CV2 Exam SS24 (Main exam)

Computer Vision II: Multiple view geometry (Technische Universität München)

Problem 1 Mathematical Background (18 credits)

0 1 2

1.1 How are the *kernel* and the *range* of a matrix $A \in \mathbb{R}^{n \times m}$ defined?

$\ker(A) = \{x \in \mathbb{R}^m \mid Ax = 0\}$ and $\text{range}(A) = \{Ax \mid x \in \mathbb{R}^m\}$

0 1 2

1.2 Let $A \in \mathbb{R}^{n \times n}$ be a real symmetric matrix. We have for x, y being eigenvectors belonging to distinct eigenvalues $\lambda_x \neq \lambda_y$:

$$\lambda_x \langle x, y \rangle = \langle Ax, y \rangle = \langle x, Ay \rangle = \lambda_y \langle x, y \rangle$$

What property of symmetric real matrices can be concluded from this?

From $\lambda_x \langle x, y \rangle = \lambda_y \langle x, y \rangle$ and $\lambda_x \neq \lambda_y$ it follows that $\langle x, y \rangle = 0$. Consequently, eigenvectors of a symmetric matrix associated to different eigenvalues are always orthogonal.

1.3 Which of the following statements is true for a vectorspace $(V, +)$ over $\mathbb{R}$? (1)

- [ ] The $(V, +)$ and $(V, \cdot)$ are groups.
- [ ] The scalar multiplication is defined as a mapping $\mathbb{R} \times V \to \mathbb{R}$.
- [x] The $(V, +)$ with scalar-vector multiplication has to fulfill the distributive law.
- [ ] An inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is defined.

0 1 2

1.4 Using the Frobenius inner product for matrices $A, B \in \mathbb{R}^{n \times n}$, the inner product is defined as $\langle A, B \rangle_F = \text{tr}(A^T B)$. Show, that the rotation matrix $R \in SO(n)$ does not change the inner product of two matrices $A, B \in \mathbb{R}^{n \times n}$, i.e. $\langle RA, RB \rangle_F = \langle A, B \rangle_F$.

$\langle RA, RB \rangle = \text{tr}(A^T R^T RB) = \text{tr}(A^T B) = \langle A, B \rangle$

In general, for a totally differentiable function $f(x)\colon H \to \mathbb{R}$, with H being either $\mathbb{R}^n$ or $\mathbb{R}^{m \times n}$, we have

$$\lim_{t \to 0} \frac{f(x + t\Delta x) - f(x)}{t} = \frac{df(x)}{dx} \Delta x = \langle \nabla f(x), \Delta x \rangle .$$

0 1

1.5 How is the directional derivative defined?

$lim_{t \to 0} \frac{f(x+t\Delta x) - f(x)}{t}$

0 1

1.6 What is the dimension of the gradient $\nabla f(x)$ in the above equation for the different cases of H?

n for $H = \mathbb{R}^n$ and $m \times n$ for $H = \mathbb{R}^{m \times n}$

1.7 What inner product could be used to express the directional derivative with the gradient for $H = \mathbb{R}^n$ and $H = \mathbb{R}^{m\times n}$ respectively?

- $\langle x, y\rangle = x^T y$ for $H = \mathbb{R}^n$
- $\langle A, B\rangle = \mathrm{tr}(A^T B)$ for $H = \mathbb{R}^{m\times n}$

1.8 The Frobenius norm of a matrix $M \in \mathbb{R}^{n\times m}$ is defined as: $\|M\|_F = \sqrt{\langle M, M\rangle_F}$ For a function $f(A) = \frac{1}{2}\|B - AC\|_F^2$, with $B, C \in \mathbb{R}^{n\times m}$, mapping $A \in \mathbb{R}^{n\times n}$ to $\mathbb{R}$. What is the gradient $\nabla f(A)$?

$\nabla f(A) = -(B - AC)C^T$

1.9 Determine which statement is correct for the sets. (2)

- $B_1 = \{\widehat{e}_1, \widehat{e}_2, \widehat{e}_3\}$, where e_i is the i-th unit vector in $\mathbb{R}^3$ and $\widehat{\cdot} : \mathbb{R} \to \mathrm{skew}(3)$ is mapping to the skew-symmetric matrix.
- $B_2 = \left\{ \begin{pmatrix}1\\0\\2\end{pmatrix}, \begin{pmatrix}-3\\1\\-6\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix} \right\}$

☐ B_2 form a basis of $\mathbb{R}^3$.

☐ B_1 spans $\mathbb{R}^3$.

☒ $\mathrm{span}(B_2) = \left\{ \begin{pmatrix}x\\y\\z\end{pmatrix} \in \mathbb{R}^3 \mid x - 0.5z = 0 \right\}$.

☐ B_1 is a basis of SO(3).

1.10 Determine which statement is correct (2)

☐ SO(3) is a subgroup of $\mathbb{R}^{3\times 3}$ with the matrix multiplication.

☐ SO(3) is a vectorspace over $\mathbb{R}$ with the matrix addition.

☐ SO(3) is a subvectorspace of $\mathbb{R}^{3\times 3}$ vectorspace over $\mathbb{R}$ with the matrix addition.

☒ $\mathrm{sym}(3) = \{A \in \mathbb{R}^{3\times 3} | A = A^T\}$ is a subvectorspace of $\mathbb{R}^{3\times 3}$ over $\mathbb{R}$ with the matrix addition.

☐ SO(3) is a subgroup of $\mathbb{R}^3$ with the matrix multiplication.

1.11 Let $\lambda_1, \dots, \lambda_k \in \mathbb{R}$. The set of vectors $\{v_1, \dots, v_k\} \subset \mathbb{R}^n$ is called *linearly independent* if (2)

☐ $\sum_{i=1}^{k} \lambda_i v_i = 0 \quad \Rightarrow \quad v_i = 0 \quad \forall\, i = 1, \dots, k$

☒ $\sum_{i=1}^{k} \lambda_i v_i = 0 \quad \Rightarrow \quad \lambda_i = 0 \quad \forall\, i = 1, \dots, k$

☐ $v_1 = 0$

☐ $v_i \neq v_j \quad \forall\, i, j = 1, \dots, k$ with $i \neq j$

Problem 2 Representing a Moving Scene (22 credits)

0 1 2

2.1 Write down expressions for the preservation properties of a rigid body motion.

For any two vectors $(v, w) \in \mathbb{R}^3$:
1) Norm preservation: $\|g(v)\| = \|v\|$
2) Cross product preservation: $g(v \times w) = g(v) \times g(w)$

0 1 2 3

2.2 How is the set of ridig-body motions defined? What is the inverse of $g \in SE(3)$? Your answer should have the following form:

$$SE(3) = \left\{ g = \begin{pmatrix} & \end{pmatrix} \middle| \qquad \right\}$$

$$g^{-1} = \begin{pmatrix} & \end{pmatrix}$$

$$SE(3) = \left\{ g = \begin{pmatrix} R & T \\ 0 & 1 \end{pmatrix} \middle| R \in SO(3), T \in \mathbb{R}^3 \right\}$$

$$g^{-1} = \begin{pmatrix} R^\top & -R^\top T \\ 0 & 1 \end{pmatrix}$$

0 1 2 3

2.3 We recall that the matrix of the rotation in the plane of angle θ is written as:

$$\begin{pmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{pmatrix}.$$

Let $R \in \mathbb{R}^{3\times3}$, corresponding to the rotation around the z-axis by the angle α, followed by a rotation around the x-axis by the angle γ. Write R in terms of α and γ.

$$R = \boxed{\begin{pmatrix} cos(\alpha) & -sin(\alpha) & 0 \\ cos(\gamma)sin(\alpha) & cos(\gamma)cos(\alpha) & -sin(\gamma) \\ sin(\gamma)sin(\alpha) & sin(\gamma)cos(\alpha) & cos(\gamma) \end{pmatrix}}$$

(A wrong answer which deserves 2 pts: Rotation around x-axis followed by rotation around z-axis:

$$\begin{pmatrix} cos(\alpha) & -sin(\alpha)cos(\gamma) & sin(\alpha)sin(\gamma) \\ sin(\alpha) & cos(\alpha)cos(\gamma) & -cos(\alpha)sin(\gamma) \\ 0 & sin(\gamma) & cos(\gamma) \end{pmatrix})$$

2.4 In the lecture you have seen the groups O(n) (Orthogonal Group), SL(n) (Special Linear Group) A(n) (Affine Group), E(n) (Euclidean Group), SE(n) (Special Euclidean Group), SO(n) (Special Orthogonal Group), GL(n) (General Linear Group) with $n \in \mathbb{N}$.

Which subset relation is wrong? 2

- ☐ $SE(n) \subset GL(n+1)$
- ☒ $SL(n) \subset O(n)$
- ☐ $E(n) \subset A(n)$
- ☐ $SO(n) \subset O(n)$

2.5 Consider the Lie groups SE(n) (Special Euclidean Group) and SO(n) (Special Orthogonal Group) with $n \in \mathbb{N}$. The corresponding Lie algebras $\mathfrak{se}(n)$ and $\mathfrak{so}(n)$ are defined as follows :

$$\mathfrak{so}(n) = \{A \in \mathbb{R}^{n\times n} | A^T = -A\}$$
$$\mathfrak{se}(n) = \left\{ \begin{pmatrix} A & v \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{(n+1)\times(n+1)} | A \in \mathfrak{so}(n), v \in \mathbb{R}^n \right\}$$

Which of the following statements is true? 3

- [ ] $\dim(\mathfrak{se}(3)) = 3$
- [x] $\dim(\mathfrak{se}(n)) - \dim(\mathfrak{so}(n)) = n$
- [ ] $\dim(\mathfrak{so}(3)) - \dim(\mathfrak{so}(2)) = 1$

2.6 Draw a schematic visualization of Lie Group and Lie Algebra, including the respective mappings (logarithmic map log and exponential map exp).

See lecture chapter 2, slide 14.

2.7 Let $R \in SO(3)$ the rotation around the y-axis with an angle of $\frac{\pi}{2}$. Give a vector $v \in \mathbb{R}^3$ such that $\widehat{v} = \log(R)$. No justification is required.

$v = \pm\frac{\pi}{2}(0, 1, 0)$

In the next questions, we consider two cameras whose poses, denoted respectively by (R_1, T_1) and (R_2, T_2), transform a point in world coordinates into camera coordinates.

2.8 By using $(R_i, T_i)_{i=1,2}$, write an expression of the rotation R and translation T of the relative pose, which transforms a point from camera 1 coordinates to camera 2 coordinates.

$R = R_2 R_1^\top$, and $T = T_2 - R_2 R_1^\top T_1$

2.9 Let C_1, C_2 be the positions of the two cameras, and $V \in \mathbb{R}^3$ the vector containing the coordinates of $\overrightarrow{C_1 C_2}$ in the world frame. Without justification, write an expression of V with respect to $(R_i, T_i)_{i=1,2}$.

$V = R_1^\top T_1 - R_2^\top T_2$

Problem 3 Perspective projection (22 credits)

In the following, no justifications are required, and only the final result matters.

3.1 A 3D point is given in the world coordinate frame by $P = (-9, 2, 12)^\top$. The camera pose which maps from camera coordinate to the world coordinate, and the intrinsic parameter matrix are respectively given by:

$$[R, T] = \begin{pmatrix} 0 & -1 & 0 & -1 \\ 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 4 \end{pmatrix} \qquad K = \begin{pmatrix} 300 & 0 & 200 \\ 0 & 300 & 200 \\ 0 & 0 & 1 \end{pmatrix}$$

Write down the pixel-position of the projected point in the image.

$(u, v) = (200, 500)$

3.2 We consider a camera whose optical center is at $C = (1, 0, 0)^\top$. The rotation matrix R, mapping a vector in world coordinate into camera coordinate, and intrinsic parameter matrix K are given respectively by

$$R = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad K = \begin{pmatrix} 500 & 0 & 320 \\ 0 & 400 & 240 \\ 0 & 0 & 1 \end{pmatrix}.$$

Write down the 3D point in the world coodinate frame, corresponding to the pixel $p = (70, 440)$ with depth 2.

$P = (3, 1, 1)^\top$

Given is a camera, which has its optical center at $C = (0, 1, 0)^\top$. The 3D points $X_1 = (0, -2, 2)^\top$, $X_2 = (1, 0, 1)^\top$ in the world coordinate frame, appear in the image at pixel coordinates $x_1 = (50, 100)^\top$ and $x_2 = (60, 130)^\top$. Assuming a perspective pinhole camera with no skew, the intrinsic matrix and rotation mapping vectors from camera to world coordinates are given respectively by:

$$K = \begin{pmatrix} f_x & 0 & o_x \\ 0 & f_y & o_y \\ 0 & 0 & 1 \end{pmatrix} \qquad R = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$$

We would like to estimate the intrinsic parameters.

3.3 What is the value of f_x and f_y?

$f_x = 10$, and $f_y = 90$

3.4 What is the value of o_x and o_y?

$o_x = 50$, and $o_y = 40$

3.5 Which of the following statements about radial distortion is true? 2

- [x] Radial distortion is caused by imperfect lenses.
- [ ] Radial distortion is caused by image sensor noise.
- [ ] Radial distortion is caused by non-rectangular pixels.
- [ ] The FOV (or ATAN) model for radial distortion is not invertible in closed form.

3.6 If the camera exhibits radial distortion, one can obtain a rectified virtual image through undistortion. We assume a perfect undistortion. Which of the following statements about a straight line oriented in any direction in 3D is true? 2

- [x] It will either appear as a straight line or a point in the rectified virtual image.
- [ ] It will appear as a straight line in the rectified virtual image.
- [ ] The angle between the line and another line in 3D is preserved in the rectified virtual image.
- [ ] It will appear as a point in the rectified virtual image.

3.7 Consider the radial distortion model

0 1 2 3

$$x_d = g(r)x \,, \tag{3.1}$$

which maps coordinates $x \in \mathbb{R}^2$ in the normalized image plane to distorted coordinates $x_d \in \mathbb{R}^2$. The factor $g(r)$ depends on the radius $r = |x|$ and is given as

$$g(r) = \frac{1}{1 + kr} \,, \tag{3.2}$$

where $k \in \mathbb{R}$ is a model parameter. Write down $f(r_d)$ solely with respect to k and r_d, such that the inverse distortion can be written as

$$x = f(r_d)x_d \,, \tag{3.3}$$

where $r_d = |x_d|$.

$f(r_d) = 1/(1 - kr_d)$

Problem 4 Estimating Point Correspondence (16 credits)

0
1

4.1 The brightness constancy assumption states that the intensity of a point in an image does not change over time. What is the mathematical formulation of this assumption? Do not differentiate the term.

The brightness constancy assumption is formulated as $\frac{d}{dt} I(\mathbf{x}(t), t) = 0$.

0
1
2

4.2 The assumption stated above leads to the optical flow constraint, which is given by

$$\nabla I(\mathbf{x}(t), t)^T \mathbf{v} + \frac{\partial I}{\partial t} = 0, \tag{4.1}$$

where $\mathbf{v} := \frac{d\mathbf{x}(t)}{dt}$ is the velocity vector. Explain the aperture problem and why we can not determine the velocity vector if we consider only the gradient at one pixel. Use solely the equation above to explain your answer. You can use the brief notation $\nabla I := \nabla I(\mathbf{x}(t), t)$.

The velocity vector is projected onto the gradient of the image intensity. $\nabla I^T \mathbf{v} = -\frac{\partial I}{\partial t}$. Therefore, we can only determine the velocity parallel to the gradient.

The optimization problem at a given point to estimate the velocity vector $\mathbf{v}$ at this point is given by

$$\min_{\mathbf{v}} \|\nabla I^T \mathbf{v} + \frac{\partial I}{\partial t}\|_2^2$$

and leads to the optimality condition

$$\nabla I \nabla I^T \mathbf{v} + \nabla I \frac{\partial I}{\partial t} = 0.$$

0
1
2

4.3 What is the structure tensor M' and the error vector q' in this context at one pixel location (no integration)?

The structure tensor M' is given by $M' = \nabla I \nabla I^T$ and the error vector q' is given by $q' = \nabla I \frac{\partial I}{\partial t}$.

0
1
2

4.4 Let $\nabla I^{\perp}$ be an orthogonal vector of ∇I. Show that $\nabla I^{\perp} \in Ker(M')$. What is the meaning of this statement?

$M' \nabla I^{\perp} = \nabla I \nabla I^T \nabla I^{\perp} = \nabla I \langle \nabla I, \nabla I^{\perp} \rangle = 0$

It restates the aperture problem, i.e. that we can not determine the velocity orthogonal to the gradient, if the structure tensor is build from the gradient at only one pixel location.

0
1
2

4.5 The Lukas Kanade algorithm uses the optical flow constraint (see equation above) emerging from the brightness constancy assumption to estimate the velocity vector $\mathbf{v}$. In addition, the method uses another assumption to formulate the optimization problem. What is the assumption?

Constant motion in a local neighborhood.

4.6 Using the convolution operator $*$, the gaussian kernel G, the structure tensor M', and the error vector q', state

- the modified optimization problem and
- the corresponding optimality constraint,

which are used by the Lukas Kanade Optical Flow Algorithm.

0 1 2 3

Optimization problem:

$$\min_{\mathbf{v}} G * \left(\nabla I^T \mathbf{v} + \frac{\partial I}{\partial t}\right)$$

Optimality constraint:

$$(G * M')\mathbf{v} = -G * q'$$

4.7 Consider the eigenvalues λ_1 and λ_2 of the structure tensor $M = G * M'$ and the three different cases:

0 1 2

1. $\lambda_1 \approx \lambda_2 \approx 0$,
2. $\lambda_1 \gg 0, \lambda_2 \approx 0$,
3. $\lambda_1 \gg 0, \lambda_2 \gg 0$.

We define the neighborhood of a pixel to be the surrounding area of a pixel with the size of the gaussian weighting kernel G. What can we say about the image gradients present in the neighborhood of a pixel?

1. the neighborhood has no dominant gradient directions, i.e. is a flat region
2. the neighborhood has one dominant gradient directions, i.e. is an edge
3. the neighborhood has more than one dominant gradient directions, i.e. is a corner

4.8 An edge detector would like to detect edges in an image. Which of the regions described above is of interest for the edge detector?

0 1 2

In the second case, where $\lambda_1 \gg 0, \lambda_2 \approx 0$.

Problem 5 3D reconstruction (21 credits)

The 8-point algorithm computes the rotation R and translation T (up to a scalar factor) between two images by solving the equation $\chi E^s = 0$, where E^s is the stacked essential matrix E and χ is computed from $n \geq 8$ point pair observations $\mathbf{x}_1^j$ and $\mathbf{x}_2^j$, for $j = 1, \dots, n$.

0 1 2

5.1 In practice, once you are given χ, how do you calculate E^s?

E^s is calculated as the ninth column of V_χ in the SVD $\chi = U_\chi \Sigma_\chi V_\chi^\top$.

0 1 2

5.2 If the camera have the same position for the two viewpoints, can E^s be used to accurately estimate the rotation? Justify your answer.

In this case, the translation would be zero, hence the same goes for the essential matrix ($E = \widehat{T}R$). Consequently, no rotation can be estimated from such an essential matrix.

0 1 2 3

5.3 The essential matrix E and hence the translation T are only defined up to an arbitrary scale $\gamma \in \mathbb{R}^+$, with $||E|| = ||T|| = 1$. After recovering R and T from the essential matrix, we therefore have the relation

$$\lambda_2^j \mathbf{x}_2^j = \lambda_1^j R \mathbf{x}_1^j + \gamma T \quad \forall j = 1, \dots, n,$$

with unknown scale parameters λ_i^j. To recover the depth of each point in the first camera coordinate system, we can solve the equation $M\vec{\lambda} = 0$ with $\vec{\lambda} = (\lambda_1^1, \lambda_1^2, \dots, \lambda_1^n, \gamma)^T \in \mathbb{R}^{n+1}$. Write down $M \in \mathbb{R}^{3n \times (n+1)}$, without justification.

$$M = \begin{pmatrix} \widehat{\mathbf{x}_2^1} R \mathbf{x}_1^1 & 0 & 0 & \dots & 0 & \widehat{\mathbf{x}_2^1} T \\ 0 & \widehat{\mathbf{x}_2^2} R \mathbf{x}_1^2 & 0 & \dots & 0 & \widehat{\mathbf{x}_2^2} T \\ 0 & 0 & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & \dots & \widehat{\mathbf{x}_2^{n-1}} R \mathbf{x}_1^{n-1} & 0 & \widehat{\mathbf{x}_2^{n-1}} T \\ 0 & \dots & \dots & 0 & \widehat{\mathbf{x}_2^n} R \mathbf{x}_1^n & \widehat{\mathbf{x}_2^n} T \end{pmatrix}$$

0 1 2

5.4 Assume you would like to compute the 3D rigid body motion (with unknown scale) between two images taken of the same static scene. What is the minimal number of 2D to 2D point-correspondences needed, if you know the intrinsic camera calibration?

The minimal number of 2D to 2D point-correspondences needed is 5 (see lecture chapter 5 slide 15).

For the next two questions, we consider two cameras whose intrinsics are denoted respectively by K_1 and K_2, and the relative pose mapping from camera 1 to camera 2 coordinates is denoted by (R, T).

0 1 2 3

5.5 Using the aforementioned matrices, state the epipolar constraint between two pixels $p_1 \in \mathbb{R}^3$ and $p_2 \in \mathbb{R}^3$ respectively of image 1 and 2, in homogeneous coordinates.

$p_2^\top K_2^{-\top} \widehat{T} R K_1^{-1} p_1 = 0$

5.6 In this question, we assume that the aforementioned intrinsics and extrinsics are given by:

$$[R, T] = \begin{pmatrix} 1 & 0 & 0 & 5 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 1 \end{pmatrix} \qquad K_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad K_2 = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

What is the equation of the epipolar line associated to the pixel $p_2 = (40, 30, 1)^\top$ of the second image? 4

- ☐ $v = -25u + 600$
- ☐ $v = -30u + 305$
- ☐ $v = -45u + 615$
- ☐ $v = -20u + 700$
- ☒ None of the other choices
- ☐ $v = -50u + 630$

5.7 We assume that there are three cameras whose rotations are all the identity matrix $\mathbb{I}_{3\times 3}$. Let T_{ij} the translation of the relative pose mapping from camera i to camera j coordinates, where $1 \leqslant i < j \leqslant 3$. We consider the pixels in homogeneous coordinates $p_1 = (10, 6, 1)^\top$, $p_2 = (5, 15, 1)^\top$ and $p_3 = (8, 9, 1)^\top$ of respectively image 1,2 and 3. And we assume that:

$$p_j^\top \widehat{T}_{ij} p_i = 0 \quad \forall (i,j) \in \{1,2,3\}^2 \text{ with } i < j$$

A statement is suggested: "There exist a unique 3D point whose projections in the different images correspond to p_1, p_2 and p_3". Is this statement true, wrong, or is it impossible to assess the validity of this statement? Justify your answer. 3

We can't assess the validity of this statement in general. Indeed, we have two cases:
1) All the intrinsics are the identity matrix:
The equations being satisfied correspond to the pairwise epipolar constraints, since the vectors p_1, p_2 and p_3 are not coplanar, this ensure the existence and uniqueness of the preimage, i.e. those pixels are projections of a unique 3D point.
2) Some of the intrinsic matrices are not the identity:
The equations have no significant meaning (the epipolar constraint should be $p_j^\top K_j^{-\top} \widehat{T}_{ij} K_i^{-1} p_i = 0$).

5.8 Consider a point $X \in \mathbb{R}^3$ which is observed in $m \geq 2$ images. Let $x_1, \dots, x_m \in \mathbb{R}^3$ denote the respective observations in homogeneous coordinates, and $\Pi_1, \dots, \Pi_m \in \mathbb{R}^{3\times 4}$ the multiple-view projection matrices, projecting a point into the respective image. Let

$$N_p := \begin{pmatrix} \Pi_1 & x_1 & 0 & \cdots & 0 \\ \Pi_2 & 0 & x_2 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Pi_m & 0 & 0 & \cdots & x_m \end{pmatrix} \in \mathbb{R}^{3m \times (m+4)}$$

What does the rank of N_p tell you regarding existence and uniqueness of a reconstruction X?

If $rank(N_p) = m + 4$ (full rank), there are no solutions.
If $rank(N_p) = m + 3$, the solution is unique (up to scale).
If $rank(N_p) < m + 3$, the space of the solutions have a dimension bigger than 1.

Problem 6 Bundle Adjustment and Nonlinear Optimization (17 credits)

0
1

6.1 With $r : \mathbb{R}^n \to \mathbb{R}^m$ being a differentiable function. When optimizing a function $f(x) = \|r(x)\|_2^2$, a common approach is to approximate the function locally by a quadratic function. What does the quadratic approximation of $f(x)$ for $x = x_0 + \Delta x$ look like if we apply the Taylor expansion of $f(x)$ at the point x_0? Express it in terms of the Hessian H, the gradient ∇f, and the function value f at the point x_0.

$$f(x) \approx f(x_0) + \nabla f(x_0)^T \Delta x + \frac{1}{2} \Delta x^T H(x_0) \Delta x$$

0
1
2

6.2 What does the quadratic approximation of $f(x)$ look like if we apply the first-order Taylor expansion of the function $r(x)$ at the point x_0? Express it in terms of the residual value r and the Jacobian J_r at the point x_0 of the function $r(x)$. By identifying terms with the answer to the previous question: What is the approximation of the Hessian H of $f(x)$ in this case?

$$f(x) = \|r(x)\|_2^2 \approx \|r(x_0) + J_r \Delta x\|_2^2 = \|r(x_0)\|_2^2 + 2r(x_0)^T J_r \Delta x + \|J_r \Delta x\|_2^2$$

$$H \approx 2 J_r^T J_r$$

0
1
2
3

6.3 State the optimization problem that is solved in the Gauss-Newton algorithm and the Levenberg-Marquardt algorithm. Use the diagonal matrix $D \in \mathbb{R}^{n \times n}$ and weight $\lambda \in \mathbb{R}$ to express the regularization term.

$$\min_{\Delta x} \|r(x_0) + J_r \Delta x\|_2^2$$

for Gauss-Newton and

$$\min_{\Delta x} \|r(x_0) + J_r \Delta x\|_2^2 + \lambda \|D \Delta x\|_2^2$$

for Levenberg-Marquardt

0
1

6.4 Are the Gauss-Newton and Levenberg-Marquardt methods iterative optimization algorithms? No justification is needed.

Yes, both methods are iterative optimization algorithms.

6.5 Let $g(\Delta x) = \|r(x_0) + J_r \Delta x\|_2^2$ be the function that approximates f locally at x_0. When computing the gradient of g w.r.t Δx, we get the following expression

$$\nabla g = J_r^T r(x_0) + J_r^T J_r \Delta x$$

0 1 2 3

Compute the gradient of the optimization objective for the case where a regularizer is present (the Levenberg-Marquardt algorithm is used) in problem 6.3, then state the update step for the Levenberg-Marquardt algorithm.

The gradient is given by

$$\nabla g + \lambda D^T D \Delta x = J_r^T r(x_0) + J_r^T J_r \Delta x + \lambda D^T D \Delta x$$

and the update step is given by

$$\Delta x = -(J_r^T J_r + \lambda D^T D)^{-1} J_r^T r(x_0)$$

for Levenberg-Marquardt.

6.6 Using the notation from the previous exercise. Which of the following is correct? (Solving the previous exercise helps you to solve this problem) 3

- ☐ The term $J_r^T J_r$ is the Hessian of the residual function r.
- ☒ For high values of λ, the update step Δx will be small, we stay in trust region.
- ☐ The regularization does not help to stabilize the computation of the inverse of $J_r^T J_r$.
- ☐ For high values of λ, and D being the identity the update step will not become closer to the direction of the gradient of the $\|r(x)\|^2$.

0
1
2
3

6.7 Consider the case where

- the camera intrinsics are known and that the 3D points are projected by the projection function $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ mapping a point from camera coordinates to pixel coordinates.
- we have two views only.

Let $X_j \in \mathbb{R}^3$, $j = 1, ..., N$ be a set of 3D points in the first camera coordinate system, whose corresponding 2D image coordinates are $\tilde{x}_i^j \in \mathbb{R}^2$ for images $i = 1, 2$. The goal is to estimate the 3D points X_j and the relative camera transformation $R \in SO(3)$ and $T \in \mathbb{R}^3$ mapping from the first to the second viewpoint coordinates that minimize the reprojection error, given the correspondences. The reprojection error is defined as the squared difference between the observed 2D coordinates and the projection of the feature's 3D coordinate. State the cost function for the bundle adjustment problem between two views.

$$E(R, T, X_1, \dots X_N) = \sum_{j=1}^{N} \|\tilde{x}_1^j - \pi(X_j)\|^2 + \|\tilde{x}_2^j - \pi(RX_j + T)\|^2. \quad (6.1)$$

0
1

6.8 Name the main advantage of solving the bundle adjustment using nonlinear least squares compared to the 8-point algorithm assuming a good initialization for the iterative optimization is given.

The 8-point algorithm is very unstable and sensitive to noise, while bundle adjustment is more robust and can handle noise and outliers better.