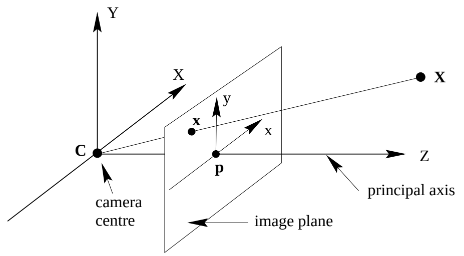
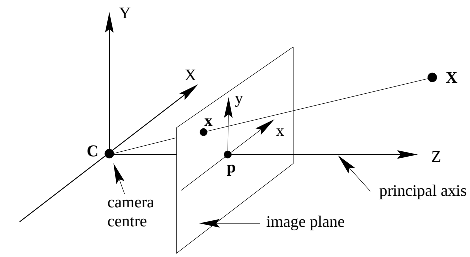
Camera models

- A camera model describes a projection from $\mathcal{P}^3$ to $\mathcal{P}^2$.
- The most basic camera model is the *pinhole camera* where a 3D world point $\mathbf{X}$ is projected onto a 2D image point $\mathbf{x}$ in the *image plane* through a *camera center* $\mathbf{C}$.
- In the physical camera, a mirror image is formed behind the camera center. However, often a virtual image plane is depicted in front of the camera center.
- The points $\mathbf{X}$, $\mathbf{C}$ and (the 3D point) $\mathbf{x}$ are *collinear*, i.e. lie on the same line in space.



Synonyms

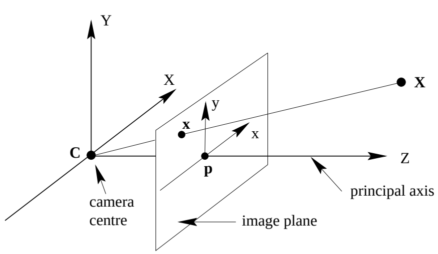
- The camera center is also called *projection center*, *perspective center*, *optical center*, or *focus*.
- The image plane is also called *focal plane*.
- The distance between the camera center and the image plane is called *focal distance*, *principal distance*, *camera constant*, or *film-focus-distance*.
- The ray through the camera center orthogonal to the image plane is called *principal axis*, *principal ray*, *camera axis* or *central ray*.
- The intersection between the principal ray and the image plane is called *principal point* or *foot point*.
- The plane through the camera center parallel with the image plane is called the *principal plane*.

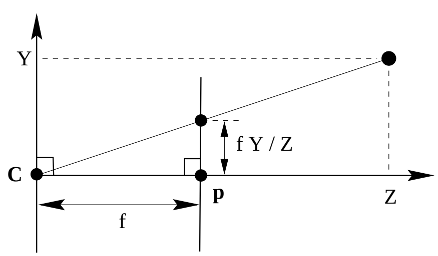


Central projection

- If the camera center is at the origin and the image plane is the plane $Z = f$ the world coordinate $(X, Y, Z)^\top$ is mapped to the point $(fX/Z, fY/Z, f)^\top$ in space or $(fX/Z, fY/Z)$ in the image plane, i.e.

$$(X, Y, Z)^\top \mapsto (fX/Z, fY/Z)^\top$$





Central projection in homogenous coordinates

- The corresponding expression in homogenous coordinates may be written as

$$\begin{bmatrix}X\\Y\\Z\\1\end{bmatrix} \mapsto \underset{\mathbf{x}}{\begin{bmatrix}fX\\fY\\Z\end{bmatrix}} = \underset{\mathtt{P}}{\begin{bmatrix} f & & & 0\\ & f & & 0\\ & & 1 & 0\end{bmatrix}} \underset{\mathbf{X}}{\begin{bmatrix}X\\Y\\Z\\1\end{bmatrix}}.$$

- The matrix $\mathtt{P}$ is called the *camera matrix* and maps the world point $\mathbf{X}$ onto the image point $\mathbf{x}$.
- In more compact form $\mathtt{P}$ may be written as

$$\mathtt{P} = \text{diag}(f, f, 1)[\mathtt{I} \mid \mathbf{0}],$$

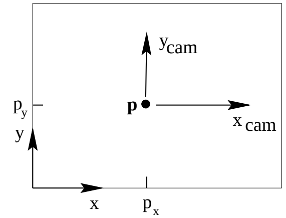
where $\text{diag}(f, f, 1)$ is a diagonal matrix and $\mathtt{I}$ is the identity matrix.

The principal point

- If the principal point is not at the origin of the image coordinate system, the mapping becomes

$$(X, Y, Z)^\top \mapsto (fX/Z + p_x, fY/Z + p_y)^\top,$$

where $(p_x, p_y)^\top$ are the image coordinates for the principal point.



- In homogenous coordinates

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} fX + Zp_x \\ fY + Zp_y \\ Z \end{bmatrix} = \begin{bmatrix} f & & p_x & 0 \\ & f & p_y & 0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

The camera calibration matrix

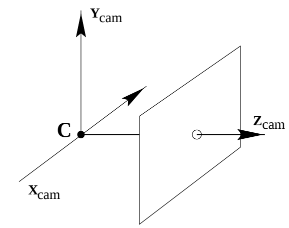
- If we write

$$\mathtt{K} = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix}$$

the projection may be written as

$$\mathbf{x} = \mathtt{K}[\mathtt{I} \mid \mathbf{0}]\mathbf{X}_{\text{cam}}.$$

- The matrix $\mathtt{K}$ is called the *camera calibration matrix*. The notation $\mathbf{X}_{\text{cam}}$ emphasizes that the world coordinates are with respect to the camera, i.e. the camera center is at the origin, the X and Y axes coincide and the principal axis coincides with the Z axis.



The camera position and orientation

- Usually the *camera coordinate system* does not coincide with the *world coordinate system*. The two coordinate systems are related via a rotation and a translation.
- If the point $\tilde{\mathbf{X}}$ denotes an cartesian 3D point in world coordinates and $\tilde{\mathbf{X}}_{\text{cam}}$ denotes the same point in camera coordinates, then their relation is given by

$$\tilde{\mathbf{X}}_{\text{cam}} = \mathtt{R}(\tilde{\mathbf{X}} - \tilde{\mathbf{C}}),$$

where $\tilde{\mathbf{C}}$ is the camera center in world coordinates and $\mathtt{R}$ is a 3×3 rotation matrix describing the orientation of the camera in space.

- In homogenous coordinates

$$\mathbf{X}_{\text{cam}} = \begin{bmatrix} \mathtt{R} & -\mathtt{R}\tilde{\mathbf{C}} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} \mathtt{R} & -\mathtt{R}\tilde{\mathbf{C}} \\ 0 & 1 \end{bmatrix} \mathbf{X},$$

or

$$\mathbf{x} = \mathtt{KR}[\mathtt{I} \mid -\tilde{\mathbf{C}}]\mathbf{X}.$$

Internal and external parameters

- The projection equation

$$\mathbf{x} = \mathtt{KR}[\mathtt{I} \mid -\tilde{\mathbf{C}}]\mathbf{X}$$

that describes the general projection for a pinhole camera has 9 degrees of freedom: 3 in $\mathtt{K}$ (the elements f, p_x, p_y), 3 in $\mathtt{R}$ (rotation angles) and 3 for $\tilde{\mathbf{C}}$.

- The elements of $\mathtt{K}$ describes properties internal to the camera while the parameters of $\mathtt{R}$ and $\tilde{\mathbf{C}}$ describe the relation between the camera and the world.
- The parameters are therefore called one of

$\mathtt{K}$	$\mathtt{R}, \tilde{\mathbf{C}}$
internal parameters	*external parameters*
internal orientation	*external orientation*
intrinsic parameters	*extrinsic parameters*
sensor model	*platform model*

- Sometimes it is practical not to write the camera center explicitly but instead describe the world-to-camera transformation as $\tilde{\mathbf{X}}_{\text{cam}} = \mathtt{R}\tilde{\mathbf{X}} + \mathbf{t}$ with camera matrix

$$\mathtt{P} = \mathtt{K}[\mathtt{R} \mid \mathbf{t}],$$

where $\mathbf{t} = -\mathtt{R}\tilde{\mathbf{C}}$.

Aspect ratio

- If we have different scale in the x and y directions, i.e. the pixels are not square, we have to include that deformation into the equation.
- Let m_x and m_y be the number of pixels per unit in the x and y direction of the image. Then the camera calibration matrix becomes

$$\begin{aligned}\mathtt{K} &= \begin{bmatrix} m_x & & \\ & m_y & \\ & & 1 \end{bmatrix} \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} \\ &= \begin{bmatrix} m_x f & & m_x p_x \\ & m_y f & m_y p_y \\ & & 1 \end{bmatrix} = \begin{bmatrix} \alpha_x & & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix},\end{aligned}$$

where $\alpha_x = f m_x$ and $\alpha_y = f m_y$ is the focal length in pixels in the x and y directions and $\tilde{\mathbf{x}}_0 = (x_0, y_0)^\top = (m_x p_x, m_y p_y)^\top$ is the principal point in pixels.
- A camera with unknown aspect ratio has 10 degrees of freedom.

Skew

- For an even more general camera model we can add a skew parameter s to describe any non-orthogonality between the image axis. Then the camera matrix becomes

$$\mathtt{K} = \begin{bmatrix} \alpha_x & s & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix}.$$

- A camera matrix

$$\mathtt{P} = \mathtt{KR}[\mathtt{I} \mid -\tilde{\mathbf{C}}] = [\mathtt{KR} \mid -\mathtt{KR}\tilde{\mathbf{C}}]$$

where the camera calibration matrix $\mathtt{K}$ is in the form above is called a *finite projective camera*.
- A finite projective camera has 11 degrees of freedom, the same as a 3×4 homogenous matrix.
- For a finite projective camera $\mathtt{P}$ the left 3×3 block $\mathtt{KR}$ is always non-singular.
- If we remove this restriction we get a *general projective camera* described by a 3×4 matrix of rank 3.

The general projective camera

- A general projective camera $\mathtt{P}$ maps world points $\mathbf{X}$ onto image points $\mathbf{x}$ according to

$$\mathbf{x} = \mathtt{P}\mathbf{X},$$

where the camera matrix $\mathtt{P}$ is a 3×4 homogenous matrix of rank 3.
- A general projective camera may be blocked as

$$\mathtt{P} = [\mathtt{M} \mid \mathbf{p}_4],$$

where $\mathtt{M}$ is a 3×3 matrix. We will see that $\mathtt{P}$ is a finite camera iff $\mathtt{M}$ is non-singular

The camera center

- The camera matrix $\mathtt{P}$ has a 1-dimensional null-space. Let the vector $\mathbf{C}$ be a basis for the null-space, i.e. $\mathtt{P}\mathbf{C} = \mathbf{0}$. We will see that $\mathbf{C}$ represents the camera center in homogenous coordinates.
- Study the line through $\mathbf{C}$ and an arbitrary point $\mathbf{A}$ in space. Points on this line may be described by

$$\mathbf{X}(\lambda) = \lambda\mathbf{A} + (1-\lambda)\mathbf{C}, \; -\infty < \lambda < \infty.$$

- Under the mapping $\mathbf{x} = \mathtt{P}\mathbf{X}$ points on the line are mapped to

$$\mathbf{x} = \mathtt{P}\mathbf{X}(\lambda) = \lambda\mathtt{P}\mathbf{A} + (1-\lambda)\underbrace{\mathtt{P}\mathbf{C}}_{=0} = \lambda\mathtt{P}\mathbf{A}, \; -\infty < \lambda < \infty.$$

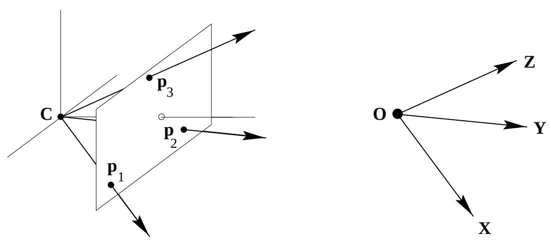
- This means that all points on the line are mapped to the same point $\mathtt{P}\mathbf{A}$. Thus, the line has to be a ray through the camera center. Since $\mathbf{A}$ was arbitrary, $\mathbf{C}$ must be a homogenous representation of the camera center.
- For finite cameras this is given from the fact that $\mathbf{C} = (\tilde{\mathbf{C}}^\top, 1)^\top$ is a null-vector to $\mathtt{P} = \mathtt{KR}[\mathtt{I} \mid -\tilde{\mathbf{C}}]$.
- Another line of reasoning concludes that the projection $\mathtt{P}\mathbf{C} = (0,0,0)^\top$ is undefined and the camera center is the only point with an undefined projection.

The camera column vectors

- Let $\mathbf{p}_i, i = 1, \ldots, 4$ be the columns of $\mathtt{P}$. Then $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ are the mappings of the axis directions in the world coordinate system.
- E.g. the X axis $\mathbf{D} = (1, 0, 0, 0)^\top$ is mapped to

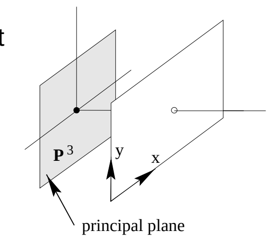
$$\mathtt{P}\mathbf{D} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & \mathbf{p}_4 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{p}_1.$$

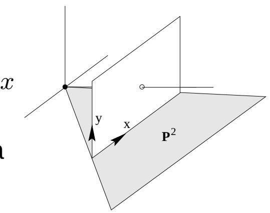
- The column vector $\mathbf{p}_4$ is the mapping of the world origin $(0, 0, 0, 1)^\top$.



The camera row vectors

- Let $\mathbf{P}^{i\top}, i = 1, \ldots, 3$ be the rows of $\mathtt{P}$.
- The principal plane of the camera is the plane through the camera center parallel with the image plane. It consists of all points $\mathbf{X}$ that are mapped to the line at infinity in the image plane, i.e. $\mathtt{P}\mathbf{X} = (x, y, 0)^\top$.
- A point is on the principal plane iff $\mathbf{P}^{3\top}\mathbf{X} = 0$. Thus the vector $\mathbf{P}^3$ represents the principal plane. Especially $\mathbf{C}$ is in the principal plane since $\mathtt{P}\mathbf{C} = \mathbf{0}$.
- All points $\mathbf{X}$ on the plane $\mathbf{P}^2$ satisfy $\mathbf{P}^{2\top}\mathbf{X} = 0$ and are therefore mapped to $\mathtt{P}\mathbf{X} = (x, 0, w)^\top$ on the image x axis. Since $\mathtt{P}\mathbf{C} = 0$ the plane $\mathbf{P}^2$ the plane may be defined by the camera center and the image x axis.
- Similarly the plane $\mathbf{P}^1$ may be defined by the camera center and the image y axis.
- The vectors $\mathbf{P}^1$ and $\mathbf{P}^2$ depend on the internal coordinate system of the camera. $\mathbf{P}^3$ does not.





he principal point and the principal axis

- The principal axis is the line through the camera center $\mathbf{C}$ orthogonal to the principal plane $\mathbf{P}^3$.
- The axis intersects the image plane in the principal point.
- The normal of the principal plane $\mathbf{P}^3 = (p_{31}, p_{32}, p_{33}, p_{34})^\top$ has direction $\hat{\mathbf{P}}^3 = (p_{31}, p_{32}, p_{33}, 0)^\top$.
- The projection of this point gives the principal point $\mathtt{P}\hat{\mathbf{P}}^3$.
- If $\mathtt{P} = [\mathtt{M} \mid \mathbf{p}_4]$ the principal point is calculated as

$$\mathbf{x}_0 = \mathtt{M}\mathbf{m}^3,$$

where $\mathbf{m}^{3\top}$ is the third row of $\mathtt{M}$.

The direction of the principal axis

- In theory all points $\mathbf{X}$ outside the principal plane are projected to a finite point in the image plane according to $\mathbf{x} = \mathtt{P}\mathbf{X}$. In reality only points *in front of* the camera.
- Let $\mathtt{P} = [\mathtt{M} \mid \mathbf{p}_4]$. We know that $\mathbf{m}_3$ points in the direction of the principal axis. However, the sign of $\mathtt{P}$ is undefined so the question is if $\mathbf{m}^3$ or $-\mathbf{m}^3$ points in the positive direction.
- Study the projection

$$\mathbf{x} = \mathtt{P}_{\text{cam}}\mathbf{X}_{\text{cam}} = \mathtt{K}[\mathtt{I} \mid \mathbf{0}]\mathbf{X}_{\text{cam}},$$

where $\mathbf{X}_{\text{cam}}$ is a 3D point in camera coordinates. Note that the vector $\mathbf{v} = \det(\mathtt{M})\mathbf{m}^3 = (0, 0, 1)^\top$ points *forward* in the direction of the principal axis, independently of the scaling of $\mathtt{P}_{\text{cam}}$ since

$$\mathtt{P}_{\text{cam}} \rightarrow k\mathtt{P}_{\text{cam}} \Rightarrow \mathbf{v} \rightarrow k^4\mathbf{v}$$

with unchanged sign.
- In world coordinates the camera matrix is $\mathtt{P} = k\mathtt{K}[\mathtt{R} \mid -\mathtt{R}\mathbf{p}_4]$, where $\mathtt{M} = k\mathtt{K}\mathtt{R}$. Since $\det(\mathtt{R}) = +1$ the direction of $\mathbf{v} = \det(\mathtt{M})\mathbf{m}^3$ is unchanged.

ward and backward projection of points

- A world point $\mathbf{X}$ is forward projected (mapped) onto an image point according to $\mathbf{x} = \mathtt{P}\mathbf{X}$.
- Points at infinity (*vanishing points*) $\mathbf{D} = (\mathbf{d}^\top, 0)^\top$ are mapped to

$$\mathbf{x} = \mathtt{P}\mathbf{D} = [\mathtt{M} \mid \mathbf{p}_4]\mathbf{D} = \mathtt{M}\mathbf{d}$$

and are only affected by the left 3×3 block $\mathtt{M}$ of $\mathtt{P}$.
- An image point is backward projected as a ray in space. We know two points on this ray: The camera center $\mathbf{C}$ and the point $\mathtt{P}^+\mathbf{x}$, where $\mathtt{P}^+$ is the *psuedo inverse* of $\mathtt{P}$. The psuedo inverse of $\mathtt{P}$ is the matrix $\mathtt{P}^+ = \mathtt{P}^\top(\mathtt{P}\mathtt{P}^\top)^{-1}$ such that $\mathtt{P}\mathtt{P}^+ = \mathtt{I}$.
- The point $\mathtt{P}^+\mathbf{x}$ is on the ray since it is projected onto $\mathbf{x}$, since $\mathtt{P}(\mathtt{P}^+\mathbf{x}) = \mathtt{I}\mathbf{x} = \mathbf{x}$.
- The ray can thus be describes as the line

$$\mathbf{X}(\lambda) = \mathtt{P}^+\mathbf{x} + \lambda\mathbf{C}.$$

Point depth

- Study a camera $\mathtt{P} = [\mathtt{M} \mid \mathbf{p}_4]$ that projects a point $\mathbf{X} = (X, Y, Z, 1)^\top = (\tilde{\mathbf{X}}^\top, 1)^\top$ in $\mathcal{R}^3$ on the image point $\mathbf{x} = w(x, y, 1)^\top = \mathtt{P}\mathbf{X}$.
- Let $\mathbf{C} = (\tilde{\mathbf{C}}^\top, 1)^\top$ be the camera center. Then

$$w = \mathbf{P}^{3\top}\mathbf{X} = \mathbf{P}^{3\top}(\mathbf{X} - \mathbf{C})$$

since $\mathtt{P}\mathbf{C} = \mathbf{0}$ for the camera center $\mathbf{C}$.
- Since

$$\mathbf{P}^{3\top}(\mathbf{X} - \mathbf{C}) = \mathbf{m}^{3\top}(\tilde{\mathbf{X}} - \tilde{\mathbf{C}}),$$

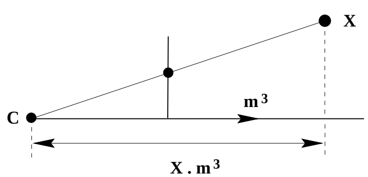
where $\mathbf{m}^3$ is the direction of the principal axis,

$$w = \mathbf{m}^{3\top}(\tilde{\mathbf{X}} - \tilde{\mathbf{C}})$$

may be interpreted as the inner product between the direction of the principal axis and the ray through the camera center and $\mathbf{X}$.

Point depth

- If the camera matrix has been normalized such that $\det(\mathtt{M}) > 0$ and $\|\mathbf{m}^3\| = 1$ the $\mathbf{m}^3$ is a unit vector in the positive principal axis direction. Then w may be interpreted as the depth of the point $\mathbf{X}$ from the camera center $\mathbf{C}$ in the direction of the principal axis



- For an unnormalized camera matrix, the depth of a point $\mathbf{X} = (X, Y, Z, T)^\top$ with respect to a camera $\mathtt{P} = [\mathtt{M} \mid \mathbf{p}_4]$ may instead be calculated as

$$\text{depth}(\mathbf{X}; \mathtt{P}) = \frac{\text{sign}(\det(\mathtt{M}))w}{T\|\mathbf{m}^3\|},$$

where

$$\mathtt{P}(X, Y, Z, T)^\top = w(x, y, 1)^\top.$$

- If $\mathtt{P}$ is a finite camera, then all points satisfying

$$\text{sign}(\det(\mathtt{M}))w/T > 0$$

Factorizing the camera matrix

- Given a camera matrix $\mathtt{P}$ the camera center is given from the nullspace $\mathtt{P}\mathbf{C} = \mathbf{0}$.
- For a finite camera

$$\mathtt{P} = [\mathtt{M} \mid -\mathtt{M}\tilde{\mathbf{C}}] = \mathtt{K}[\mathtt{R} \mid -\mathtt{R}\tilde{\mathbf{C}}]$$

with $\det(\mathtt{M}) > 0$, $\mathtt{K}$ and $\mathtt{R}$ may be calculated by the *RQ factorization* of $\mathtt{M}$.
- The RQ factorization of a matrix $\mathtt{A}$ gives an upper triangular matrix $\mathtt{K}$ (“R”) and an orthogonal matrix $\mathtt{R}$ (“Q”) such that $\mathtt{K}\mathtt{R} = \mathtt{A}$.
- The matrix $\mathtt{R}$ describes the orientation and $\mathtt{K}$ the calibration of the camera.
- If $\mathtt{K}$ is chosen to have positive diagonal elements the factorization is unique.
- If the calibration matrix $\mathtt{K}$ is scaled such that $\mathtt{K}_{33} = 1$ it may be interpreted as

$$\mathtt{K} = \begin{bmatrix} \alpha_x & s & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix}.$$

Affine cameras

- An affine camera $\mathtt{P} = [\mathtt{M} \mid \mathbf{p}_4]$ is an infinite camera, i.e. $\mathtt{M}$ has rank 2.
- An affine camera $\mathtt{P}$ has last row $\mathbf{P}^{3\top} = (0, 0, 0, 1)^\top$.
- Points at infinity in $\mathcal{P}^3$ are mapped onto points at infinity in $\mathcal{P}^2$.
- The camera center $\mathbf{C}$ is on the plane at infinity.
- The principal plane is the point at infinity.
- The principal point is undefined.
- The canonical form $[\mathtt{I} \mid \mathbf{0}]$ is replaced by the parallel projection matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

- The camera caibration matrix $\mathtt{K}$ is replaced by

$$\begin{bmatrix} \mathtt{K}_{2\times 2} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}.$$