EECE 5550 Mobile Robotics Lab #5

David M. Rosen

Due: Dec 9, 2022

Problem 1: Camera pose estimation

As we mentioned in class, cameras are fundamentally *bearing* sensors: they measure the *direction* of an incoming ray of light connecting an imaged point with the camera's optical center. Thus, a simple (but very general) mathematical abstraction of a camera is the mapping π that projects each 3D point onto the sphere centered at the camera's optical center. We can write this map more concretely as:

$$\begin{aligned} \pi &\colon \mathbb{R}^3 - \{0\} \to S^2 \\ \pi({}_C p) &= \frac{{}_C p}{\|{}_C p\|} \end{aligned} \tag{1}$$

where ${}_C p \in \mathbb{R}^3$ are the coordinates of the 3D point measured in the camera's body-centric frame.

In this exercise, you will apply the camera model (1) together with some simple manifold geometry and optimization to perform *camera localization*: estimating the pose of a camera from observations of a set of known landmarks.

(a) Let ${}_C p \in \mathbb{R}^3$ be the coordinates of a point p as measured in the camera's body-centric coordinate frame. The *derivative* $d\pi \colon T_{{}_C p}(\mathbb{R}^3 - \{0\}) \to T_{\pi({}_C p)}(S^2)$ of the camera projection map (1) describes (to first order) how the image $\pi({}_C p) \in S^2$ of this point varies as a function of the camera pose. Calculate this derivative.

(b) Now suppose that we are given the coordinates ${}_W p \in \mathbb{R}^3$ of point p expressed in the *world* coordinate frame, and the pose $X_{CW} = (t_{CW}, R_{CW}) \in \mathrm{SE}(3)$ of the camera in the world frame. Derive an expression for ${}_C p$, the coordinates of point p expressed in the camera's *body-centric* coordinate frame.

(c) Using the result of part (b), compute $\frac{\partial {}_C p}{\partial X_{CW}}$, the derivative of point p's coordinates ${}_C p$ in the camera frame as a function of the camera pose X_{CW}. You may express your answer in terms of $\frac{\partial {}_C p}{\partial t_{CW}}$ and $\frac{\partial {}_C p}{\partial R_{CW}}$, the derivatives of ${}_C p$ with respect to the position and orientation of the camera (respectively).

(Hint: For computing the derivative with respect to the camera orientation R_{CW}, you may find it helpful to take advantage of the identity:

$$\mathrm{vec}(AB) = (B^{\mathsf{T}} \otimes I)\,\mathrm{vec}(A), \tag{2}$$

to express the derivative in terms of the vectorization $r_{CW} \triangleq \mathrm{vec}(R_{CW})$ of R_{CW}. This will let you write the camera orientation as a *vector*, rather than a *matrix*.)

(d) Define:

$$\psi \colon \mathbb{R}^3 \times \mathrm{SE}(3) \to S^2 \tag{3}$$

to be the mapping that accepts as input (i) the coordinates ${}_W p \in \mathbb{R}^3$ of a point p expressed in the *world* frame, and (ii) the pose $X_{CW} \in \mathrm{SE}(3)$ of the camera, and returns the image $u \in S^2$ of point p in the camera. Write down an expression for ψ (you may leave your answer in terms of the camera projection map π.)

Now suppose that there are m landmarks in the environment, at known positions ${}_W p_1, \ldots, {}_W p_m \in \mathbb{R}^3$ in the world frame, and that after taking an image with the camera (at an *unknown* pose $X_{CW} \in SE(3)$), these landmarks appear at the points $\tilde{u}_1, \ldots, \tilde{u}_m \in S^2$ in the image. Our goal is to *determine* the camera pose $X_{CW} \in \mathrm{SE}(3)$, given the landmark observations $\tilde{u}_1, \ldots, \tilde{u}_m$.

Given a *guess* $X_{CW} \in \mathrm{SE}(3)$ for the camera pose, a natural way to measure the quality of this guess is to calculate the sum of squared *reprojection errors*:

$$
\begin{aligned}
L\colon\ &\mathrm{SE}(3) \to \mathbb{R} \\
L(X_{CW}) &= \sum_{k=1}^{m} d_{S^2}\left(\psi({}_W p_k, X_{CW}),\ \tilde{u}_k\right)^2.
\end{aligned} \tag{4}
$$

In words: the function $L(X_{CW})$ reports the sum of the squared distances $d_{S^2}(u_k, \tilde{u}_k)$ (measured on S^2) between the *predicted* projection:

$$
u_k \triangleq \psi\left({}_W p_k, X_{CW}\right) \tag{5}
$$

of the kth landmark into a camera at pose X_{CW}, and the *actual* observed location $\tilde{u}_k$ of the kth landmark in the image. The best estimate $\hat{X}_{CW}$ for the camera pose is then the one that makes the sum of squared reprojection errors as small as possible:

$$
\hat{X}_{CW} = \underset{X_{CW} \in \mathrm{SE}(3)}{\operatorname{argmin}}\ L(X_{CW}). \tag{6}
$$

(e) The distance function on the sphere S^2 is:

$$
\begin{aligned}
d_{S^2}\colon S^2 \times S^2 &\to \mathbb{R} \\
d_{S^2}(x, y) &= \arccos\left(x^{\mathsf{T}} y\right).
\end{aligned} \tag{7}
$$

Let

$$
\begin{aligned}
\ell_k\colon S^2 \times S^2 &\to \mathbb{R} \\
\ell_k(u) &\triangleq d_{S^2}(u, \tilde{u}_k)^2
\end{aligned} \tag{8}
$$

be the function that returns the squared distance from the point $u \in S^2$ to the observed location of the kth landmark $\tilde{u}_k \in S^2$ in the image. What is the derivative of ℓ_k with respect to u?

(f) What is the derivative of L with respect to X_{CW}? You may express your result in terms of $\frac{\partial L}{\partial t_{CW}}$ and $\frac{\partial L}{\partial R_{CW}}$, the derivatives of L with respect to the position and orientation of the camera (respectively), and leave you answer in terms of the derivatives $\frac{\partial \ell_k}{\partial u}$ and $\frac{\partial \pi}{\partial_C p}$ you calculated in parts (a) and (e).

[Hint: You may find it convenient to start by rewriting (4) as:

$$
L(X_{CW}) = \sum_{k=1}^{m} \ell_k(u_k), \tag{9}
$$

with u_k as defined in (5), and then calculate derivatives of L with respect to t_{CW} and R_{CW} via repeated application of the Chain Rule, using your results from (a), (c), and (e).]

(g) Now let us consider the specific example in which we have 4 landmarks, located at the vertices of the standard simplex in $\mathbb{R}^3$:

$$_W p_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad _W p_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad _W p_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad _W p_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{10}$$

Suppose that after taking an image of these points with our camera, these landmarks appear at the following locations in the image:

$$\tilde{u}_1 = \begin{pmatrix} .866 \\ .289 \\ -.408 \end{pmatrix}, \quad \tilde{u}_2 = \begin{pmatrix} -.945 \\ .189 \\ -.267 \end{pmatrix}, \quad \tilde{u}_3 = \begin{pmatrix} -.189 \\ -.567 \\ .802 \end{pmatrix}, \quad \tilde{u}_4 = \begin{pmatrix} -.289 \\ -.866 \\ -.408 \end{pmatrix}. \tag{11}$$

Use the Manopt library to construct a model of the optimization problem (6), and solve it using gradient descent. Report the estimated camera pose $\hat{X}_{CW} = (\hat{t}_{CW}, \hat{R}_{CW}) \in \mathrm{SE}(3)$ obtained from (6), and submit the code that you wrote to implement the optimization.