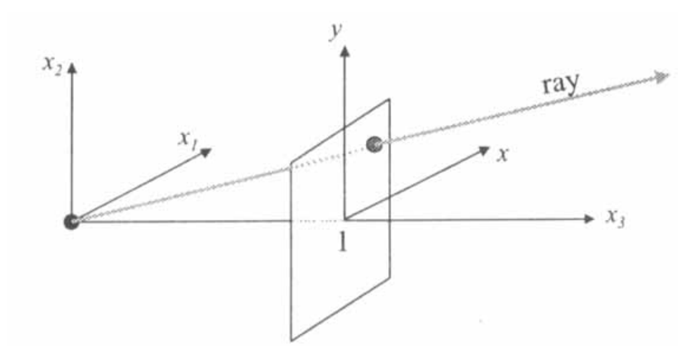
Smallest eigenvalue of A^TA

Solution for x

$$SVD(A) = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \ldots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} D_0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \ldots & \mathbf{v}_n \end{bmatrix}^T$$

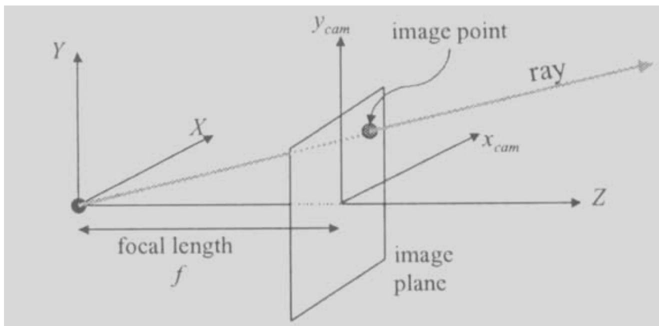
Homogeneous Coordinates

- Consider the 2D case
  - In normal inhomogeneous coordinates a 2D point is represented by a 2-vector $(x,y)^T$.
  - In homogeneous coordinates a 2D point is represented by a 3-vector $(x_1, x_2, x_3)^T$.



Homogeneous Coordinates

- This is the same as the pinhole camera model!
- Only here we scale the 3rd component to the focal length f.



Homogeneous Coordinates

Formally:

- A point in $\mathcal{P}^n$ is represented by the (n+1)-tuple

  $x = (x_0, x_1, \ldots, x_n)$

  where two point x and y are considered to be equal if there exists a non-zero k, such that

  $x = k\,y$
- E.g. in $\mathcal{P}^1$ the following holds:

  (1,2) = (2,4) = (p,2p) ¸ . . .

  (3,5) = (4.5,7.5) = (300,500) ¸ . . .
- Any representation of x in $\mathcal{P}^n$ will be called homogeneous coordinates for x.
- To obtain the inhomogeneous coordinates, divide each of n-tuple by the last number.

Homogeneous Coordinates

- Normally it doesn't make sense to talk about a point at infinity. With homogeneous coordinates we can!
- We can also say in which direction this distant point lies.

$$\tilde{p} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 0 \end{bmatrix}$$

Zero here implies the point is located at an infinite distance away

- Even parallel lines meet.

Homogeneous Coordinates

- Consider 2 parallel lines on a plane:

  Inhomogeneous (Euclidean) coord:

  ax+by+c =0

  ax+by+c' =0

  Homogeneous (Projective) coord:

  ax+by+cz =0

  ax+by+c'z =0

- Solve for (x,y,z). Solution is given by k(1, -a/b, 0).
  - This is the point at infinity. No need to have special case for parallel lines in this more general (projective) geometry!
- In fact, the intersection of two lines $\boldsymbol{l}$ and $\boldsymbol{l'}$ can be obtained by $\boldsymbol{x} = \boldsymbol{l} \times \boldsymbol{l'}$. The simplicity of this expression is again a consequence of the use of homogeneous representation.

Homogeneous Coordinates

- Another example:
  - Consider the simple problem of determining the intersection of the lines x=1 and y=1.
  - The line x=1 is equivalent to $-1x + 1 = 0$, and thus has homogeneous representation $\boldsymbol{l} = (-1,0,1)^T$. The line y=1 is equivalent to $-1y + 1 = 0$, and thus has homogeneous representation $\boldsymbol{l'} = (0,-1,1)^T$.
  - The intersection of 2 lines $\boldsymbol{l}$ and $\boldsymbol{l'}$ is the point $\boldsymbol{x} = \boldsymbol{l} \times \boldsymbol{l'} = (1,1,1)^T$, which is the inhomogeneous point $(1,1)^T$ as required.
- On the other hand, in projective geometry, angles and lengths do not make sense. Only concepts like point, line, incidence are retained. Felix Klein's Erlangen Program

Homogeneous Coordinates

- Homogeneous coordinate systems will allow us to transform between reference frames with a single matrix multiplication.

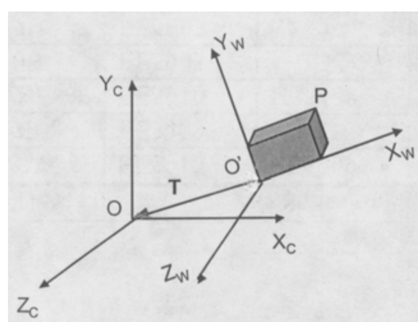
Coord of pt in camera frame

Coord of pt in world frame

$$\mathbf{P}_c = \mathrm{R}\,(\mathbf{P}_w - \mathbf{T})$$

Rotation matrix describing orientation of world frame wrt camera frame.

Coord of camera origin in world frame



The above follows the textbook's convention on p35, that transformation is defined by translation followed by rotation. The usual convention is however $P_c = RP_w + T$, i.e., rotation precedes translation (see p126 textbook, and p142 HZ). You have to take note of this if you try to derive the relation on p163 textbook.

Homogeneous Coordinates

- Using homogeneous coordinates, we append a 1 on the end of each set of coordinates

$$\tilde{P} = \begin{bmatrix} P_X \\ P_Y \\ P_Z \\ 1 \end{bmatrix}$$

- This allows us to write the transformation

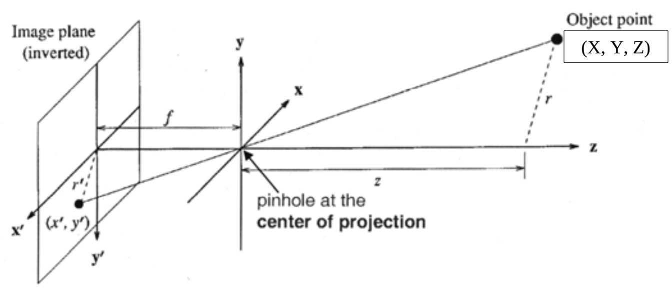
$$\begin{pmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{pmatrix} = \begin{bmatrix} \mathrm{R} & -\mathrm{RT} \\ 0^{\mathrm{T}} & 1 \end{bmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

Geometry of Perspective Imaging

- Given a point or line in space, where will we see it in the image?

- Conversely, can we recover 3-D information from 2-D images?

Perspective Imaging – pinhole camera model

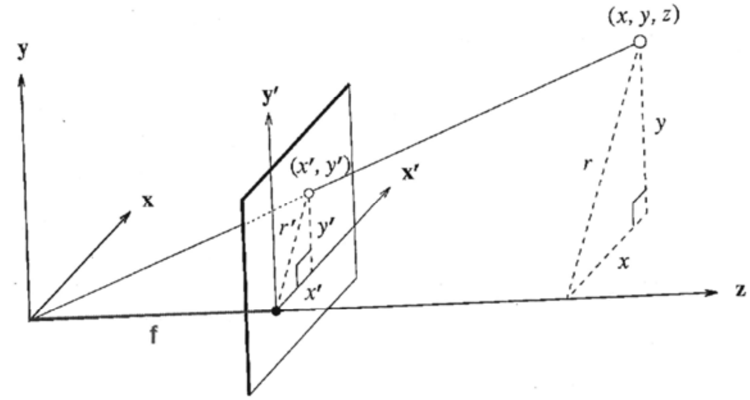


- The point on the image plane that corresponds to a particular point in the scene is found by following the line that passes through the scene point and the center of projection.

Perspective Imaging

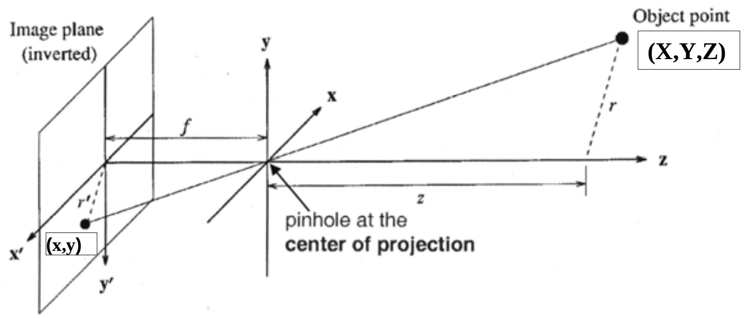
- Imaging system center of projection (COP) coincides with the origin of the 3-D coordinate system for the camera.
- Line of sight to a point in the scene is the line through the COP to that point
- Image plane is parallel to the x-y plane:
  - ❑ Distance to the image plane is f – focal length
  - ❑ This inverts the image
  - ❑ Move the image plane in front of the COP (A similar switching is done in the brain)

Perspective Imaging



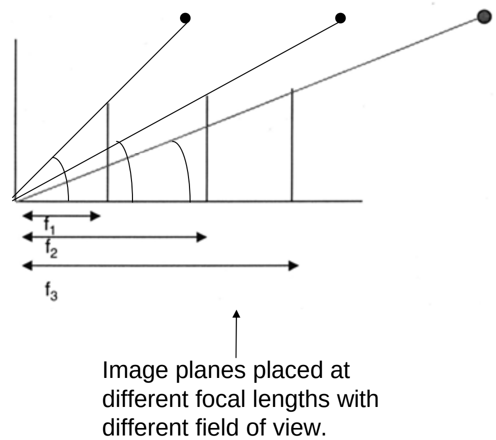
Perspective Imaging

- Using similar triangle, we can obtain the perspective projection equation: x=fX/Z; y=fY/Z



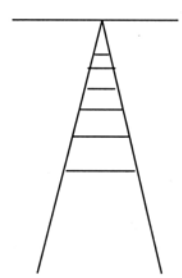
Properties of perspective Imaging

- As f gets smaller, image becomes more wide angle
- As f gets larger, image becomes more telescopic.



Properties of perspective Imaging

- Under perspective projection, lengths, angles,parallelism become "distorted".
- Mathematically described by projective geometry.
- Consider a set of railroad tracks:
  - ❑Their actual shape (Euclidean geometry):
    - ➢Tracks are parallel
    - ➢Ties are perpendicular to the track
    - ➢Ties are evenly spaced along the track
  - ❑Their appearance (Projective geometry)
    - ➢Tracks converge to a point on the horizon
    - ➢Tracks don't meet ties at right angle
    - ➢Ties becomes closer and closer towards the horizon



Camera Calibration

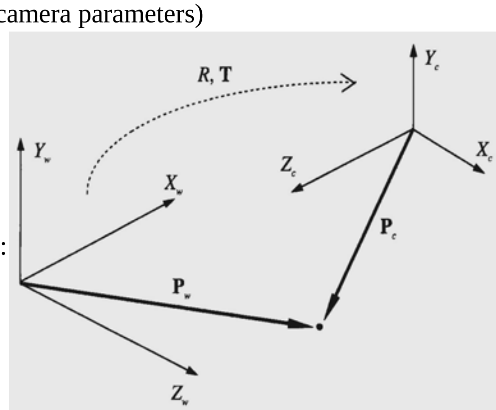
- Camera calibration is the problem of determining the elements that govern the relationship between the 2D image that a camera perceives and the 3D information of the imaged object.

3x4 Projection Matrix

There are 3 coordinate systems involved --- world, camera, and image.

1. **World**: (extrinsic/external camera parameters)

$\mathbf{P}_c = R\,(\mathbf{P}_w - \mathbf{T})$



- Can write as a linear mapping:

$$\begin{pmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{pmatrix} = \begin{bmatrix} R & -RT \\ 0^T & 1 \end{bmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

- Why not always use the camera ref frame as the world ref frame?
  - Camera might be moving, or we might have several cameras.

3x4 Projection Matrix

2. Camera: perspective projection.

$x_c = fX_c/Z_c;\ y_c = fY_c/Z_c$

This can be written as a linear mapping between homogeneous coordinates (the equation is only up to a scale factor):

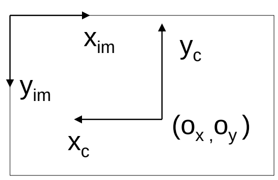
$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{pmatrix}$$

where the 3x4 **projection matrix** represents a map from 3D to 2D, and $x_c = x/w$, $y_c = y/w$.

3x4 Projection Matrix

3. Image:

$$x_c = -(x_{im} - o_x)s_x$$
$$y_c = -(y_{im} - o_y)s_y$$



1. The principal point (o_x, o_y) which is the point where the optic axis intersects the image plane.
2. The scaling in the image x and y directions, s_x and s_y. The aspect ratio is s_y / s_x.
3. Together with focal length f, these parameters are known as intrinsic/internal camera parameters.

Once these parameters are known, the camera is termed calibrated.

3x4 Projection Matrix

3. The relationship can be written in a matrix form:

$$\begin{pmatrix} x_{im} \\ y_{im} \\ 1 \end{pmatrix} = \begin{bmatrix} \frac{-1}{s_x} & 0 & o_x \\ 0 & \frac{-1}{s_y} & o_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_c \\ y_c \\ 1 \end{pmatrix}$$

Sometimes an additional skew factor *k* is included:

$$\begin{pmatrix} x_{im} \\ y_{im} \\ 1 \end{pmatrix} = \begin{bmatrix} \frac{-1}{s_x} & k & o_x \\ 0 & \frac{-1}{s_y} & o_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x_c \\ y_c \\ 1 \end{pmatrix}$$

The factor *k* is equal to 1/tan θ , where θ is the angle between the image axis. We will not consider *k* in this course.

Concatenating all stages

$$\begin{pmatrix} x_{im} \\ y_{im} \\ 1 \end{pmatrix} = \begin{bmatrix} \frac{-1}{s_x} & k & o_x \\ 0 & \frac{-1}{s_y} & o_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathrm{R} & -\mathrm{R\,T} \\ 0^{\mathrm{T}} & 1 \end{bmatrix} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix}$$

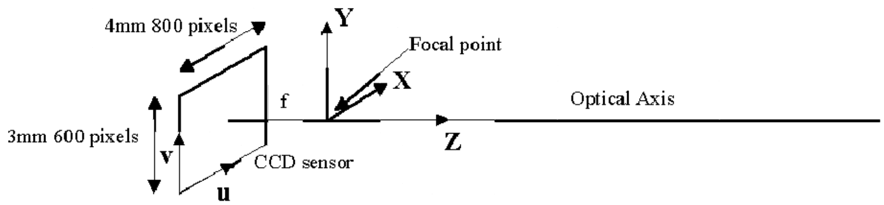
and regrouping the matrices:

$$\begin{pmatrix} x_{im} \\ y_{im} \\ 1 \end{pmatrix} = M_{\mathrm{int}} M_{ext} \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \quad \text{where} \quad \begin{cases} M_{ext} = \begin{bmatrix} R & -RT \end{bmatrix} \\ M_{\mathrm{int}} = \begin{pmatrix} \frac{-f}{s_x} & 0 & o_x \\ 0 & \frac{-f}{s_y} & o_y \\ 0 & 0 & 1 \end{pmatrix} \end{cases}$$

where M_{int} M_{ext} defines the 3x4 projection matrix from 3D-space to an image.

- if M_{int} is known, the camera is said to be calibrated,
- otherwise, it is an uncalibrated camera.

E.g. The figure below shows a typical pinhole camera model. We construct a camera frame of reference centered at the focal point of the camera with the z axis along the principal optical axis of the device. The focal length of this camera, f, is 8mm, the CCD sensor on the focal plane is 4mm wide and 3mm tall as shown. The sensor is divided into a grid of 800 pixels in the x direction and 600pixels in the y direction. The center of the imaging array corresponds to the point where the optical axis cuts the focal plane. The pixel coordinates are indexed from the bottom right hand corner as shown.



Compute the matrix of intrinsic parameters which relates the coordinates of features given with respect to the camera frame to the pixel coordinates of the projection of that point on the sensor.

Image aberration

- Aberrations
  - are variations in focal length due to imperfections in the lens. They result in a blurred image.

i. Spherical aberration:
  - focal length decreases on moving from the center of the lens to the periphery.

ii. Astigmatism:
  - focal length varies on moving around the center of the lens.

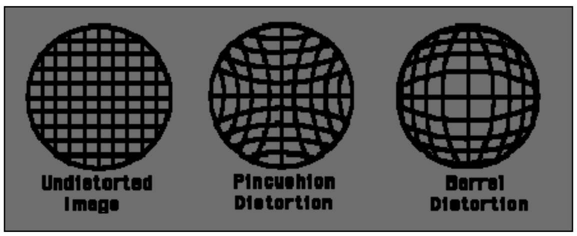
iii. Chromatic aberration:
  - focal length varies with the wavelength of the light.

Image distortion

Distortions are variations in positions of points in the image due to imperfections in the lens.

i. Radial distortion: small near the center but increasing towards the periphery. Depends very much on the camera.

- $\Delta r = K_1 r^3 + K_2 r^5 + K_3 r^7$



ii. Tangential distortion: Usually negligible.

Weak perspective camera

- Let max difference in depth between any two scene points = ΔZ
- Let average depth of scene points = Z_{ave}
- If $\Delta Z << Z_{ave}$, then for each scene point:

$$x \approx fX / Z_{ave} \qquad y \approx fY / Z_{ave}$$

- This gives linear camera model, corresponding to:
  - Orthographic projection (x=X,y=Y); followed by
  - Isotropic scaling by factor f/ Z_{ave}
- Viable model when $\Delta Z < Z_{ave} /20$
- Known as scaled orthography or weak perspective model – some sense of perspective retained (as opposed to being totally lost in orthographic model)

Reference

- Chapter 2 and Appendix of Textbook.
- Handout (Introduction to Visual Science).