Closed-form Solutions for Vision-aided Inertial Navigation

Tue-Cuong Dong-Si and Anastasios I. Mourikis
Dept. of Electrical Engineering, University of California, Riverside
E-mail: tdongsi@ee.ucr.edu, mourikis@ee.ucr.edu

April 12, 2011

Abstract

This report focuses on motion estimation using inertial measurements and observations of naturally occurring point features. To date, this task has primarily been addressed using filtering methods, which track the system state starting from known initial conditions. However, when no prior knowledge of the initial system state is available, (e.g., at the onset of the system's operation), the existing approaches are not applicable. To address this problem, in this work we present algorithms for computing all the observable quantities (platform attitude and velocity, feature positions, IMU biases, and IMU-camera calibration) *in closed form* directly from the sensor measurements, without any prior knowledge. As a key contribution of this work, we identify and analyze the properties of *minimal problems* that have a finite number of solutions, as well as *singular trajectories*, in which solutions cannot be computed. Additionally, to address the presence of noise in the measurements, we present a quadratically constrained least-squares solution and an iterative maximum-likelihood estimator.

Contents

1 Introduction

In recent years, the topic of motion estimation using visual and inertial measurements (often termed *vision-aided inertial navigation*) has attracted significant research interest (e.g., see [1–6] and references therein). Both cameras and MEMS inertial sensors are compact, inexpensive, and have low power requirements. Moreover, these sensors can operate in virtually any environment, and allow for full-3D pose estimation, thus providing a very versatile solution for navigation. The vast majority of existing techniques for navigation using camera and IMU measurements employ either a recursive Bayesian estimation approach [1–4], or a smoothing formulation [5]. In both cases, an accurate initial guess (prior estimate) for the state is necessary for reliable estimation. This is due to the fact that both types of methods rely on linearization of the measurement models, and thus in the absence of an accurate initial estimate, large linearization errors can lead to divergence.

In current practice, to initialize any of the state estimation methods discussed above, one typically uses domain-specific knowledge on a case-by-case basis. For instance, in certain applications, additional sensors (e.g., inclinometer and/or GPS) may be available, or it may be known that the platform is initially at rest. However, no generally applicable methods exist for determining all the observable system states directly from the sensor data, without use of a prior or of domain-specific knowledge. To address this problem, this report presents the following key contributions:

- We propose algorithms for computing, *in closed form*, all the observable quantities of the system directly from the visual and inertial measurements. In this work, we do *not* require that the 3D positions of the features are known in advance (i.e., we do not utilize fiducial points). Instead, the feature coordinates with respect to the camera are also computed as part of the solution.
- We carry out a detailed analysis of the properties of the problem, and identify *minimal problems* for which a solution exists. Moreover, we determine *singular cases* for the platform trajectory and the features' spatial configuration, which result in either multiple or infinite solutions. For instance, we prove that when the camera is moving with a constant acceleration, *two* discrete solutions for the trajectory exist in general.
- In the presence of noise, the formulation proposed here results in a quadratically-constrained least-squares problem, which can be solved analytically. This has the advantage of not requiring any prior information about the state, but it does not carry any optimality properties. To properly treat the noise in the sensor data, we propose a nonlinear iterative maximum-likelihood estimator (MLE) for estimating the observable quantities of the system.

Our main motivation for the closed-form solutions presented in this work is the initialization of recursive (e.g., [1–3]) and iterative estimators (e.g., the MLE described in Section 5.2). However, the methods presented here can also be applied for improving the robustness of vision-aided inertial navigation methods. For instance, the solutions of the minimal problems can be utilized within a RANSAC framework for outlier rejection. Moreover, the closed-form solutions can be run in parallel to a filtering algorithm (e.g., the EKF) as a sanity check, in order to detect possible divergence. In what follows, we present the details of our work.

2 Related work

In recent publications, the observability properties of the vision-aided inertial navigation system have been examined [1,2]. These works show that, in the absence of reference points with known global coordinates, the global position of the IMU, as well as the rotation about the axis of gravity (i.e., the yaw) are not observable. On the other hand, the following quantities are *in general* observable:

(O1) The IMU attitude with respect to the horizontal plane (i.e., the roll and pitch),

(O2) The IMU trajectory (position, velocity, and orientation) with respect to the initial IMU frame,

(O3) The feature positions with respect to the initial IMU frame.

(O4) The IMU biases, and

(O5) The transformation between the IMU and camera frames (ie., the camera-to-IMU calibration).

These results can be justified intuitively: if no known landmarks are available, the only source of absolute pose information is the gravity vector, which is sensed by the IMU accelerometers. Therefore, we can estimate the *global* orientation with respect to the horizontal plane (i.e., the plane normal to gravity), while all other quantities can be determined only in the *local* frame of the first IMU pose.

The results of [1, 2] provide valuable intuition into the properties of the problem at hand, and serve as a motivation for further study. Since we know which quantities are observable in general, the next step is to identify (i) the minimal conditions (number of features and images) required for a solution, and (ii) the number of possible solutions. The aforementioned works do not address these questions. The same holds for [5], in which an iterative method for determining the IMU orientation and attitude is proposed, but the conditions under which a solution exists are not studied. In Section 3.2, these limitations will be addressed.

In a paper that will soon appear [7], the visual and inertial measurements are used to analytically determine the initial velocity and attitude, as well as the feature positions. Compared to that work, in this report we (i) offer a detailed characterization of the special cases, under which a solution exists, (ii) propose solutions that can address the presence of noise in the sensor data, and (iii) we describe solutions for the observable quantities O4-O5 described above.

We point out that methods for motion estimation from features have been an important area of computer vision research (e.g., determination of the epipolar geometry from five point correspondences [8,9]). However, it is important to note that, in the absence of fiducial points, camera-only approaches can only estimate the camera motion up to an unknown scale. In contrast, the availability of the inertial measurements makes it possible to estimate (i) the *absolute* scale of the platform's motion, as well as (ii) the *absolute* orientation of the platform with respect to the horizontal plane. In recent publications, the second of these properties has been exploited. Specifically, [10, 11] examine the problem of motion estimation with a *known* vertical direction, and derive algorithms for its solution. However, these approaches do not use the inertial measurements to estimate the scale of the camera motion, and require that the camera is at rest when images are recorded (to allow for estimating its orientation using the accelerometer measurements). In our work, both of these limitations are lifted.

3 Closed-form solution: Known biases and IMU-Camera calibration

We first study the case in which the IMU biases and the transformation between the IMU and camera are known (e.g., from prior calibration), and our goal is to estimate the observable quantities O1-O3 defined in Section 2. The case where all quantities are unknown and need to be estimated from the data is treated in Section 4.

Consider the case where N images are recorded at the time instants $t_0, t_1, \ldots, t_{N-1}$. By employing a suitable image processing algorithm (e.g., KLT tracking [12]), we track M feature points in the images. For the solution presented here, we do *not* require all features to be tracked in all images. In addition to the feature observations, the IMU (gyroscope and accelerometer) measurements for the time interval $[t_0, t_{N-1}]$ are available. In the remainder of the section, we show how the feature observations and IMU measurements can be used to estimate the observable quantities O1-O3.

The measurements of the IMU gyroscopes and accelerometers are given by the following equations [13]:

$$\boldsymbol{\omega}_m(t) = {}^B\boldsymbol{\omega}(t) + \mathbf{b}_g + \mathbf{n}_\omega(t) \tag{1}$$

$$\mathbf{a}_m(t) = {}^B_G\mathbf{C}(t)({}^G\mathbf{a}(t) - {}^G\mathbf{g}) + \mathbf{b}_a + \mathbf{n}_a(t) \tag{2}$$

where[1] ${}^B\boldsymbol{\omega}(t)$ denotes the 3D rotational velocity vector expressed in the IMU frame, ${}^G\mathbf{a}(t)$ is the IMU acceleration in the global frame, ${}^G\mathbf{g}$ is the gravitational acceleration vector expressed in the global frame, $\mathbf{b}_g$, and $\mathbf{b}_a$ are the gyroscope and accelerometer biases, while $\mathbf{n}_\omega$, and $\mathbf{n}_a$ represent the noise in the gyroscope and accelerometer measurements, respectively. In the remainder of this section, we assume that the biases are known and removed from the measurements. Moreover, in order to obtain a closed-form solution, we will ignore the measurement noise, which is addressed properly in Section 5.

To estimate the IMU orientation change in the interval $[t_0, t_i]$, we integrate the following differential equation [14]:

$${}^{B_0}_{B}\dot{\mathbf{C}}(t) = -{}^{B_0}_{B}\mathbf{C}(t)\lfloor\boldsymbol{\omega}_m(t)\times\rfloor \;\;,\;\; {}^{B_0}_{B}\mathbf{C}(t_0) = \mathbf{I}_3 \tag{3}$$

in $[t_0, t_i]$. This yields the rotation matrix ${}^{B_0}_{B_i}\mathbf{C} = {}^{B_0}_{B}\mathbf{C}(t_i)$, which describes the IMU rotation between times t_0 and t_i. Note that in practice, we may not directly integrate the above equation, but use a quaternion-based representation of the orientation, or roll-pitch-yaw angles. These representations are equivalent for our purposes, and we here use (3) to simplify the presentation. The global position of the IMU at time t_i is computed as:

$${}^G\mathbf{p}(t_i) = {}^G\mathbf{p}(t_0) + {}^G\mathbf{v}(t_0)\Delta t_i + \int_{t_0}^{t_i}\int_{t_0}^{\tau} {}^G\mathbf{a}(\varsigma)d\varsigma d\tau \tag{4}$$

[1] Throughout this report, the IMU (body) frame is denoted by $\{B\}$, the camera frame by $\{C\}$, and the global (inertial) frame by $\{G\}$. ${}^X\mathbf{y}$ denotes the vector $\mathbf{y}$ expressed with respect to frame $\{X\}$, and ${}^X_Y\mathbf{C}$ denotes the rotation matrix transforming vectors from frame $\{Y\}$ into frame $\{X\}$. $\mathbf{I}_n$ is the $n \times n$ identity matrix, and finally, $\lfloor\mathbf{y}\times\rfloor$ is the skew-symmetric matrix associated with the 3×1 vector $\mathbf{y}$.

where ${}^{G}\mathbf{v}(t_0)$ is the initial velocity, and $\Delta t_i = t_i - t_0$. Using Eq. (2) and the above equation we obtain (see Appendix A):

$$ {}^{G}\mathbf{p}(t_i) = {}^{G}\mathbf{p}(t_0) + {}^{G}\mathbf{v}(t_0)\Delta t_i + {}^{G}\mathbf{g}\frac{\Delta t_i^2}{2} + {}^{G}_{B_0}\mathbf{C}\,\mathbf{s}(t_i) \tag{5} $$

$$ \text{where} \qquad \mathbf{s}(t_i) = \int_{t_0}^{t_i}\int_{t_0}^{\tau} {}^{B_0}_{B}\mathbf{C}(\varsigma)\mathbf{a}_m(\varsigma)\,d\varsigma d\tau \tag{6} $$

Note that Eq. (5) involves the global position, velocity, and orientation of the IMU, which are impossible to determine using only visual and inertial data. Therefore, our next step is to remove these quantities. To this end, we note that the IMU position at time t_i with respect to $\{B_0\}$ is given by ${}^{B_0}\mathbf{p}_{B_i} = {}^{G}_{B_0}\mathbf{C}^T({}^{G}\mathbf{p}(t_i) - {}^{G}\mathbf{p}(t_0))$. Using this expression, we can re-arrange Eq. (5) to obtain:

$$ {}^{B_0}\mathbf{p}_{B_i} = {}^{B_0}\mathbf{v}_0\Delta t_i + {}^{B_0}\mathbf{g}\frac{\Delta t_i^2}{2} + \mathbf{s}(t_i) \tag{7} $$

where ${}^{B_0}\mathbf{v}_0 = {}^{G}_{B_0}\mathbf{C}^T\,{}^{G}\mathbf{v}(t_0)$ is the IMU velocity at time t_0 expressed with respect to frame $\{B_0\}$, and ${}^{B_0}\mathbf{g} = {}^{G}_{B_0}\mathbf{C}^T\,{}^{G}\mathbf{g}$ is the gravity vector expressed with respect to the same frame.

Eq. (7) demonstrates that in order to estimate the platform trajectory using the inertial measurements, the initial velocity, ${}^{B_0}\mathbf{v}_0$, as well as the gravity vector in the local frame, ${}^{B_0}\mathbf{g}$ are required (note that $\mathbf{s}(t_i)$ depends *only* on the IMU measurements [Eq. (6) and (3)]). If these two vectors are known, we can employ Eq. (7) to determine the platform position in the local frame for any time instant of interest. By differentiation of the position we can also determine the platform velocity. Moreover, the IMU orientation in the local frame can be determined by integration of Eq. (3), which only requires the gyroscope measurements. The above discussion shows that to determine the platform trajectory in the local frame (O2 in Section 2), it suffices to know the vectors ${}^{B_0}\mathbf{v}_0$ and ${}^{B_0}\mathbf{g}$. Moreover, note that, since the gravity vector is normal to the horizontal plane, ${}^{B_0}\mathbf{g}$ defines the *orientation of the horizontal plane* with respect to $\{B_0\}$. Thus, estimating ${}^{B_0}\mathbf{g}$ is equivalent to estimating the IMU orientation with respect to the horizontal plane (O1 in Section 2).

In summary, we see that if we are able to determine ${}^{B_0}\mathbf{v}_0$ and ${}^{B_0}\mathbf{g}$, we can immediately obtain the observable quantities O1 and O2 described in Section 2. We next show how we can employ the feature observations to achieve this goal, as well as to compute the feature positions (O3). Assuming an intrinsically calibrated camera, the observation of the j-th feature at time t_i is described by the perspective camera model:

$$ \mathbf{z}_{ij} = \begin{bmatrix} u_{ij} \\ v_{ij} \end{bmatrix} = \begin{bmatrix} \frac{{}^{C_i}x_j}{{}^{C_i}z_j} \\ \frac{{}^{C_i}y_j}{{}^{C_i}z_j} \end{bmatrix} + \mathbf{n}_{ij} \;\;,\;\; {}^{C_i}\mathbf{p}_j = \begin{bmatrix} {}^{C_i}x_j \\ {}^{C_i}y_j \\ {}^{C_i}z_j \end{bmatrix} \tag{8} $$

where ${}^{C_i}\mathbf{p}_j$ is the position of the j-th feature with respect to the camera frame at time t_i, and $\mathbf{n}_{ij}$ is the measurement noise. We use the set $\mathcal{S}_m$ to describe all pairs of indices $\{i, j\}$ that describe the available measurements.

Using basic properties of frame transformations, we can express the vector ${}^{C_i}\mathbf{p}_j$ as follows:

$$ {}^{C_i}\mathbf{p}_j = {}^{C}_{B}\mathbf{C}\;{}^{B_i}_{B_0}\mathbf{C}\;({}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{p}_{B_i}) + {}^{C}\mathbf{p}_B \tag{9} $$

where ${}^{B_0}\mathbf{p}_j$ is the position of the feature with respect to $\{B_0\}$, while $\{{}^{C}_{B}\mathbf{C}, {}^{C}\mathbf{p}_B\}$ denotes the constant transformation (rotation and translation) between the IMU and camera frames. Using Eq. (7), we can rewrite (9) as:

$$ {}^{C_i}\mathbf{p}_j = {}^{C}_{B}\mathbf{C}\;{}^{B_i}_{B_0}\mathbf{C}\left({}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{v}_0\Delta t_i - {}^{B_0}\mathbf{g}\frac{\Delta t_i^2}{2} - \mathbf{s}(t_i)\right) + {}^{C}\mathbf{p}_B $$

On the right-hand side of above equation the unknown quantities are the vectors ${}^{B_0}\mathbf{p}_j$, ${}^{B_0}\mathbf{v}_0$, and ${}^{B_0}\mathbf{g}$, while all other terms are known. Proceeding further, we employ Eq. (8), to obtain (ignoring the measurement noise):

$$ \begin{bmatrix} 1 & 0 & -u_{ij} \\ 0 & 1 & -v_{ij} \end{bmatrix} {}^{C_i}\mathbf{p}_j = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{10} $$

By re-arranging terms, the above equation can be written as $\mathbf{A}_{ij}\mathbf{x} = \mathbf{b}_{ij}$, where $\mathbf{x}$ is the following $(3M + 6) \times 1$ vector:

$$ \mathbf{x} = \begin{bmatrix} {}^{B_0}\mathbf{p}_1^T & {}^{B_0}\mathbf{p}_2^T & \ldots & {}^{B_0}\mathbf{p}_M^T & {}^{B_0}\mathbf{v}_0^T & {}^{B_0}\mathbf{g}^T \end{bmatrix}^T \tag{11} $$

and

$$\mathbf{A}_{ij} = \mathbf{J}_{ij} \begin{bmatrix} \mathbf{0}_{3\times 3} & \dots & \underbrace{\mathbf{I}_3}_{j-\text{th block}} & \dots \mathbf{0}_{3\times 3} & -\Delta t_i \mathbf{I}_3 & -\frac{\Delta t_i^2}{2}\mathbf{I}_3 \end{bmatrix} \tag{12}$$

$$\mathbf{b}_{ij} = \mathbf{J}_{ij}\ \mathbf{s}(t_i) - \begin{bmatrix} 1 & 0 & -u_{ij} \\ 0 & 1 & -v_{ij} \end{bmatrix} {}^{C}\mathbf{p}_B \tag{13}$$

$$\mathbf{J}_{ij} = \begin{bmatrix} 1 & 0 & -u_{ij} \\ 0 & 1 & -v_{ij} \end{bmatrix} {}^{C}_{B}\mathbf{C}\ {}^{B_i}_{B_0}\mathbf{C} \tag{14}$$

We therefore see that from the observation of the j-th feature in the i-th image we obtain one equation of the form $\mathbf{A}_{ij}\mathbf{x} = \mathbf{b}_{ij}$. By collecting the equations resulting from all feature measurements, we obtain the linear system

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{15}$$

where $\mathbf{A}$ is a matrix with block rows $\mathbf{A}_{ij}$, and $\mathbf{b}$ is a block vector with block elements $\mathbf{b}_{ij}$, for all $\{i, j\} \in \mathcal{S}_m$. It is important to observe that both $\mathbf{A}$ and $\mathbf{b}$ can be computed using (i) the feature measurements, (ii) the IMU measurements, and (iii) the known IMU-camera transformation, while all the unknown quantities are included in the vector $\mathbf{x}$, which appears *linearly* in the above equation.

3.1 Solution of the linear system in the absence of noise

If the matrix $\mathbf{A}$ has full column rank, ie., if $\mathrm{rank}(\mathbf{A}) = 3M + 6$, then the solution to Eq. (15) is unique, given by:

$$\mathbf{x}^{\star} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \tag{16}$$

An interesting observation is that this solution allows us to determine not only the direction of the gravity vector in $\{B_0\}$, but also the *magnitude* of the gravitational acceleration, $g = ||{}^{B_0}\mathbf{g}||_2$. While this may be useful in certain settings, in the vast majority of applications the norm of the gravity vector is known in advance with high precision. Knowing the magnitude of gravity provides us with an additional constraint, which we can use to expand the situations under which a solution is possible.

Specifically, if the rank of $\mathbf{A}$ equals $3M+5$ (i.e., one less than the number of its columns), the matrix has a nullspace of dimension one: $\dim(\mathcal{N}(\mathbf{A})) = 1$. In this case Eq. (15) has an infinite number of solutions, described by $\mathbf{x} = \mathbf{A}^{+}\mathbf{b} + \alpha\mathbf{n}$, where $\mathbf{A}^{+}$ is a pseudoinverse of $\mathbf{A}$, $\mathbf{n}$ is a basis vector for $\mathcal{N}(\mathbf{A})$, and α is an arbitrary scalar [15]. However, if we know the magnitude of the gravitational acceleration, g, then we can enforce the constraint that the norm of the vector formed by the last three elements of $\mathbf{x}$ is equal to g [Eq. (11)]:

$$||\mathbf{\Pi}\left(\mathbf{A}^{+}\mathbf{b} + \alpha\mathbf{n}\right)||_2 = g, \text{ with } \mathbf{\Pi} = [\mathbf{0}_{3\times(3M+3)}\ \ \mathbf{I}_3] \tag{17}$$

This is a quadratic equation in α and has two real roots, α_1 and α_2. Thus, we obtain two discrete solutions for $\mathbf{x}$:

$$\mathbf{x}_1^{\star} = \mathbf{A}^{+}\mathbf{b} + \alpha_1\mathbf{n}, \quad \mathbf{x}_2^{\star} = \mathbf{A}^{+}\mathbf{b} + \alpha_2\mathbf{n} \tag{18}$$

It should be noted that if the nullspace of $\mathbf{A}$ is of dimension higher than one, then the solutions of (15) contain more than one arbitrary scalars. In that case, even if the known-gravity constraint is used, it is not possible to obtain a finite number of solutions. We now summarize the above results as follows:

Result 1. If $\mathbf{A}$ has full column rank (i.e., $\dim(\mathcal{N}(\mathbf{A})) = 0$), there exists a unique solution for $\mathbf{x}$, given by Eq. (16). On the other hand, if $\dim(\mathcal{N}(\mathbf{A})) = 1$, there exist two solutions in general. These can be found by numerically computing $\mathbf{n}$ and $\mathbf{A}^{+}$, solving Eq. (17) for α, and substituting in (18).

In either case, these solutions provide, *in closed form*, the position of all the features, the initial IMU velocity, and the gravity vector, expressed with respect to the first IMU frame. As previously explained, this allows us to compute the observable quantities O1-O3 in the system, which is our key objective.

3.2 Analysis of the rank of $\mathbf{A}$ and minimal cases

We next examine the rank of $\mathbf{A}$ and the dimension of $\mathcal{N}(\mathbf{A})$ in important cases of interest, and identify the minimal requirements (in terms of the number of features and images) that allow for computing a finite number of solutions.

We first derive a condition for the number of measurements needed. Since $\mathbf{A}$ contains two rows for each feature observation [Eq. (12)], the total number of rows in the matrix is $2K$, where K is the total number of measurements available: $K = \sum_{j=1}^{M} n_j$, where n_j is the number of times feature j has been observed. Therefore, $\mathrm{rank}(\mathbf{A}) \leq 2K$, which, by application of the rank-nullity theorem [15], yields the inequality $\dim(\mathcal{N}(\mathbf{A})) \geq 3M + 6 - 2K$. Using this inequality and Result 1, we obtain the following result:

Result 2. For a unique solution to exist, a *necessary*, but not sufficient, condition is that the number of measurements satisfies $K \geq \frac{3}{2}M + 3$. On the other hand, for a finite number of solutions (one or two) to exist, a necessary condition is $K \geq \frac{3}{2}M + \frac{5}{2}$. If neither of the above necessary conditions are met, then it is guaranteed that infinite solutions exist.

If the feature measurements meet the above necessary conditions, then we can proceed to compute $\dim(\mathcal{N}(\mathbf{A}))$, to verify its dimension (since the above conditions are not sufficient, when they are met it may still be possible that infinite solutions exist). For arbitrary trajectories and feature placements, the dimension of $\mathcal{N}(\mathbf{A})$ can be numerically computed, on a case-by-case basis. However, by employing an analytical approach to determining the rank of $\mathbf{A}$, we can identify an important special case. Specifically, in Appendix B.1, we show that if the camera moves with a constant acceleration, that is, if

$$
{}^{B_0}\mathbf{p}_{C_i} = {}^{B_0}\mathbf{p}_{C_0} + \Delta t_i \mathbf{v} + \frac{\Delta t_i^2}{2}\mathbf{a}\,, \tag{19}
$$

then $\mathbf{A}$ has a nullspace of dimension at least one, and the following vector is one basis vector for this nullspace:

$$
\mathbf{n} = \begin{bmatrix} \left({}^{B_0}\mathbf{p}_1 - {}^{B_0}\mathbf{p}_{C_0}\right)^T & \ldots & \left({}^{B_0}\mathbf{p}_M - {}^{B_0}\mathbf{p}_{C_0}\right)^T & \mathbf{v}^T & \mathbf{a}^T \end{bmatrix}^T
$$

If enough feature measurements are available, this will be the only basis vector for the nullspace of $\mathbf{A}$, and thus $\dim(\mathcal{N}(\mathbf{A})) = 1$. Thus, by application of Result 1, we conclude that:

Result 3. When the camera moves with a constant acceleration, two discrete solutions for the camera trajectory exist in general, given by (18).

Further analyzing this result, it is interesting to explore what happens if $\mathbf{a} = \mathbf{0}$, i.e., if the camera moves with a constant velocity. In that case, the last 3×1 block of $\mathbf{n}$ is zero, and α vanishes from Eq. (17). In turn, this means that α cannot be determined, and infinite solutions exist, given by $\mathbf{x} = \mathbf{A}^{+}\mathbf{b} + \alpha\mathbf{n}$. It is important to observe, however, that since the last three elements of $\mathbf{n}$ are zero, these solutions all have *the same* value for ${}^{B_0}\mathbf{g}$. In other words:

Result 4. When the camera is moving with a constant velocity, the orientation of the platform with respect to the horizontal plane can be uniquely determined, while an infinite number of solutions exist for the platform velocity and the feature positions. The physical interpretation of the infinite set of solutions is that the *scale* of the trajectory cannot be determined *[1]*.

We now turn our attention to identifying the minimal conditions, in terms of the number of features and images, that make it possible to obtain discrete solutions for the observable states. We start by noting that if only one image is available, the number of equations in Eq. (15) is always less than the number of unknowns. Moreover, in Appendix B.2.2, we show that if only $N = 2$ images are available, the dimension of the nullspace of $\mathbf{A}$ is at least 3, regardless of the number of features. The minimum number of images needed for a finite number of solutions is $N = 3$. For this case, by comparing the number of unknowns to the number of measurements, we conclude that the number of features must be at least $M = 2$. On the other hand, if only one feature is available ($M = 1$), we obtain the condition $N \geq 4$, as the lower bound on the number of images (Appendix B.2.2).

From the above we see that two *minimal problems* can be identified: (i) two features seen in three images, and (ii) one feature seen in four images. Both of these problems are minimal, since in the first case we use the minimum number of images possible, while in the second case we use the minimum number of features. In what follows, we describe the properties of these problems.

3.2.1 Two features seen in three images

In Appendix B.2.3, we prove that when two features are observed in three images, then in general it is $\dim(\mathcal{N}(\mathbf{A})) = 1$. The only exception is when the three camera positions and the two feature points lie in the same plane, in which case

$\dim(\mathcal{N}(\mathbf{A})) = 2$. Barring this singular configuration, the nullspace of the matrix $\mathbf{A}$ is spanned by:

$$\mathbf{n} = \begin{bmatrix} {}^{B_0}\mathbf{p}_1 - {}^{B_0}\mathbf{p}_{C_0} \\ {}^{B_0}\mathbf{p}_2 - {}^{B_0}\mathbf{p}_{C_0} \\ \frac{\Delta t_2({}^{B_0}\mathbf{p}_{C_1} - {}^{B_0}\mathbf{p}_{C_0})}{\Delta t_1(\Delta t_2 - \Delta t_1)} - \frac{\Delta t_1({}^{B_0}\mathbf{p}_{C_2} - {}^{B_0}\mathbf{p}_{C_0})}{\Delta t_2(\Delta t_2 - \Delta t_1)} \\ \frac{-2}{\Delta t_2 - \Delta t_1}\left(\frac{{}^{B_0}\mathbf{p}_{C_1} - {}^{B_0}\mathbf{p}_{C_0}}{\Delta t_1} - \frac{{}^{B_0}\mathbf{p}_{C_2} - {}^{B_0}\mathbf{p}_{C_0}}{\Delta t_2}\right) \end{bmatrix} \tag{20}$$

From the above expression, we see that if the camera is moving with a constant velocity, then the last three elements of $\mathbf{n}$ are zero, and the scale of the trajectory is undefined, while the attitude is uniquely determined. In all other cases, two distinct solutions exist.

3.2.2 One feature seen in four images

By following a similar analysis in Appendix B.2.4, we show that the results for the minimal problem where one feature is seen in four different images are analogous to the previous case. The results for both cases can be summarized as follows:

Result 5. When two features are observed in three images *or* one feature is observed in four images, then we can compute two distinct solutions for all observable quantities, using Eq. (18). There exist two singular cases when an infinite number of solutions exist: (i) when all camera positions and the feature(s) lie in the same plane, and (ii) when the camera is moving with a constant velocity. In the latter case, the orientation is uniquely defined, but the scale is unobservable.

Fig. 1 depicts the two solutions arising in an example case where one feature is observed in four images. It is interesting to observe that, even though this is difficult to see in the plot, the two trajectories are not coplanar. As a final remark, we note that if either more images, or more features become available in either one of the two minimal scenarios discussed, then the problems become over-determined, and thus a unique solution can be obtained in general (barring singular cases, such as the one described in Result 3). This makes it possible to determine the conditions for obtaining a unique solution.

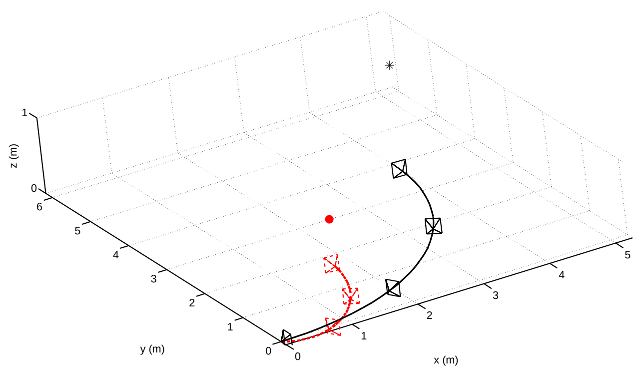


Figure 1: One feature observed in four camera images: This plot shows the two possible solutions for the trajectory and the feature position, both of which satisfy the measurements exactly. The two trajectories are not coplanar.

4 Unknown biases and/or IMU-Camera calibration

Up to this point, we have examined the case where the extrinsic calibration between the IMU and the camera, as well as the IMU biases, are known in advance. We here study the scenario where one or both of these quantities are not kwown, and need to be estimated. As shown in [1], both the extrinsic calibration as well as the IMU biases are observable, for general camera trajectories.

4.1 Unknown extrinsic calibration, known IMU biases

In this scenario, we consider the case where the transformation (rotation and translation) between the IMU and camera is not available while the IMU biases are known. By using the corrected IMU measurements, we can estimate the IMU orientation change ${}^{B_i}_{B_j}\mathbf{C}$ in the time interval $[t_i, t_j]$, as in Eq. (3). In addition, by using only feature observations, we can independently estimate the relative camera orientation during that interval, ${}^{C_i}_{C_j}\mathbf{C}$, by using image-based motion estimation algorithms such as [8]. We can then employ the following equation for the unknown ${}^{C}_{B}\mathbf{C}$:

$$ {}^{C_i}_{C_j}\mathbf{C} = {}^{C}_{B}\mathbf{C}\, {}^{B_i}_{B_j}\mathbf{C}\, {}^{B}_{C}\mathbf{C} \Rightarrow {}^{C_i}_{C_j}\mathbf{C}\, {}^{C}_{B}\mathbf{C} = {}^{C}_{B}\mathbf{C}\, {}^{B_i}_{B_j}\mathbf{C} \tag{21} $$

The above equation can be solved by transforming it into its equivalent unit-quaternion representation [16, 17]. Specifically, using this representation, we have [18]:

$$ \begin{aligned} &{}^{C_i}_{C_j}\bar{\mathbf{q}} \otimes {}^{C}_{B}\bar{\mathbf{q}} = {}^{C}_{B}\bar{\mathbf{q}} \otimes {}^{B_i}_{B_j}\bar{\mathbf{q}} \\ \Rightarrow &\mathcal{L}({}^{C_i}_{C_j}\bar{\mathbf{q}})\, {}^{C}_{B}\bar{\mathbf{q}} = \mathcal{R}({}^{B_i}_{B_j}\bar{\mathbf{q}})\, {}^{C}_{B}\bar{\mathbf{q}} && (22) \\ \Rightarrow &\big(\mathcal{L}({}^{C_i}_{C_j}\bar{\mathbf{q}}) - \mathcal{R}({}^{B_i}_{B_j}\bar{\mathbf{q}})\big)\, {}^{C}_{B}\bar{\mathbf{q}} = \mathbf{0} && (23) \end{aligned} $$

where for a 4×1 unit quaternion $\bar{\mathbf{q}}$:

$$ \bar{\mathbf{q}} = \begin{bmatrix} q_1 & q_2 & q_3 & q_4 \end{bmatrix}^T = \begin{bmatrix} \mathbf{q}^T & q_4 \end{bmatrix}^T \tag{24} $$

we define

$$ \mathcal{L}(\bar{\mathbf{q}}) = \begin{bmatrix} q_4\mathbf{I}_3 - \lfloor \mathbf{q}\times \rfloor & \mathbf{q} \\ -\mathbf{q}^T & q_4 \end{bmatrix} \tag{25} $$

$$ \mathcal{R}(\bar{\mathbf{q}}) = \begin{bmatrix} q_4\mathbf{I}_3 + \lfloor \mathbf{q}\times \rfloor & \mathbf{q} \\ -\mathbf{q}^T & q_4 \end{bmatrix} \tag{26} $$

Therefore, we have a linear system with ${}^{C}_{B}\bar{\mathbf{q}}$, the quaternion representing the IMU-camera rotation, as the unknown. By using all the available pairs of images, we construct an over-constrained linear system $\mathbf{B}\, {}^{C}_{B}\bar{\mathbf{q}} = \mathbf{0}$, whose least-squares solution ${}^{C}_{B}\bar{\mathbf{q}}$ is the right singular vector corresponding to the smallest singular value of $\mathbf{B}$. For the solution to be unique, at least two pairs of images where the system rotates about different axes are required [17]. Therefore, at least three images are needed, in which at least six features (to ensure a unique solution for ${}^{C_i}_{C_j}\mathbf{C}$) are tracked. After ${}^{C}_{B}\bar{\mathbf{q}}$ is estimated, we can directly obtain ${}^{C}_{B}\mathbf{C}$ [18]. Subsequently, by treating ${}^{C}\mathbf{p}_B$ in Eq. (9) as an extra unknown, we can construct a similar linear system as in Section 3, except that the vector of unknowns becomes (see Appendix A.2):

$$ \mathbf{x} = \begin{bmatrix} {}^{B_0}\mathbf{p}_1^T & \dots & {}^{B_0}\mathbf{p}_M^T & {}^{B_0}\mathbf{v}_0^T & {}^{C}\mathbf{p}_B^T & {}^{B_0}\mathbf{g}^T \end{bmatrix}^T \tag{27} $$

By solving this modified linear system, we can solve for the observable quantities O1-O3 discussed in Section 2 as well as the IMU-camera relative translation.

4.2 Unknown extrinsic calibration, unknown IMU biases

We next consider the case where we have no prior information about either the IMU-camera transformation or the IMU biases. We here assume that the gyroscope and accelerometer biases are constant over the short time interval used for initialization. Similar to the previous subsection, by using only feature observations, we can estimate the relative camera pose change ${}^{C_i}_{C_j}\mathbf{C}$ in the time interval $[t_i, t_j]$. Let ${}^{C_i}_{C_j}\bar{\mathbf{q}}$ and ${}^{B_i}_{B_j}\bar{\mathbf{q}}$ be the quaternions that represent the relative pose changes of the camera and IMU, respectively:

$$ {}^{C_i}_{C_j}\bar{\mathbf{q}} = \begin{bmatrix} {}^{C_i}_{C_j}\mathbf{q} \\ {}^{C_i}_{C_j}q_4 \end{bmatrix}, \; {}^{B_i}_{B_j}\bar{\mathbf{q}} = \begin{bmatrix} {}^{B_i}_{B_j}\mathbf{q} \\ {}^{B_i}_{B_j}q_4 \end{bmatrix} \tag{28} $$

To derive an equation that will enable us to obtain a closed-form solution, we will employ the relationship between a quaternion and its corresponding rotation matrix [18]:

$$ \mathbf{C}(\bar{\mathbf{q}}) = \mathbf{I}_3(2q_4^2 - 1) - 2q_4\lfloor \mathbf{q}\times \rfloor + 2\mathbf{q}\mathbf{q}^T \tag{29} $$

Using basic properties of rotation matrices, we have:

$$ {}^{C_i}_{C_j}\mathbf{C} = {}^{C}_{B}\mathbf{C}\, {}^{B_i}_{B_j}\mathbf{C}\, {}^{B}_{C}\mathbf{C} \tag{30} $$

$$\Rightarrow {}^{C_i}_{C_j}\mathbf{C} = {}^{C}_{B}\mathbf{C}\left(\mathbf{I}_3(2\,{}^{B_i}_{B_j}q_4^2 - 1) - 2\,{}^{B_i}_{B_j}q_4\lfloor {}^{B_i}_{B_j}\mathbf{q}\times\rfloor + 2\,{}^{B_i}_{B_j}\mathbf{q}\,{}^{B_i}_{B_j}\mathbf{q}^T\right){}^{B}_{C}\mathbf{C}$$
$$\Rightarrow {}^{C_i}_{C_j}\mathbf{C} = \mathbf{I}_3(2\,{}^{B_i}_{B_j}q_4^2 - 1) - 2\,{}^{B_i}_{B_j}q_4\lfloor {}^{C}_{B}\mathbf{C}\,{}^{B_i}_{B_j}\mathbf{q}\times\rfloor + 2({}^{C}_{B}\mathbf{C}\,{}^{B_i}_{B_j}\mathbf{q})({}^{C}_{B}\mathbf{C}\,{}^{B_i}_{B_j}\mathbf{q})^T$$

Converting to the equivalent quaternion representation, we obtain:

$$\begin{bmatrix} {}^{C_i}_{C_j}\mathbf{q} \\ {}^{C_i}_{C_j}q_4 \end{bmatrix} = \begin{bmatrix} {}^{C}_{B}\mathbf{C}\,{}^{B_i}_{B_j}\mathbf{q} \\ {}^{B_i}_{B_j}q_4 \end{bmatrix} \tag{31}$$

The above equation shows that the last element of the quaternion of the true IMU rotation (${}^{B_i}_{B_j}q_4$) can be estimated from feature observations (${}^{C_i}_{C_j}q_4$). If we let $\bar{\boldsymbol{\omega}}_{ij}$ denote the average true rotational velocity in the time interval $[t_i, t_j]$, and $\bar{\boldsymbol{\omega}}_{m_{ij}}$ represent the average measured rotational velocity in the same time, then, using the results of [18], we can write:

$$||\bar{\boldsymbol{\omega}}_{ij}\,t_{ij}|| \simeq 2\arccos({}^{C_i}_{C_j}q_4)$$
$$\Rightarrow ||\bar{\boldsymbol{\omega}}_{m_{ij}} - \mathbf{b}_g|| \simeq \frac{2\arccos({}^{C_i}_{C_j}q_4)}{t_{ij}}$$
$$\Rightarrow ||\bar{\boldsymbol{\omega}}_{m_{ij}}||^2 - 2\,\bar{\boldsymbol{\omega}}_{m_{ij}}^T\,\mathbf{b}_g + ||\mathbf{b}_g||^2 = \frac{4\arccos({}^{C_i}_{C_j}q_4)^2}{t_{ij}^2}$$

where $t_{ij} = t_j - t_i$. Choosing $j = i+1, \forall i = 1\ldots(N-1)$, we obtain $(N-1)$ equations of the above form. By subtracting the first one from all other equations, we get the linear system:

$$\begin{array}{rcl} (2\bar{\boldsymbol{\omega}}_{m_{23}} - 2\bar{\boldsymbol{\omega}}_{m_{12}})^T\mathbf{b}_g & = & d_2 \\ (2\bar{\boldsymbol{\omega}}_{m_{34}} - 2\bar{\boldsymbol{\omega}}_{m_{12}})^T\mathbf{b}_g & = & d_3 \\ \vdots & = & \vdots \\ (2\bar{\boldsymbol{\omega}}_{m_{(N-1)N}} - 2\bar{\boldsymbol{\omega}}_{m_{12}})^T\mathbf{b}_g & = & d_{N-1} \end{array} \tag{32}$$

where

$$d_i = ||\bar{\boldsymbol{\omega}}_{m_{i(i+1)}}||^2 - \frac{4\arccos({}^{C_i}_{C_{i+1}}q_4)^2}{t_{i(i+1)}^2} - ||\bar{\boldsymbol{\omega}}_{m_{12}}||^2 + \frac{4\arccos({}^{C_1}_{C_2}q_4)^2}{t_{12}^2} \tag{33}$$

Note that all quantities except for $\mathbf{b}_g$ are known from the IMU and camera measurements. Therefore, the bias in the gyroscope measurements, $\mathbf{b}_g$, can be estimated by solving for the least-squares solution of the linear system constructed in Eq. (32). For a unique solution, at least five images are needed. After we have computed the gyro bias, we can obtain the corrected gyroscope measurements, and proceed to solve for the IMU-camera rotation matrix as presented in Section 4.1. Subsequently, we can solve for the remaining quantities linearly. Specifically, by taking into account the accelerometer bias, we rewrite Eq. (7) as follows (see Appendix A):

$${}^{B_0}\mathbf{p}_{B_i} = {}^{B_0}\mathbf{v}_0\Delta t_i + {}^{B_0}\mathbf{g}\frac{\Delta t_i^2}{2} - \mathbf{S}(t_i)\mathbf{b}_a + \mathbf{s}(t_i) \tag{34}$$

$$\text{where} \qquad \mathbf{S}(t_i) = \int_{t_0}^{t_i}\int_{t_0}^{\tau} {}^{B_0}_{B}\mathbf{C}(\varsigma)\,d\varsigma d\tau \tag{35}$$

By treating ${}^{C}\mathbf{p}_B$ and $\mathbf{b}_a$ as unknowns, we can construct a similar linear system as in Section 3, except that the vector $\mathbf{x}$ in Eq. (11) is extended as (see Appendix A.3):

$$\mathbf{x} = \begin{bmatrix} {}^{B_0}\mathbf{p}_1^T & \ldots & {}^{B_0}\mathbf{p}_M^T & {}^{B_0}\mathbf{v}_0^T & \mathbf{b}_a^T & {}^{C}\mathbf{p}_B^T & {}^{B_0}\mathbf{g}^T \end{bmatrix}^T \tag{36}$$

and $\mathbf{A}_{ij}$ becomes:

$$\mathbf{J}_{ij}\begin{bmatrix} \ldots & \mathbf{I}_3 & \ldots & -\Delta t_i\mathbf{I}_3 & \mathbf{S}(t_i) & {}^{B_0}_{B_i}\mathbf{C}{}^{B}_{C}\mathbf{C} & -\frac{\Delta t_i^2}{2}\mathbf{I}_3 \end{bmatrix}$$

$$\text{where} \quad \mathbf{J}_{ij} = \begin{bmatrix} 1 & 0 & -u_{ij} \\ 0 & 1 & -v_{ij} \end{bmatrix} {}^{C}_{B}\mathbf{C}\ {}^{B_i}_{B_0}\mathbf{C} \tag{37}$$

$$\text{and} \quad \mathbf{b}_{ij} = \mathbf{J}_{ij}\ \mathbf{s}(t_i) \tag{38}$$

Solving this modified linear system, we can estimate the accelerometer bias and IMU-camera relative translation in addition to the observable quantities O1-O3 in Section 2.

5 Solution in the presence of noise

The analysis in Section 3 did not account for the presence of noise, which will inevitably exist in any practical application. In this section, we show how noisy measurements can be employed for obtaining an estimate of the observable quantities of the system. Specifically, in Section 5.1 we describe two least-squares solutions for the linear system in Eq. (15), which can be found in closed form. Then, in Section 5.2 we present a maximum-likelihood estimator (MLE), which employs a probabilistic modeling of the measurement noise, to obtain statistically optimal estimates.

5.1 Linear system solutions

As shown in Sections 3 and 4, the feature positions, IMU velocity, local gravity vector, accelerometer bias, and camera-IMU translation can be found by solving a linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$. When $\mathbf{A}$ and $\mathbf{b}$ are constructed using many noisy sensor measurements, the system is an over-determined one, and no exact solution can be found. Instead, we can compute a least-squares solution, i.e., we can minimize the cost function $||\mathbf{A}\mathbf{x}-\mathbf{b}||_2$. It is a well-known result that the optimal value of $\mathbf{x}$ for this problem is given by $\mathbf{x}^\star = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$, which is identical to Eq. (16). Note however, that this solution does not take advantage of the fact that the norm of the gravitational acceleration vector may be known in advance. To exploit this additional information, we can formulate a *constrained* least-squares problem:

$$\begin{aligned} \text{minimize } \; & ||\mathbf{A}\mathbf{x}-\mathbf{b}||_2 = \left|\left|[\mathbf{A}_1 \;\; \mathbf{A}_2]\begin{bmatrix}\mathbf{x}_1 \\ {}^{B_0}\mathbf{g}\end{bmatrix} - \mathbf{b}\right|\right|_2 \\ \text{subject to } \; & ||{}^{B_0}\mathbf{g}||_2 = g \end{aligned} \tag{39}$$

where g is the known value of the gravitational acceleration, $\mathbf{x}_1$ is a vector comprising the landmark positions and the IMU velocity [Eq. (11)], as well as IMU-camera translation and accelerometer bias if unknown [Eq. (27), (36)], and the partitioning of $\mathbf{A}$ is compatible with that of $\mathbf{x}$. The above problem is a quadratically-constrained least-squares problem. In the remainder of this section, we use $\mathbf{x}_2 = {}^{B_0}\mathbf{g}$ in the derivations for clarity. Using the Lagrange multiplier approach, the solution to the above problem can be found by finding the stationary points of the function:

$$\min L(\mathbf{x},\lambda) = \left|\left|[\mathbf{A}_1 \;\; \mathbf{A}_2]\begin{bmatrix}\mathbf{x}_1 \\ \mathbf{x}_2\end{bmatrix} - \mathbf{b}\right|\right|_2^2 - \lambda(\mathbf{x}_2^T\mathbf{x}_2 - g^2) \tag{40}$$

After taking partial derivatives, we obtain:

$$\frac{\partial L}{\partial \mathbf{x}_1}^T = 2\mathbf{A}_1^T\mathbf{A}_1\mathbf{x}_1 + 2\mathbf{A}_1^T\mathbf{A}_2\mathbf{x}_2 - 2(\mathbf{b}^T\mathbf{A}_1)^T = \mathbf{0} \tag{41}$$

$$\frac{\partial L}{\partial \mathbf{x}_2}^T = 2\mathbf{A}_2^T\mathbf{A}_2\mathbf{x}_2 + 2\mathbf{A}_2^T\mathbf{A}_1\mathbf{x}_1 - 2(\mathbf{b}^T\mathbf{A}_2)^T - 2\lambda\mathbf{x}_2 = \mathbf{0} \tag{42}$$

$$\frac{\partial L}{\partial \lambda} = \mathbf{x}_2^T\mathbf{x}_2 - g^2 = 0 \tag{43}$$

$$\Rightarrow \left\{ \begin{array}{rcl} \mathbf{A}_1^T\mathbf{A}_1\mathbf{x}_1 + \mathbf{A}_1^T\mathbf{A}_2\mathbf{x}_2 & = & \mathbf{A}_1^T\mathbf{b} \\ \mathbf{A}_2^T\mathbf{A}_1\mathbf{x}_1 + (\mathbf{A}_2^T\mathbf{A}_2 - \lambda\mathbf{I})\mathbf{x}_2 & = & \mathbf{A}_2^T\mathbf{b} \\ \mathbf{x}_2^T\mathbf{x}_2 & = & g^2 \end{array}\right. \tag{44}$$

Solving the first of the above equations for $\mathbf{x}_1$, we get:

$$\mathbf{x}_1 = -(\mathbf{A}_1^T\mathbf{A}_1)^{-1}\mathbf{A}_1^T\mathbf{A}_2\mathbf{x}_2 + (\mathbf{A}_1^T\mathbf{A}_1)^{-1}\mathbf{A}_1^T\mathbf{b} \tag{45}$$

Using the above equation in the second equation in (44), we have:

$$\begin{gathered} \mathbf{D}\mathbf{x}_2 - \lambda\mathbf{x}_2 = \mathbf{d} \\ \mathbf{x}_2^T\mathbf{x}_2 = g^2 \end{gathered} \tag{46}$$

where

$$\mathbf{D} = \mathbf{A}_2^T\left(\mathbf{I} - \mathbf{A}_1(\mathbf{A}_1^T\mathbf{A}_1)^{-1}\mathbf{A}_1^T\right)\mathbf{A}_2 \tag{47}$$

$$\mathbf{d} = \mathbf{A}_2^T\left(\mathbf{I} - \mathbf{A}_1(\mathbf{A}_1^T\mathbf{A}_1)^{-1}\mathbf{A}_1^T\right)\mathbf{b} \tag{48}$$

We thus see that to determine $\mathbf{x}$, we can first determine $\mathbf{x}_2$ from (46), and then substitute in (45) to find $\mathbf{x}_1$. Finding the solution to (46) is equivalent to finding the solution to the following optimization problem:

$$\begin{aligned} \text{minimize} \;\; & \mathbf{x}_2^T \mathbf{D} \mathbf{x}_2 - 2\mathbf{d}^T \mathbf{x}_2 \\ \text{subject to} \;\; & \mathbf{x}_2^T \mathbf{x}_2 = g \end{aligned} \tag{49}$$

The equivalence arises from the fact that the equations in (46) are obtained by forming the Lagrange function of the problem in (49), and setting its derivatives equal to zeros. Furthermore, in [19, 20], it is shown that the optimal value of $\mathbf{x}_2$ in (49) is the same as the optimal value of $\mathbf{x}_2$ in the following problem:

$$\begin{aligned} \text{minimize} \;\; & \lambda \\ \text{subject to} \;\; & \mathbf{D}\mathbf{x}_2 = \lambda \mathbf{x}_2 + \mathbf{d} \\ & \mathbf{x}_2^T \mathbf{x}_2 = g^2 \end{aligned} \tag{50}$$

Theorem 5.1 in [19] states that if $\{\lambda, \mathbf{x}_2\}$ satisfy the constraint equations in (50), then there exists a vector $\mathbf{e} \neq \mathbf{0}$ such that the following holds:

$$(\mathbf{D} - \lambda \mathbf{I}_3)^2 \mathbf{e} = \frac{1}{g^2} \mathbf{d}\mathbf{d}^T \mathbf{e} \tag{51}$$

The above equation describes a *quadratic eigenvalue problem*. From this equation, we conclude that:

$$\det\left((\mathbf{D} - \lambda \mathbf{I}_3)^2 - \frac{1}{g^2} \mathbf{d}\mathbf{d}^T \right) = 0 \tag{52}$$

The matrix whose determinant we compute above is a 3×3 matrix, whose elements are quadratic polynomials in λ. Therefore, (52) is a sixth-order polynomial equation in λ. Out of the six possible roots, we must keep the smallest real one, as this is the solution to (50). To find the smallest root, one can simply compute all roots (which can be done with very low computational cost and without any prior information), and choose the minimum real one. Once the minimum λ has been determined, we can solve (46) to obtain $\mathbf{x}_2$ as follows [19, Theorem 5.2]:

$$\mathbf{x}_2 = (\mathbf{D} - \lambda \mathbf{I})^{-1} \mathbf{d} \tag{53}$$

In summary, the constrained least-squares solution is given by:

$$\mathbf{x}^\star = \begin{bmatrix} -(\mathbf{A}_1^T \mathbf{A}_1)^{-1} \mathbf{A}_1^T \mathbf{A}_2 {}^{B_0}\mathbf{g}^\star + (\mathbf{A}_1^T \mathbf{A}_1)^{-1} \mathbf{A}_1^T \mathbf{b} \\ {}^{B_0}\mathbf{g}^\star \end{bmatrix} \tag{54}$$

with ${}^{B_0}\mathbf{g}^\star = (\mathbf{D} - \lambda \mathbf{I}_3)^{-1} \mathbf{d}$, where $\mathbf{D}, \mathbf{d}$ are defined in Eq. (47)-(48) and λ is the smallest solution to the equation:

$$\det\left((\mathbf{D} - \lambda \mathbf{I}_3)^2 - \frac{1}{g^2} \mathbf{d}\mathbf{d}^T \right) = 0 \tag{55}$$

Our tests have shown that when noise is present, using the known gravity information (i.e., employing the constrained-least-squares solution (54) instead of the unconstrained one (16)) results in substantially improved estimation accuracy (see Section 6).

5.2 Maximum Likelihood Estimator

The least-squares solutions presented in the previous section offer the advantage of providing the result in closed form. However, the minimization of the cost function $||\mathbf{A}\mathbf{x} - \mathbf{b}||_2$ does not properly account for the statistical properties of the noise, which affects both $\mathbf{A}$ and $\mathbf{b}$. To properly account for the noise in the measurements, we formulate a MLE for estimating a parameter vector $\boldsymbol{\theta}$ comprising (i) the IMU state (position, velocity, orientation) at each time instant where an image is recorded, (ii) the positions of all features, $\mathbf{p}_j$, $j = 1 \ldots M$, (iii) the IMU biases, if not known, and (iv) the camera-to-IMU transformation, if not known. Quantities (i)-(ii) are expressed with respect to a frame whose origin coincides with the origin of the initial IMU frame, while its z-axis is aligned with the direction of gravity.

Following standard practice, we model the image measurement noise vector, $\mathbf{n}_{ij}$, as a Gaussian zero-mean random variable, with covariance matrix $\mathbf{R}_{ij}$. Moreover, we use the IMU measurements in the time interval $[t_i, t_{i+1}]$ to compute the change in the IMU state:

$$\mathbf{x}_{\mathrm{IMU}_{i+1}} = \mathbf{f}(\mathbf{x}_{\mathrm{IMU}_i}, \boldsymbol{\omega}_m, \mathbf{a}_m) + \mathbf{w}_i \tag{56}$$

where $\mathbf{w}_i$ is a noise vector, modelled as zero-mean, Gaussian random variable with covariance matrix $\mathbf{Q}_i$. The function $\mathbf{f}$ and the covariance matrix $\mathbf{Q}_i$ are computed using numerical integration of the continuous-time motion equations [14].

Maximizing the likelihood of the measurements is equivalent to maximizing the log-likelihood, which, in turn, is equivalent to minimizing the cost function:

$$c(\boldsymbol{\theta}) = \sum_{i,j \in \mathcal{S}_m} \left|\left|\mathbf{z}_{ij} - \mathbf{h}(\mathbf{x}_{\mathrm{IMU}_i}, \mathbf{p}_j)\right|\right|_{\mathbf{R}_{ij}} + \sum_{i=0}^{N-2} \left|\left|\mathbf{x}_{\mathrm{IMU}_{i+1}} - \mathbf{f}(\mathbf{x}_{\mathrm{IMU}_i}, \boldsymbol{\omega}_m, \mathbf{a}_m)\right|\right|_{\mathbf{Q}_i} \tag{57}$$

where $\mathbf{h}(\cdot)$ is the function describing the perspective measurement model [Eq. (8)], and $\left|\left|\mathbf{u}\right|\right|_{\mathbf{M}}$ denotes the matrix-weighted norm $\mathbf{u}^T\mathbf{M}^{-1}\mathbf{u}$.

The cost function $c(\boldsymbol{\theta})$ is nonlinear, and its minimization is carried out iteratively, by application of the Gauss-Newton method [21]. At the ℓ-th iteration of this method, a correction, $\Delta\boldsymbol{\theta}^{(\ell)}$, to the current estimate, $\boldsymbol{\theta}^{(\ell)}$, is computed by minimizing the second-order Taylor-series approximation of the cost function:

$$c(\boldsymbol{\theta}^{(\ell)} + \Delta\boldsymbol{\theta}) \simeq c(\boldsymbol{\theta}^{(\ell)}) + \mathbf{b}^{(\ell)T}\Delta\boldsymbol{\theta} + \frac{1}{2}\Delta\boldsymbol{\theta}^T\mathbf{A}^{(\ell)}\Delta\boldsymbol{\theta} \tag{58}$$

where $\mathbf{b}^{(\ell)} = \nabla c(\boldsymbol{\theta}^{(\ell)})$ and $\mathbf{A}^{(\ell)} = \nabla^2 c(\boldsymbol{\theta}^{(\ell)})$ denote the Jacobian and Hessian of c with respect to $\boldsymbol{\theta}$, evaluated at the current iteration, $\boldsymbol{\theta}^{(\ell)}$. Specifically, $\mathbf{b}^{(\ell)}$ is given by:

$$\mathbf{b}^{(\ell)} = \mathbf{b}^h_{\mathcal{S}_m}(\boldsymbol{\theta}^{(\ell)}) + \mathbf{b}^f_{0:N-1}(\boldsymbol{\theta}^{(\ell)}) \tag{59}$$

where the two terms appearing in the sum are due to the feature measurements indexed by $\mathcal{S}_m$, and the IMU measurements in the time interval $[t_0, t_{N-1}]$, respectively:

$$\mathbf{b}^h_{\mathcal{S}_m}(\boldsymbol{\theta}^{(\ell)}) = -\sum_{(i,j)\in\mathcal{S}_m} \mathbf{H}_{ij}^{(\ell)T}\mathbf{R}_{ij}^{-1}\left(\mathbf{z}_{ij} - \mathbf{h}(\mathbf{x}^{(\ell)}_{\mathrm{IMU}_i}, \mathbf{p}^{(\ell)}_j)\right) \tag{60}$$

$$\mathbf{b}^f_{0:N-1}(\boldsymbol{\theta}^{(\ell)}) = \sum_{i=0}^{N-2} \mathbf{G}_i^{(\ell)T}\mathbf{Q}_i^{-1}\left(\mathbf{x}^{(\ell)}_{\mathrm{IMU}_{i+1}} - \mathbf{f}(\mathbf{x}^{(\ell)}_{\mathrm{IMU}_i}, \boldsymbol{\omega}_m, \mathbf{a}_m)\right) \tag{61}$$

In these expressions, the matrices $\mathbf{H}_{ij}^{(\ell)}$ and $\mathbf{G}_i^{(\ell)}$ are the Jacobians of the measurement function, $\mathbf{h}(\mathbf{x}_{\mathrm{IMU}_i}, \mathbf{p}_j)$, and of the function $\mathbf{g}_i = \mathbf{x}_{\mathrm{IMU}_{i+1}} - \mathbf{f}(\mathbf{x}_{\mathrm{IMU}_i}, \boldsymbol{\omega}_m, \mathbf{a}_m)$, with respect to $\boldsymbol{\theta}$, evaluated at the current iterate $\boldsymbol{\theta}^{(\ell)}$. Since both the measurement function and the motion model involve only a few states (either one camera pose and one feature, or two consecutive camera poses), the structure of both $\mathbf{H}_{ij}^{(\ell)}$ and $\mathbf{G}_i^{(\ell)}$ is very sparse.

In the Gauss-Newton method, and for small-residual problems, the Hessian matrix can be well approximated by [21]:

$$\mathbf{A}^{(\ell)} = \sum_{(i,j)\in\mathcal{S}_m} \mathbf{H}_{ij}^{(\ell)T}\mathbf{R}_{ij}^{-1}\mathbf{H}_{ij}^{(\ell)} + \sum_{i=0}^{N-2}\mathbf{G}_i^{(\ell)T}\mathbf{Q}_i^{-1}\mathbf{G}_i^{(\ell)} \tag{62}$$

The value of $\Delta\boldsymbol{\theta}^{(\ell)}$ minimizing the cost function (58) is found by solving the linear system:

$$\mathbf{A}^{(\ell)}\Delta\boldsymbol{\theta}^{(\ell)} = -\mathbf{b}^{(\ell)} \tag{63}$$

It is important to note that, since the initial position of the IMU and the rotation about the gravity vector are unobservable, the Hessian of the system that is solved at each Gauss-Newton iteration is singular. For this reason, we use regularization, in order to avoid numerical instabilities. In our testing we have observed that if the closed-form solution described in Section 5.1 is used to provide an initial guess for the iterations, the convergence is rapid and requires only a few (typically less than 10) iterations.

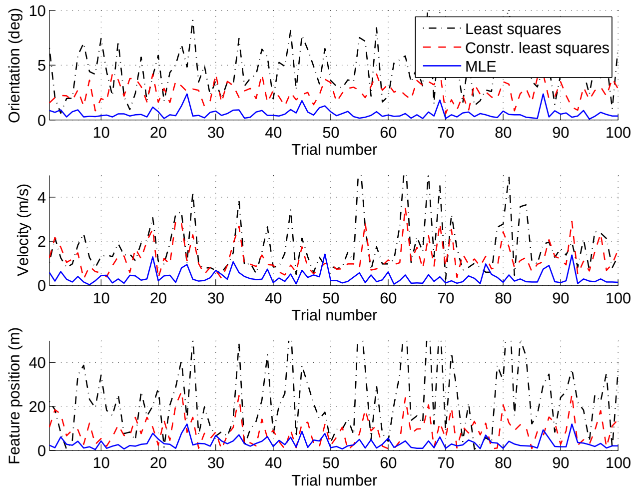


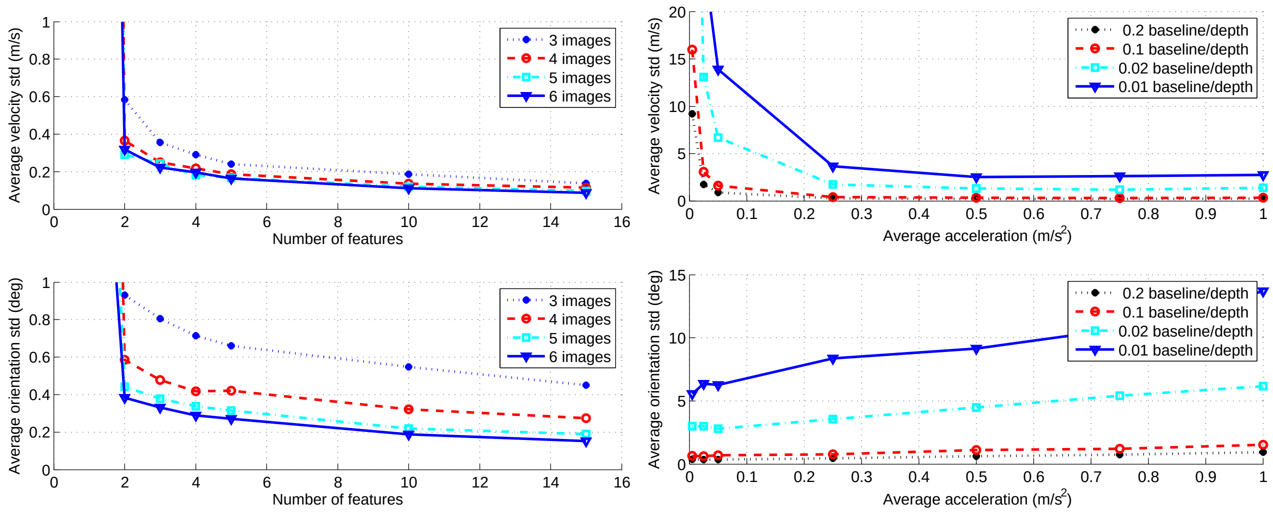
Figure 2: Comparison of the errors of the closed-form solutions vs. the MLE.

6 Results

We now present the results of Monte-Carlo simulation trials, which illustrate the accuracy of the methods described in the preceding sections, and the dependence of this accuracy on several parameters of interest. In all the results presented here, the accelerometer and gyroscope measurements are corrupted by independent zero-mean Gaussian noise processes, with standard deviation 0.05 m/s^2 and 0.05 rad/s, respectively. The image-noise standard deviation is set to 1 pixel, in a camera with a field of view of $60°$ and focal length equal to 500 pixels. IMU measurements are available at 100 Hz, while images are recorded at 1 Hz. Finally, in each trial, a trajectory is generated using a random initial orientation and velocity, as well as randomly generated acceleration and rotational velocities at each time-step. Similarly, the point features are randomly placed in the scene, such that they can be seen by all the camera poses.

We first compare the accuracy attained using the closed-form solutions described in Section 5.1, against that of the MLE. Specifically, we consider the case where four features are seen in four images, and compute the errors of three methods: (M1) unconstrained least squares, (M2) gravity-constrained least squares, and (M3) MLE. Fig. 2 shows the norm of the errors in the orientation, velocity, and feature position, in 100 Monte-Carlo trials using different random trajectories. The average errors for the methods M1-M3 are $\{4.5876, 2.4434, 0.5413\}$ degrees for orientation; $\{1.6487, 1.1579, 0.3464\}$ m/s for velocity; and $\{21.3725, 8.1121, 3.1100\}$ m for the feature positions. These results show that using the known gravity information leads to improved accuracy, compared to the case where no gravity is used, and the improvement in the error ranges from approximately 30% to 60%. On the other hand, the MLE outperforms both closed-form methods (as expected), due to the fact that it employs a probabilistic modelling of the measurement noise. Note, however, that the success of the MLE relies on having a good initial guess (provided from method M2) for the iterative minimization. If an inaccurate guess is provided as the starting point, we have observed divergence in more than 10% of the cases. This demonstrates the practical utility of using a closed-form solution.

We next examine the effect of different parameters on the estimation accuracy, starting with the impact of varying the number of images, N, and of features, M. For each selection of (N, M), we run 100 Monte Carlo trials with randomly generated trajectories and feature positions, and plot the average standard deviation of the orientation and velocity errors in Fig. 3(a). As expected, the standard deviation of the errors monotonically decreases as more images or more features become available. An interesting observation is that the improvement follows a law of diminishing return: for instance, while using more than 10 features seems to offer little benefit, increasing the number of features from one to two results in a substantial accuracy improvement. We next turn our attention to the effects of varying the "jerkiness" of the IMU motion and the features' distance relative to the the length of the camera trajectory (the baseline). To this end, we conduct a set of Monte-Carlo trials in which the control parameters are (i) the average magnitude of the random acceleration at each step (ranging in the interval $(0, 1]$ m/s^2), and (ii) the average length of the trajectory as a proportion of the average feature depth (ranging in the interval $[1/100, 1/10]$). In all trials, four images and 10 landmarks are used. For each setting we perform 100 Monte-Carlo trials, and the results are shown in Fig. 3(b). The key observation that can be made is that as the motion profile becomes "smoother," (i.e., the camera undergoes very small accelerations) the velocity estimation accuracy degrades significantly, while the accuracy of the attitude estimates remains almost constant (in fact, a small improvement can be observed). These results agree with the theoretical findings of Result 4: the closer to zero the acceleration is, the



(a) Accuracy as a function of the number of features and images

(b) Accuracy as a function of the acceleration profile and feature distance.

Figure 3: Effect of different parameters on the estimation accuracy.

less accurate the velocity estimates become. Additionally, we can also observe that as the features are placed at larger distances, the accuracy of the estimates degrades. This makes sense intuitively: at the limit where the features are at an infinite distance, their observations offer no information about the platform's displacement.

In all the results presented to this point, the IMU biases and IMU-camera transformation were treated as known constants. We now investigate the effects (in terms of estimation accuracy) of having no prior information for one or both of these quantities. For the results shown in Fig. 4, eight images and at least six features are used. From this plot, it can be observed that, as expected, more prior information will improve the accuracy of the estimates, with estimates having the smallest uncertainty when both the IMU biases and the IMU-camera transformation are known in advance, and largest when none of these are known. More interestingly, we observe that knowledge of the IMU biases appears to be more useful as it significantly reduces the uncertainty of the estimates. On the other hand, knowledge of the IMU-camera transformation results in only a small improvement of the estimates' uncertainty.

7 Conclusion

In this report we present closed-form solutions for the problem of vision-aided inertial navigation. We show that in the case of known biases and IMU-camera transformation, it is possible to analytically determine all other observable quantities by solving a quadratically-constrained linear system, which can, in general, have two discrete solutions. Through a detailed analysis of that particular system's properties, we identify the minimal problems, and show that if (i) one feature is seen in four images, or (ii) two features are seen in three images, then we can compute two solutions in closed form. Moreover, we show that two solutions exist when the IMU/camera system is moving with a constant acceleration. Furthermore, in the cases where IMU biases and/or IMU-camera transformation are unknown, we show that all observable quantities can be determined by solving a sequence of linear systems. To properly treat the measurement noise, we formulate an iterative MLE that takes the closed-form solution as the starting point. Through extensive Monte-Carlo simulations, the proposed algorithms are shown to be suitable for use in applications where state estimation has to be performed without any prior information.

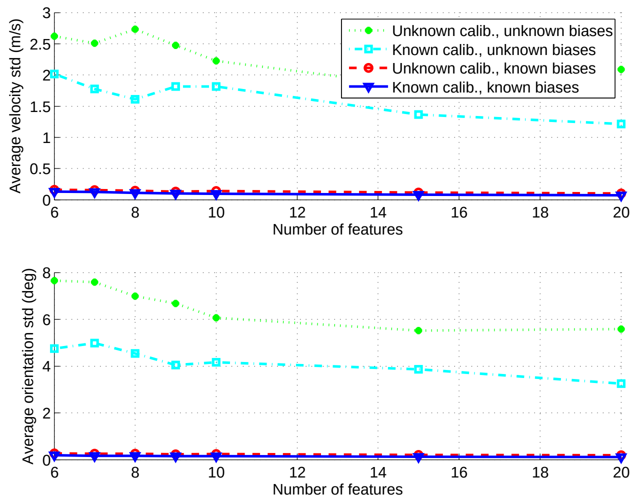


Figure 4: Accuracy as a function of the availability of prior knowledge.

A Linear system formulation

We here show how the linear system can be formulated in the general case, where the IMU-to-camera transformation and/or the IMU accelerometer bias is unknown. To obtain a closed-form solution, we will ignore the measurement noise. After removing the noise components, the measurements of the IMU gyroscopes and accelerometers are given by the following equations [13]:

$$\boldsymbol{\omega}_m(t) = {}^{B}\boldsymbol{\omega}(t) + \mathbf{b}_g \tag{64}$$

$$\mathbf{a}_m(t) = {}^{B}_{G}\mathbf{C}(t)({}^{G}\mathbf{a}(t) - {}^{G}\mathbf{g}) + \mathbf{b}_a \tag{65}$$

As discussed in Sections 3 and 4, the gyroscope measurement bias $\mathbf{b}_g$ is either known or independently estimated prior to formulating the linear system. After having the corrected gyroscope measurements, to estimate the IMU orientation change in the interval $[t_0, t_i]$, we integrate the following differential equation [14]:

$${}^{B_0}_{B}\dot{\mathbf{C}}(t) = -{}^{B_0}_{B}\mathbf{C}(t)\lfloor {}^{B}\boldsymbol{\omega}(t)\times\rfloor \ , \ {}^{B_0}_{B}\mathbf{C}(t_0) = \mathbf{I}_3 \tag{66}$$

The true acceleration of the IMU at time t in the global frame is given by:

$${}^{G}\mathbf{a}(t) = {}^{G}_{B}\mathbf{C}(t)(\mathbf{a}_m - \mathbf{b}_a) + {}^{G}\mathbf{g} \tag{67}$$

The global velocity of the IMU at time t can be found by integrating the above global acceleration:

$${}^{G}\mathbf{v}(t) = {}^{G}\mathbf{v}(t_0) + \int_{t_0}^{t} {}^{G}\mathbf{a}(\tau)\, d\tau \tag{68}$$

$$\Rightarrow {}^{G}\mathbf{v}(t) = {}^{G}\mathbf{v}(t_0) + {}^{G}\mathbf{g}(t - t_0) + \int_{t_0}^{t} {}^{G}_{B}\mathbf{C}(\tau)(\mathbf{a}_m(\tau) - \mathbf{b}_a)\, d\tau \tag{69}$$

where ${}^{G}\mathbf{v}(t_0)$ is the initial velocity. Using ${}^{G}_{B}\mathbf{C}(t) = {}^{G}_{B_0}\mathbf{C}\,{}^{B_0}_{B}\mathbf{C}(t)$, we have:

$${}^{G}\mathbf{v}(t) = {}^{G}\mathbf{v}(t_0) + {}^{G}\mathbf{g}(t - t_0) + {}^{G}_{B_0}\mathbf{C}\left(\int_{t_0}^{t} {}^{B_0}_{B}\mathbf{C}(\tau)\mathbf{a}_m(\tau)\, d\tau\right) - {}^{G}_{B_0}\mathbf{C}\left(\int_{t_0}^{t} {}^{B_0}_{B}\mathbf{C}(\tau)\, d\tau\right)\mathbf{b}_a \tag{70}$$

Similarly, the global position can be found by integrating the global velocity:

$$\begin{aligned}
{}^{G}\mathbf{p}(t) &= {}^{G}\mathbf{p}(t_0) + \int_{t_0}^{t} {}^{G}\mathbf{v}(\tau)\, d\tau && (71)\\
&= {}^{G}\mathbf{p}(t_0) + {}^{G}\mathbf{v}(t_0)(t - t_0) + {}^{G}\mathbf{g}\frac{(t - t_0)^2}{2} + {}^{G}_{B_0}\mathbf{C}\int_{t_0}^{t}\int_{t_0}^{\tau} {}^{B_0}_{B}\mathbf{C}(\varsigma)(\mathbf{a}_m(\varsigma) - \mathbf{b}_a)\, d\varsigma d\tau && (72)\\
&= {}^{G}\mathbf{p}(t_0) + {}^{G}\mathbf{v}(t_0)(t - t_0) + {}^{G}\mathbf{g}\frac{(t - t_0)^2}{2} + {}^{G}_{B_0}\mathbf{C}\left(\int_{t_0}^{t}\int_{t_0}^{\tau} {}^{B_0}_{B}\mathbf{C}(\varsigma)\mathbf{a}_m(\varsigma)\, d\varsigma d\tau\right) - {}^{G}_{B_0}\mathbf{C}\left(\int_{t_0}^{t}\int_{t_0}^{\tau} {}^{B_0}_{B}\mathbf{C}(\varsigma)\, d\varsigma d\tau\right)\mathbf{b}_a && (73)\\
&= {}^{G}\mathbf{p}(t_0) + {}^{G}\mathbf{v}(t_0)(t - t_0) + {}^{G}\mathbf{g}\frac{(t - t_0)^2}{2} + {}^{G}_{B_0}\mathbf{C}\,\mathbf{s}(t) - {}^{G}_{B_0}\mathbf{C}\,\mathbf{S}(t)\mathbf{b}_a && (74)
\end{aligned}$$

where

$$\mathbf{s}(t) = \int_{t_0}^{t}\int_{t_0}^{\tau} {}^{B_0}_{B}\mathbf{C}(\varsigma)\mathbf{a}_m(\varsigma)\, d\varsigma d\tau \tag{75}$$

$$\mathbf{S}(t) = \int_{t_0}^{t}\int_{t_0}^{\tau} {}^{B_0}_{B}\mathbf{C}(\varsigma)\, d\varsigma d\tau \tag{76}$$

At $t = t_i$ and using $\Delta t_i = t_i - t_0$, we have:

$${}^{G}\mathbf{p}(t_i) = {}^{G}\mathbf{p}(t_0) + {}^{G}\mathbf{v}(t_0)\Delta t_i + {}^{G}\mathbf{g}\frac{\Delta t_i^2}{2} + {}^{G}_{B_0}\mathbf{C}\,\mathbf{s}(t_i) - {}^{G}_{B_0}\mathbf{C}\,\mathbf{S}(t_i)\mathbf{b}_a \tag{77}$$

Next, we note that the IMU position at time t_i with respect to $\{B_0\}$ is given by ${}^{B_0}\mathbf{p}_{B_i} = {}^{G}_{B_0}\mathbf{C}^T({}^G\mathbf{p}(t_i) - {}^G\mathbf{p}(t_0))$. Using this expression, we obtain:

$$ {}^{B_0}\mathbf{p}_{B_i} = {}^{B_0}\mathbf{v}_0\Delta t_i + {}^{B_0}\mathbf{g}\frac{\Delta t_i^2}{2} + \mathbf{s}(t_i) - \mathbf{S}(t_i)\mathbf{b}_a \tag{78} $$

Using Eq. (9), we obtain:

$$ {}^{C_i}\mathbf{p}_j = {}^{C}_{B}\mathbf{C}\ {}^{B_i}_{B_0}\mathbf{C}\left({}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{v}_0\Delta t_i - {}^{B_0}\mathbf{g}\frac{\Delta t_i^2}{2} + \mathbf{S}(t_i)\mathbf{b}_a - \mathbf{s}(t_i)\right) + {}^{C}\mathbf{p}_B \tag{79} $$

Note that, similar to $\mathbf{b}_g$, the rotation matrix between the IMU and camera frames, ${}^{C}_{B}\mathbf{C}$, is either known or independently estimated prior to formulating the linear system. Proceeding further, we employ Eq. (8), to obtain:

$$ \begin{bmatrix} 1 & 0 & -u_{ij} \\ 0 & 1 & -v_{ij} \end{bmatrix} {}^{C_i}\mathbf{p}_j = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{80} $$

$$ \Rightarrow \begin{bmatrix} 1 & 0 & -u_{ij} \\ 0 & 1 & -v_{ij} \end{bmatrix} \left({}^{C}_{B}\mathbf{C}\ {}^{B_i}_{B_0}\mathbf{C}\left({}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{v}_0\Delta t_i - {}^{B_0}\mathbf{g}\frac{\Delta t_i^2}{2} + \mathbf{S}(t_i)\mathbf{b}_a - \mathbf{s}(t_i)\right) + {}^{C}\mathbf{p}_B \right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{81} $$

$$ \Rightarrow \underbrace{\begin{bmatrix} 1 & 0 & -u_{ij} \\ 0 & 1 & -v_{ij} \end{bmatrix} {}^{C}_{B}\mathbf{C}\ {}^{B_i}_{B_0}\mathbf{C}}_{\mathbf{J}_{ij}} \left({}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{v}_0\Delta t_i - {}^{B_0}\mathbf{g}\frac{\Delta t_i^2}{2} + \mathbf{S}(t_i)\mathbf{b}_a - \mathbf{s}(t_i) + {}^{B_0}_{B_i}\mathbf{C}\ {}^{B}_{C}\mathbf{C}\ {}^{C}\mathbf{p}_B\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{82} $$

$$ \Rightarrow \mathbf{J}_{ij}\left({}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{v}_0\Delta t_i - {}^{B_0}\mathbf{g}\frac{\Delta t_i^2}{2} + \mathbf{S}(t_i)\mathbf{b}_a - \mathbf{s}(t_i) + {}^{B_0}_{B_i}\mathbf{C}\ {}^{B}_{C}\mathbf{C}\ {}^{C}\mathbf{p}_B\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{83} $$

where

$$ \mathbf{J}_{ij} = \begin{bmatrix} 1 & 0 & -u_{ij} \\ 0 & 1 & -v_{ij} \end{bmatrix} {}^{C}_{B}\mathbf{C}\ {}^{B_i}_{B_0}\mathbf{C} \tag{84} $$

On the left-hand side of Eq. (83), the vectors ${}^{B_0}\mathbf{p}_j$, ${}^{B_0}\mathbf{v}_0$, $\mathbf{b}_a$, ${}^{C}\mathbf{p}_B$, and ${}^{B_0}\mathbf{g}$ appear linearly. By keeping the unknown quantities on the left-hand side, and moving the known quantities to the right-hand side, we can form a linear equation $\mathbf{A}_{ij}\mathbf{x} = \mathbf{b}_{ij}$ from the observation of the j-th feature in the i-th image. By collecting the equations resulting from all feature measurements, we obtain the linear system:

$$ \mathbf{A}\mathbf{x} = \mathbf{b} \tag{85} $$

We next examine the structure of the matrix $\mathbf{A}$ and the vector of unknown quantities $\mathbf{x}$, for the different cases studied in Sections 3 and 4.

A.1 Known biases and extrinsic IMU-Camera calibration

This scenario is discussed in Section 3. In this case, the IMU-camera translation, i.e., ${}^{C}\mathbf{p}_B$ in Eq. (83), is known and constant, while the accelerometer bias is known and removed from the accelerometer measurements, i.e., $\mathbf{b}_a$ in Eq. (83) becomes zero. Therefore, the unknown quantities in Eq. (83) are the vectors ${}^{B_0}\mathbf{p}_j$, ${}^{B_0}\mathbf{v}_0$, and ${}^{B_0}\mathbf{g}$. Then, Eq. (83) can be re-formulated as:

$$ \mathbf{J}_{ij}\left({}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{v}_0\Delta t_i - {}^{B_0}\mathbf{g}\frac{\Delta t_i^2}{2} + \underbrace{\mathbf{S}(t_i)\mathbf{b}_a}_{\mathbf{0}} - \mathbf{s}(t_i) + {}^{B_0}_{B_i}\mathbf{C}\ {}^{B}_{C}\mathbf{C}\ {}^{C}\mathbf{p}_B\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{86} $$

$$ \Rightarrow \mathbf{J}_{ij}\left({}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{v}_0\Delta t_i - {}^{B_0}\mathbf{g}\frac{\Delta t_i^2}{2}\right) = \mathbf{J}_{ij}\left(\mathbf{s}(t_i) - {}^{B_0}_{B_i}\mathbf{C}\ {}^{B}_{C}\mathbf{C}\ {}^{C}\mathbf{p}_B\right) \tag{87} $$

$$ \Rightarrow \underbrace{\mathbf{J}_{ij}\begin{bmatrix} \mathbf{0}_{3\times 3} & \cdots & \underbrace{\mathbf{I}_3}_{j-\text{th block}} & \cdots \mathbf{0}_{3\times 3} & -\Delta t_i\mathbf{I}_3 & -\frac{\Delta t_i^2}{2}\mathbf{I}_3 \end{bmatrix}}_{\mathbf{A}_{ij}} \underbrace{\begin{bmatrix} {}^{B_0}\mathbf{p}_1 \\ \vdots \\ {}^{B_0}\mathbf{p}_j \\ \vdots \\ {}^{B_0}\mathbf{p}_M \\ {}^{B_0}\mathbf{v}_0 \\ {}^{B_0}\mathbf{g} \end{bmatrix}}_{\mathbf{x}} = \underbrace{\mathbf{J}_{ij}\left(\mathbf{s}(t_i) - {}^{B_0}_{B_i}\mathbf{C}\ {}^{B}_{C}\mathbf{C}\ {}^{C}\mathbf{p}_B\right)}_{\mathbf{b}_{ij}} \tag{88} $$

$$\Rightarrow \qquad \mathbf{A}_{ij}\mathbf{x} = \mathbf{b}_{ij} \tag{89}$$

where $\mathbf{x}$ is a $(3M+6)\times 1$ vector and:

$$\mathbf{A}_{ij} = \mathbf{J}_{ij}\begin{bmatrix}\mathbf{0}_{3\times 3} & \dots & \underbrace{\mathbf{I}_3}_{j-\text{th block}} & \dots \mathbf{0}_{3\times 3} & -\Delta t_i \mathbf{I}_3 & -\frac{\Delta t_i^2}{2}\mathbf{I}_3\end{bmatrix} \tag{90}$$

$$\mathbf{b}_{ij} = \mathbf{J}_{ij}\left(\mathbf{s}(t_i) - {}^{B_0}_{B_i}\mathbf{C}\ {}^{B}_{C}\mathbf{C}\ {}^{C}\mathbf{p}_B\right) = \mathbf{J}_{ij}\ \mathbf{s}(t_i) - \begin{bmatrix}1 & 0 & -u_{ij}\\ 0 & 1 & -v_{ij}\end{bmatrix}{}^{C}\mathbf{p}_B \tag{91}$$

From this derivation, we obtain Eq. (12)-(13) in Section 3.

A.2 Unknown extrinsic calibration, known IMU biases

In the case examined in Section 4.1, the IMU-camera translation, ${}^{C}\mathbf{p}_B$, is unknown, while the accelerometer bias is known and removed from the accelerometer measurements (i.e., $\mathbf{b}_a = \mathbf{0}$). Therefore, we have the extra unknown ${}^{C}\mathbf{p}_B$, in addition to the unknown quantities ${}^{B_0}\mathbf{p}_j$, ${}^{B_0}\mathbf{v}_0$, and ${}^{B_0}\mathbf{g}$. Similarly, Eq. (83) can be re-formulated as:

$$\mathbf{J}_{ij}\left({}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{v}_0\Delta t_i - {}^{B_0}\mathbf{g}\frac{\Delta t_i^2}{2} + \underbrace{\mathbf{S}(t_i)\mathbf{b}_a}_{\mathbf{0}} - \mathbf{s}(t_i) + {}^{B_0}_{B_i}\mathbf{C}\ {}^{B}_{C}\mathbf{C}\ {}^{C}\mathbf{p}_B\right) = \begin{bmatrix}0\\0\end{bmatrix} \tag{92}$$

$$\Rightarrow \mathbf{J}_{ij}\left({}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{v}_0\Delta t_i + {}^{B_0}_{B_i}\mathbf{C}\ {}^{B}_{C}\mathbf{C}\ {}^{C}\mathbf{p}_B - {}^{B_0}\mathbf{g}\frac{\Delta t_i^2}{2}\right) = \mathbf{J}_{ij}\mathbf{s}(t_i) \tag{93}$$

$$\Rightarrow \underbrace{\mathbf{J}_{ij}\begin{bmatrix}\mathbf{0}_{3\times 3} & \dots & \underbrace{\mathbf{I}_3}_{j-\text{th block}} & \dots & \mathbf{0}_{3\times 3} & -\Delta t_i\mathbf{I}_3 & {}^{B_0}_{B_i}\mathbf{C}\ {}^{B}_{C}\mathbf{C} & -\frac{\Delta t_i^2}{2}\mathbf{I}_3\end{bmatrix}}_{\mathbf{A}'_{ij}} \underbrace{\begin{bmatrix}{}^{B_0}\mathbf{p}_1\\ \vdots \\ {}^{B_0}\mathbf{p}_j \\ \vdots \\ {}^{B_0}\mathbf{p}_M \\ {}^{B_0}\mathbf{v}_0 \\ {}^{C}\mathbf{p}_B \\ {}^{B_0}\mathbf{g}\end{bmatrix}}_{\mathbf{x}} = \underbrace{\mathbf{J}_{ij}\mathbf{s}(t_i)}_{\mathbf{b}'_{ij}} \tag{94}$$

$$\Rightarrow \qquad \mathbf{A}'_{ij}\mathbf{x} = \mathbf{b}'_{ij} \tag{95}$$

where $\mathbf{x}$ is a $(3M+9)\times 1$ vector and:

$$\mathbf{A}'_{ij} = \mathbf{J}_{ij}\begin{bmatrix}\mathbf{0}_{3\times 3} & \dots & \underbrace{\mathbf{I}_3}_{j-\text{th block}} & \dots & \mathbf{0}_{3\times 3} & -\Delta t_i\mathbf{I}_3 & {}^{B_0}_{B_i}\mathbf{C}\ {}^{B}_{C}\mathbf{C} & -\frac{\Delta t_i^2}{2}\mathbf{I}_3\end{bmatrix} \tag{96}$$

$$\mathbf{b}'_{ij} = \mathbf{J}_{ij}\ \mathbf{s}(t_i) \tag{97}$$

A.3 Unknown extrinsic calibration, unknown IMU biases

In Section 4.2, both the IMU-camera translation, ${}^{C}\mathbf{p}_B$, and the accelerometer bias, $\mathbf{b}_a$, are unknown. Therefore, we have the extra unknowns ${}^{C}\mathbf{p}_B, \mathbf{b}_a$, in addition to the unknown quantities ${}^{B_0}\mathbf{p}_j$, ${}^{B_0}\mathbf{v}_0$, and ${}^{B_0}\mathbf{g}$. Similarly to Section A.1, Eq. (83) can be re-formulated as:

$$\mathbf{J}_{ij}\left({}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{v}_0\Delta t_i - {}^{B_0}\mathbf{g}\frac{\Delta t_i^2}{2} + \mathbf{S}(t_i)\mathbf{b}_a - \mathbf{s}(t_i) + {}^{B_0}_{B_i}\mathbf{C}\ {}^{B}_{C}\mathbf{C}\ {}^{C}\mathbf{p}_B\right) = \begin{bmatrix}0\\0\end{bmatrix} \tag{98}$$

$$\Rightarrow \mathbf{J}_{ij}\left({}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{v}_0\Delta t_i + \mathbf{S}(t_i)\mathbf{b}_a + {}^{B_0}_{B_i}\mathbf{C}\ {}^{B}_{C}\mathbf{C}\ {}^{C}\mathbf{p}_B - {}^{B_0}\mathbf{g}\frac{\Delta t_i^2}{2}\right) = \mathbf{J}_{ij}\mathbf{s}(t_i) \tag{99}$$

$$\Rightarrow \underbrace{\mathbf{J}_{ij}\begin{bmatrix}\mathbf{0}_{3\times3} & \cdots & \underbrace{\mathbf{I}_3}_{j-\text{th block}} & \cdots & \mathbf{0}_{3\times3} & -\Delta t_i\mathbf{I}_3 & \mathbf{S}(t_i) & {}^{B_0}_{B_i}\mathbf{C}\;{}^{B}_{C}\mathbf{C} & -\frac{\Delta t_i^2}{2}\mathbf{I}_3\end{bmatrix}}_{\mathbf{A}''_{ij}} \underbrace{\begin{bmatrix} {}^{B_0}\mathbf{p}_1 \\ \vdots \\ {}^{B_0}\mathbf{p}_j \\ \vdots \\ {}^{B_0}\mathbf{p}_M \\ {}^{B_0}\mathbf{v}_0 \\ \mathbf{b}_a \\ {}^{C}\mathbf{p}_B \\ {}^{B_0}\mathbf{g}\end{bmatrix}}_{\mathbf{x}} = \underbrace{\mathbf{J}_{ij}\mathbf{s}(t_i)}_{\mathbf{b}''_{ij}} \tag{100}$$

$$\Rightarrow \qquad \mathbf{A}''_{ij}\mathbf{x} = \mathbf{b}''_{ij} \tag{101}$$

where $\mathbf{x}$ is a $(3M + 12) \times 1$ vector and:

$$\mathbf{A}''_{ij} = \mathbf{J}_{ij}\begin{bmatrix}\mathbf{0}_{3\times3} & \cdots & \underbrace{\mathbf{I}_3}_{j-\text{th block}} & \cdots & \mathbf{0}_{3\times3} & -\Delta t_i\mathbf{I}_3 & \mathbf{S}(t_i) & {}^{B_0}_{B_i}\mathbf{C}\;{}^{B}_{C}\mathbf{C} & -\frac{\Delta t_i^2}{2}\mathbf{I}_3\end{bmatrix} \tag{102}$$

$$\mathbf{b}''_{ij} = \mathbf{J}_{ij}\;\mathbf{s}(t_i) \tag{103}$$

B Analysis of the rank of A and minimal cases

B.1 Singular cases

We here show that if the camera moves with a constant acceleration, that is, if

$${}^{B_0}\mathbf{p}_{C_i} = {}^{B_0}\mathbf{p}_{C_0} + \Delta t_i\mathbf{v} + \frac{\Delta t_i^2}{2}\mathbf{a}\,, \tag{104}$$

then $\mathbf{A}$ has a nullspace of dimension at least one. We do so by proving that the following vector is one basis vector for this nullspace, regardless of the number of feature obsevations:

$$\mathbf{n} = \left[\left({}^{B_0}\mathbf{p}_1 - {}^{B_0}\mathbf{p}_{C_0}\right)^T \;\; \ldots \;\; \left({}^{B_0}\mathbf{p}_M - {}^{B_0}\mathbf{p}_{C_0}\right)^T \;\; \mathbf{v}^T \;\; \mathbf{a}^T\right]^T \tag{105}$$

For the observation of the j-th feature in the i-th image, we obtain two rows of the matrix $\mathbf{A}$, defined by Eq. (12)-(14) for the case of known biases and extrinsic calibration. To show that $\mathbf{n}$ is in the nullspace of $\mathbf{A}$, we first show that $\mathbf{A}_{ij}\mathbf{n} = \mathbf{0}$:

$$\mathbf{A}_{ij}\mathbf{n} = \mathbf{J}_{ij}\begin{bmatrix}\mathbf{0}_{3\times3} & \cdots & \underbrace{\mathbf{I}_3}_{j-\text{th block}} & \cdots & \mathbf{0}_{3\times3} & -\Delta t_i\mathbf{I}_3 & -\frac{\Delta t_i^2}{2}\mathbf{I}_3\end{bmatrix}\begin{bmatrix}{}^{B_0}\mathbf{p}_1 - {}^{B_0}\mathbf{p}_{C_0} \\ \vdots \\ {}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{p}_{C_0} \\ \vdots \\ {}^{B_0}\mathbf{p}_M - {}^{B_0}\mathbf{p}_{C_0} \\ \mathbf{v} \\ \mathbf{a}\end{bmatrix} \tag{106}$$

$$= \mathbf{J}_{ij}\left({}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{p}_{C_0} - \Delta t_i\mathbf{v} - \frac{\Delta t_i^2}{2}\mathbf{a}\right) \tag{107}$$

$$= \mathbf{J}_{ij}\left({}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{p}_{C_i}\right) \qquad \text{from Eq. (104)} \tag{108}$$

$$= \begin{bmatrix}1 & 0 & -u_{ij} \\ 0 & 1 & -v_{ij}\end{bmatrix}\underbrace{{}^{C}_{B}\mathbf{C}\;{}^{B_i}_{B_0}\mathbf{C}}_{{}^{C_i}_{B_0}\mathbf{C}}({}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{p}_{C_i}) \tag{109}$$

$$= \begin{bmatrix}1 & 0 & -u_{ij} \\ 0 & 1 & -v_{ij}\end{bmatrix}{}^{C_i}\mathbf{p}_j = \mathbf{0} \qquad \text{from Eq. (10)} \tag{110}$$

Since the above result holds for all $\{i, j\} \in \mathcal{S}_m$, we have:

$$\mathbf{A}\,\mathbf{n} = \mathbf{0} \tag{111}$$

Thus, the vector $\mathbf{n}$ defined in Eq. (105) is a basis vector for the nullspace of the matrix $\mathbf{A}$. If enough feature measurements are available, this will be the only basis vector for this nullspace, and thus $\dim(\mathcal{N}(\mathbf{A})) = 1$.

B.2 Minimal cases

B.2.1 Finding minimal cases

To simplify the analysis of minimal problems, we consider the cases where each of the available features is observed in all the images. In those cases, the total number of feature measurements is $K = \sum_{j=1}^{M} n_j = NM$, and thus, from Result 2, the necessary condition for the existence of a finite number of solutions is $2NM \geq 3M + 5$. We can now identify the following cases for the number of available images:

- $N = 1$: The necessary condition is not met, regardless of the number of features.
- $N = 2$: The necessary condition is met when at least $M = 5$ features are available. However, we will show that if only two images are available, the dimension the nullspace of $\mathbf{A}$ is at least 3, regardless of the number of features. Thus for $N = 2$ an infinite number of solutions always exists.
- $N = 3$: The necessary condition is met when $M \geq 2$.
- $N = 4$: The necessary condition is met when $M \geq 1$.
- $N > 4$: The necessary condition is met when $M \geq 1$.

From the above we see that two *minimal problems* can be identified: (i) two features seen in three images, and (ii) one feature seen in four images.

B.2.2 Two images

We here show that when only $N = 2$ images are available, the matrix $\mathbf{A}$ has a nullspace of dimension at least 3. For this analysis, we will employ the decomposition of $\mathbf{A}$ into a block diagonal matrix $\mathbf{D}$, and a matrix $\mathbf{K}$, such that $\mathbf{A} = \mathbf{DK}$, where:

$$\mathbf{D} = \begin{bmatrix} \ddots & & \mathbf{0} \\ & \mathbf{J}_{ij} & \\ \mathbf{0} & & \ddots \end{bmatrix}, \quad \mathbf{K} = \left[\begin{array}{ccccccc} \mathbf{I}_3 & \cdots & \mathbf{0}_{3\times3} & \cdots & \mathbf{0}_{3\times3} & -\Delta t_0 \mathbf{I}_3 & -\frac{1}{2}\Delta t_0^2 \mathbf{I}_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_{3\times3} & \cdots & \mathbf{I}_3 & \cdots & \mathbf{0}_{3\times3} & -\Delta t_0 \mathbf{I}_3 & -\frac{1}{2}\Delta t_0^2 \mathbf{I}_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_{3\times3} & \cdots & \mathbf{0}_{3\times3} & \cdots & \mathbf{I}_3 & -\Delta t_0 \mathbf{I}_3 & -\frac{1}{2}\Delta t_0^2 \mathbf{I}_3 \\ \hline \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \hline \mathbf{I}_3 & \cdots & \mathbf{0}_{3\times3} & \cdots & \mathbf{0}_{3\times3} & -\Delta t_{N-1} \mathbf{I}_3 & -\frac{1}{2}\Delta t_{N-1}^2 \mathbf{I}_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_{3\times3} & \cdots & \mathbf{I}_3 & \cdots & \mathbf{0}_{3\times3} & -\Delta t_{N-1} \mathbf{I}_3 & -\frac{1}{2}\Delta t_{N-1}^2 \mathbf{I}_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0}_{3\times3} & \cdots & \mathbf{0}_{3\times3} & \cdots & \mathbf{I}_3 & -\Delta t_{N-1} \mathbf{I}_3 & -\frac{1}{2}\Delta t_{N-1}^2 \mathbf{I}_3 \end{array}\right] \tag{112}$$

when we have M features and N images.

For $N = 2$ images, by performing Gaussian elimination on $\mathbf{K}$, we have (in the following, we use the sign $\sim$ to denote a transformation using row or column operations on the matrix):

$$\mathbf{K} = \left[\begin{array}{ccccccc}
\mathbf{I}_3 & \cdots & \mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & -\Delta t_0 \mathbf{I}_3 & -\frac{1}{2}\Delta t_0^2 \mathbf{I}_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0}_{3\times 3} & \cdots & \mathbf{I}_3 & \cdots & \mathbf{0}_{3\times 3} & -\Delta t_0 \mathbf{I}_3 & -\frac{1}{2}\Delta t_0^2 \mathbf{I}_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & \cdots & \mathbf{I}_3 & -\Delta t_0 \mathbf{I}_3 & -\frac{1}{2}\Delta t_0^2 \mathbf{I}_3 \\
\hline
\mathbf{I}_3 & \cdots & \mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & -\Delta t_1 \mathbf{I}_3 & -\frac{1}{2}\Delta t_1^2 \mathbf{I}_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0}_{3\times 3} & \cdots & \mathbf{I}_3 & \cdots & \mathbf{0}_{3\times 3} & -\Delta t_1 \mathbf{I}_3 & -\frac{1}{2}\Delta t_1^2 \mathbf{I}_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & \cdots & \mathbf{I}_3 & -\Delta t_1 \mathbf{I}_3 & -\frac{1}{2}\Delta t_1^2 \mathbf{I}_3
\end{array}\right] \tag{113}$$

Subtract each block row in the upper partition from the corresponding block row in the lower partition

$$\sim \left[\begin{array}{ccccccc}
\mathbf{I}_3 & \cdots & \mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & -\Delta t_0 \mathbf{I}_3 & -\frac{1}{2}\Delta t_0^2 \mathbf{I}_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0}_{3\times 3} & \cdots & \mathbf{I}_3 & \cdots & \mathbf{0}_{3\times 3} & -\Delta t_0 \mathbf{I}_3 & -\frac{1}{2}\Delta t_0^2 \mathbf{I}_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & \cdots & \mathbf{I}_3 & -\Delta t_0 \mathbf{I}_3 & -\frac{1}{2}\Delta t_0^2 \mathbf{I}_3 \\
\hline
\mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & -(\Delta t_1 - \Delta t_0) \mathbf{I}_3 & -(\frac{1}{2}\Delta t_1^2 - \frac{1}{2}\Delta t_0^2) \mathbf{I}_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & -(\Delta t_1 - \Delta t_0) \mathbf{I}_3 & -(\frac{1}{2}\Delta t_1^2 - \frac{1}{2}\Delta t_0^2) \mathbf{I}_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & -(\Delta t_1 - \Delta t_0) \mathbf{I}_3 & -(\frac{1}{2}\Delta t_1^2 - \frac{1}{2}\Delta t_0^2) \mathbf{I}_3
\end{array}\right] \tag{114}$$

Subtract the first block row in the lower partition from other block rows in the lower partition

$$\sim \left[\begin{array}{ccccccc}
\mathbf{I}_3 & \cdots & \mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & -\Delta t_0 \mathbf{I}_3 & -\frac{1}{2}\Delta t_0^2 \mathbf{I}_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0}_{3\times 3} & \cdots & \mathbf{I}_3 & \cdots & \mathbf{0}_{3\times 3} & -\Delta t_0 \mathbf{I}_3 & -\frac{1}{2}\Delta t_0^2 \mathbf{I}_3 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & \cdots & \mathbf{I}_3 & -\Delta t_0 \mathbf{I}_3 & -\frac{1}{2}\Delta t_0^2 \mathbf{I}_3 \\
\hline
\mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & -(\Delta t_1 - \Delta t_0) \mathbf{I}_3 & -(\frac{1}{2}\Delta t_1^2 - \frac{1}{2}\Delta t_0^2) \mathbf{I}_3 \\
\mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
\mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & \cdots & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3}
\end{array}\right] \tag{115}$$

From the Gauss-reduced form (row echelon form) above, we conclude that:

$$\mathrm{rank}(\mathbf{K}) = 3M + 3 \tag{116}$$

for any number of features M in $N = 2$ images. We will use the following results [15, 4.5.1]:

$$\begin{aligned}
\mathrm{rank}(\mathbf{A}) &= \mathrm{rank}(\mathbf{DK}) \\
&= \mathrm{rank}(\mathbf{K}) - \dim(\mathcal{N}(\mathbf{D}) \cap \mathcal{R}(\mathbf{K}))
\end{aligned}$$

from which we can write:

$$\mathrm{rank}(\mathbf{A}) = \mathrm{rank}(\mathbf{K}) - \dim(\mathcal{N}(\mathbf{D}) \cap \mathcal{R}(\mathbf{K})) \leq \mathrm{rank}(\mathbf{K}) = 3M + 3 \tag{117}$$

Applying the rank-nullity theorem, we have:

$$\operatorname{rank}(\mathbf{A}) + \dim(\mathcal{N}(\mathbf{A})) = 3M + 6 \Rightarrow \dim(\mathcal{N}(\mathbf{A})) \geq 3 \tag{118}$$

Therefore, we conclude that if only two images are available ($N = 2$), the dimension the nullspace of $\mathbf{A}$ is at least 3, regardless of the number of features. In the following subsections, we will analyze the rank of the matrix $\mathbf{A}$ in the two minimal cases.

B.2.3 Three images, two features

In this section, we consider the first minimal case where two features are observed and tracked in three images. We find the nullspace of $\mathbf{A}$, by finding all the solution vectors $\mathbf{a}$ for the following equation:

$$\mathbf{A}\mathbf{a} = \mathbf{0} \tag{119}$$

where $\mathbf{a}$ is a $3M + 6 \times 1$ vector, and it can be decomposed as follows:

$$\mathbf{a} = [\mathbf{a}_1^T \ \ \dots \ \ \mathbf{a}_M^T \ \ \mathbf{a}_v^T \ \ \mathbf{a}_g^T]^T, \ \ \mathbf{a}_1, \dots, \mathbf{a}_M, \mathbf{a}_v, \mathbf{a}_g \in \mathbb{R}^3 \tag{120}$$

Using Eq. (12), we obtain:

$$\begin{aligned}
& \mathbf{A}\mathbf{a} = \mathbf{0} && (121)\\
\Rightarrow\ & \mathbf{A}_{ij}\mathbf{a} = \mathbf{0}, \ \ \forall \{i, j\} \in \mathcal{S}_m && (122)\\
\Rightarrow\ & \mathbf{J}_{ij}\left(\mathbf{a}_j \Delta t_i \mathbf{a}_v - \frac{1}{2}\Delta t_i^2 \mathbf{a}_g\right) = \mathbf{0} && (123)
\end{aligned}$$

From the definition of $\mathbf{J}_{ij}$ in (14), we can conclude that the nullspace of $\mathbf{J}_{ij}$ is spanned by the vector $({}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{p}_{C_i})$. Therefore, (123) yields:

$$\mathbf{a}_j \Delta t_i \mathbf{a}_v - \frac{1}{2}\Delta t_i^2 \mathbf{a}_g = k_{ij}({}^{B_0}\mathbf{p}_j - {}^{B_0}\mathbf{p}_{C_i}) \tag{124}$$

For two features $j = 1, 2$ and three images $i = 0, 1, 2$, we have the following system of equations:

$$\begin{cases}
k_{01}({}^{B_0}\mathbf{p}_1 - {}^{B_0}\mathbf{p}_{C_0}) & = \mathbf{a}_1 - \Delta t_0 \mathbf{a}_v - \frac{1}{2}\Delta t_0^2 \mathbf{a}_g \\
k_{02}({}^{B_0}\mathbf{p}_2 - {}^{B_0}\mathbf{p}_{C_0}) & = \mathbf{a}_2 - \Delta t_0 \mathbf{a}_v - \frac{1}{2}\Delta t_0^2 \mathbf{a}_g \\
k_{11}({}^{B_0}\mathbf{p}_1 - {}^{B_0}\mathbf{p}_{C_1}) & = \mathbf{a}_1 - \Delta t_1 \mathbf{a}_v - \frac{1}{2}\Delta t_1^2 \mathbf{a}_g \\
k_{12}({}^{B_0}\mathbf{p}_2 - {}^{B_0}\mathbf{p}_{C_1}) & = \mathbf{a}_2 - \Delta t_1 \mathbf{a}_v - \frac{1}{2}\Delta t_1^2 \mathbf{a}_g \\
k_{21}({}^{B_0}\mathbf{p}_1 - {}^{B_0}\mathbf{p}_{C_2}) & = \mathbf{a}_1 - \Delta t_2 \mathbf{a}_v - \frac{1}{2}\Delta t_2^2 \mathbf{a}_g \\
k_{22}({}^{B_0}\mathbf{p}_2 - {}^{B_0}\mathbf{p}_{C_2}) & = \mathbf{a}_2 - \Delta t_2 \mathbf{a}_v - \frac{1}{2}\Delta t_2^2 \mathbf{a}_g
\end{cases} \tag{125}$$

With respect to the unknowns $\{k_{01}, k_{02}, k_{11}, k_{12}, k_{21}, k_{22}, \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_v, \mathbf{a}_g\}$, the above equations can be written in linear-system form:

$$\underbrace{\begin{bmatrix}
\mathbf{p}_{01} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{I}_3 & \mathbf{0}_{3\times1} & \Delta t_0 \mathbf{I}_3 & \Delta {t_0}^2 \mathbf{I}_3 \\
\mathbf{0}_{3\times1} & \mathbf{p}_{02} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{I}_3 & \Delta t_0 \mathbf{I}_3 & \Delta {t_0}^2 \mathbf{I}_3 \\
\mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{11} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{I}_3 & \mathbf{0}_{3\times1} & \Delta t_1 \mathbf{I}_3 & \Delta {t_1}^2 \mathbf{I}_3 \\
\mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{12} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{I}_3 & \Delta t_1 \mathbf{I}_3 & \Delta {t_1}^2 \mathbf{I}_3 \\
\mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{21} & \mathbf{0}_{3\times1} & \mathbf{I}_3 & \mathbf{0}_{3\times1} & \Delta t_2 \mathbf{I}_3 & \Delta {t_2}^2 \mathbf{I}_3 \\
\mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{22} & \mathbf{0}_{3\times1} & \mathbf{I}_3 & \Delta t_2 \mathbf{I}_3 & \Delta {t_2}^2 \mathbf{I}_3
\end{bmatrix}}_{\mathbf{\Gamma}}
\underbrace{\begin{bmatrix} k_{01} \\ k_{02} \\ k_{11} \\ k_{12} \\ k_{21} \\ k_{22} \\ \mathbf{a}_1 \\ \mathbf{a}_2 \\ -\mathbf{a}_v \\ -\frac{1}{2}\mathbf{a}_g \end{bmatrix}}_{\mathbf{y}} = \mathbf{0} \tag{126}$$

where the vector $\mathbf{y}$ contains all the unknowns in the equations, and we define:

$$\mathbf{p}_{ij} \ \ = {}^{B_0}\mathbf{p}_{C_i} - {}^{B_0}\mathbf{p}_j, \ \ i \in \{0, 1, 2\}, j \in \{1, 2\} \tag{127}$$

We will identify the conditions under which the homogeneous equation $\mathbf{\Gamma y} = \mathbf{0}$ has a non-zero solution by analyzing the rank of the 18×18 matrix $\mathbf{\Gamma}$. We perform the following elementary row and column operations on $\mathbf{\Gamma}$:

$$\mathbf{\Gamma} = \begin{bmatrix} \mathbf{p}_{01} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{I}_3 & \mathbf{0}_{3\times3} & \Delta t_0 \mathbf{I}_3 & \Delta {t_0}^2 \mathbf{I}_3 \\ \mathbf{0}_{3\times1} & \mathbf{p}_{02} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_0 \mathbf{I}_3 & \Delta {t_0}^2 \mathbf{I}_3 \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{11} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{I}_3 & \mathbf{0}_{3\times3} & \Delta t_1 \mathbf{I}_3 & \Delta {t_1}^2 \mathbf{I}_3 \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{12} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_1 \mathbf{I}_3 & \Delta {t_1}^2 \mathbf{I}_3 \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{21} & \mathbf{0}_{3\times1} & \mathbf{I}_3 & \mathbf{0}_{3\times3} & \Delta t_2 \mathbf{I}_3 & \Delta {t_2}^2 \mathbf{I}_3 \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{22} & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_2 \mathbf{I}_3 & \Delta {t_2}^2 \mathbf{I}_3 \end{bmatrix} \tag{128}$$

Add the block column for $\mathbf{a}_1$ to the block column for $\mathbf{a}_2$

$$\sim \begin{bmatrix} \mathbf{p}_{01} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{I}_3 & \mathbf{I}_3 & \Delta t_0 \mathbf{I}_3 & \Delta {t_0}^2 \mathbf{I}_3 \\ \mathbf{0}_{3\times1} & \mathbf{p}_{02} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_0 \mathbf{I}_3 & \Delta {t_0}^2 \mathbf{I}_3 \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{11} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{I}_3 & \mathbf{I}_3 & \Delta t_1 \mathbf{I}_3 & \Delta {t_1}^2 \mathbf{I}_3 \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{12} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_1 \mathbf{I}_3 & \Delta {t_1}^2 \mathbf{I}_3 \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{21} & \mathbf{0}_{3\times1} & \mathbf{I}_3 & \mathbf{I}_3 & \Delta t_2 \mathbf{I}_3 & \Delta {t_2}^2 \mathbf{I}_3 \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{22} & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_2 \mathbf{I}_3 & \Delta {t_2}^2 \mathbf{I}_3 \end{bmatrix} \tag{129}$$

Subtract the second block row from the first block row

$$\sim \begin{bmatrix} \mathbf{p}_{01} & -\mathbf{p}_{02} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times1} & \mathbf{p}_{02} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_0 \mathbf{I}_3 & \Delta {t_0}^2 \mathbf{I}_3 \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{11} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{I}_3 & \mathbf{I}_3 & \Delta t_1 \mathbf{I}_3 & \Delta {t_1}^2 \mathbf{I}_3 \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{12} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_1 \mathbf{I}_3 & \Delta {t_1}^2 \mathbf{I}_3 \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{21} & \mathbf{0}_{3\times1} & \mathbf{I}_3 & \mathbf{I}_3 & \Delta t_2 \mathbf{I}_3 & \Delta {t_2}^2 \mathbf{I}_3 \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{22} & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_2 \mathbf{I}_3 & \Delta {t_2}^2 \mathbf{I}_3 \end{bmatrix} \tag{130}$$

We now perform Gauss elimination for the first two block rows. We multiply the first block row with the matrix:

$$\mathbf{T}_{12} = \begin{bmatrix} \mathbf{p}_{01}^T \\ \mathbf{p}_{02}^T \\ (\mathbf{p}_{01} \times \mathbf{p}_{02})^T \end{bmatrix}$$

and perform row operations as follows:

$$\mathbf{\Gamma}_{1:2} = \begin{bmatrix} \mathbf{p}_{01} & -\mathbf{p}_{02} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times1} & \mathbf{p}_{02} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_0 \mathbf{I}_3 & \Delta {t_0}^2 \mathbf{I}_3 \end{bmatrix}$$

Multiply the first block row by $\mathbf{T}_{12}$

$$\sim \begin{bmatrix} \mathbf{p}_{01}^T\mathbf{p}_{01} & -\mathbf{p}_{01}^T\mathbf{p}_{02} & 0 & 0 & 0 & 0 & \mathbf{p}_{01}^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ \mathbf{p}_{02}^T\mathbf{p}_{01} & -\mathbf{p}_{02}^T\mathbf{p}_{02} & 0 & 0 & 0 & 0 & \mathbf{p}_{02}^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ 0 & 0 & 0 & 0 & 0 & 0 & (\mathbf{p}_{01} \times \mathbf{p}_{02})^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ \mathbf{0}_{3\times1} & \mathbf{p}_{02} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_0 \mathbf{I}_3 & \Delta {t_0}^2 \mathbf{I}_3 \end{bmatrix}$$

Add the first row multiplied by $-\dfrac{\mathbf{p}_{02}^T\mathbf{p}_{01}}{\mathbf{p}_{01}^T\mathbf{p}_{01}}$ to the second row, and define $s_1 = \mathbf{p}_{02}^T\mathbf{p}_{02} - \dfrac{(\mathbf{p}_{02}^T\mathbf{p}_{01})^2}{\mathbf{p}_{01}^T\mathbf{p}_{01}}$

$$\sim \begin{bmatrix} \mathbf{p}_{01}^T\mathbf{p}_{01} & -\mathbf{p}_{01}^T\mathbf{p}_{02} & 0 & 0 & 0 & 0 & \mathbf{p}_{01}^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ 0 & -s_1 & 0 & 0 & 0 & 0 & \mathbf{p}_{02}^T - \mathbf{p}_{01}^T \frac{\mathbf{p}_{02}^T\mathbf{p}_{01}}{\mathbf{p}_{01}^T\mathbf{p}_{01}} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ 0 & 0 & 0 & 0 & 0 & 0 & (\mathbf{p}_{01} \times \mathbf{p}_{02})^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ \mathbf{0}_{3\times1} & \mathbf{p}_{02} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_0 \mathbf{I}_3 & \Delta {t_0}^2 \mathbf{I}_3 \end{bmatrix}$$

Add the second row multiplied by $\dfrac{1}{s_1}\mathbf{p}_{02}$ to the second block row (rows 4-6)

$$\sim \begin{bmatrix} \mathbf{p}_{01}^T\mathbf{p}_{01} & -\mathbf{p}_{01}^T\mathbf{p}_{02} & 0 & 0 & 0 & 0 & \mathbf{p}_{01}^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ 0 & -s_1 & 0 & 0 & 0 & 0 & \mathbf{p}_{02}^T - \mathbf{p}_{01}^T \frac{\mathbf{p}_{02}^T\mathbf{p}_{01}}{\mathbf{p}_{01}^T\mathbf{p}_{01}} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ 0 & 0 & 0 & 0 & 0 & 0 & (\mathbf{p}_{01} \times \mathbf{p}_{02})^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \frac{1}{s_1}\mathbf{p}_{02}(\mathbf{p}_{02}^T - \mathbf{p}_{01}^T \frac{\mathbf{p}_{02}^T\mathbf{p}_{01}}{\mathbf{p}_{01}^T\mathbf{p}_{01}}) & \mathbf{I}_3 & \Delta t_0 \mathbf{I}_3 & \Delta {t_0}^2 \mathbf{I}_3 \end{bmatrix}$$

$$\sim \begin{bmatrix} \mathbf{p}_{01}^T\mathbf{p}_{01} & -\mathbf{p}_{01}^T\mathbf{p}_{02} & 0 & 0 & 0 & 0 & \mathbf{p}_{01}^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ 0 & -s_1 & 0 & 0 & 0 & 0 & \mathbf{p}_{02}^T - \mathbf{p}_{01}^T\frac{\mathbf{p}_{02}^T\mathbf{p}_{01}}{\mathbf{p}_{01}^T\mathbf{p}_{01}} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ 0 & 0 & 0 & 0 & 0 & 0 & (\mathbf{p}_{01}\times\mathbf{p}_{02})^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{F}_1 & \mathbf{I}_3 & \Delta t_0\mathbf{I}_3 & \Delta {t_0}^2\mathbf{I}_3 \end{bmatrix}$$

where

$$\mathbf{F}_1 = \frac{1}{s_1}\mathbf{p}_{02}\left(\mathbf{p}_{02}^T - \mathbf{p}_{01}^T\frac{\mathbf{p}_{02}^T\mathbf{p}_{01}}{\mathbf{p}_{01}^T\mathbf{p}_{01}}\right) = \frac{1}{s_1}\mathbf{p}_{02}\mathbf{p}_{02}^T\left(\mathbf{I}_3 - \frac{\mathbf{p}_{01}\mathbf{p}_{01}^T}{\mathbf{p}_{01}^T\mathbf{p}_{01}}\right)$$

Continuing similarly for the third-fourth and fifth-sixth block rows, after performing elimination and re-arranging the rows, we obtain:

$$\mathbf{\Gamma} \sim \left[\begin{array}{cccccc|cccc} \mathbf{p}_{01}^T\mathbf{p}_{01} & -\mathbf{p}_{01}^T\mathbf{p}_{02} & 0 & 0 & 0 & 0 & \mathbf{p}_{01}^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ 0 & -s_1 & 0 & 0 & 0 & 0 & \mathbf{p}_{02}^T - \mathbf{p}_{01}^T\frac{\mathbf{p}_{02}^T\mathbf{p}_{01}}{\mathbf{p}_{01}^T\mathbf{p}_{01}} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ 0 & 0 & \mathbf{p}_{11}^T\mathbf{p}_{11} & -\mathbf{p}_{11}^T\mathbf{p}_{12} & 0 & 0 & \mathbf{p}_{11}^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ 0 & 0 & 0 & -s_2 & 0 & 0 & \mathbf{p}_{12}^T - \mathbf{p}_{11}^T\frac{\mathbf{p}_{12}^T\mathbf{p}_{11}}{\mathbf{p}_{11}^T\mathbf{p}_{11}} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ 0 & 0 & 0 & 0 & \mathbf{p}_{21}^T\mathbf{p}_{21} & -\mathbf{p}_{21}^T\mathbf{p}_{22} & \mathbf{p}_{21}^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ 0 & 0 & 0 & 0 & 0 & -s_3 & \mathbf{p}_{22}^T - \mathbf{p}_{21}^T\frac{\mathbf{p}_{22}^T\mathbf{p}_{21}}{\mathbf{p}_{21}^T\mathbf{p}_{21}} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & (\mathbf{p}_{01}\times\mathbf{p}_{02})^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ 0 & 0 & 0 & 0 & 0 & 0 & (\mathbf{p}_{11}\times\mathbf{p}_{12})^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ 0 & 0 & 0 & 0 & 0 & 0 & (\mathbf{p}_{21}\times\mathbf{p}_{22})^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{F}_1 & \mathbf{I}_3 & \Delta t_0\mathbf{I}_3 & \Delta {t_0}^2\mathbf{I}_3 \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{F}_3 & \mathbf{I}_3 & \Delta t_1\mathbf{I}_3 & \Delta {t_1}^2\mathbf{I}_3 \\ \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{F}_2 & \mathbf{I}_3 & \Delta t_2\mathbf{I}_3 & \Delta {t_2}^2\mathbf{I}_3 \end{array}\right] \tag{131}$$

where

$$s_2 = \mathbf{p}_{12}^T\mathbf{p}_{12} - \frac{(\mathbf{p}_{12}^T\mathbf{p}_{11})^2}{\mathbf{p}_{11}^T\mathbf{p}_{11}}, \;\; s_3 = \mathbf{p}_{22}^T\mathbf{p}_{22} - \frac{(\mathbf{p}_{22}^T\mathbf{p}_{21})^2}{\mathbf{p}_{21}^T\mathbf{p}_{21}} \tag{132}$$

$$\mathbf{F}_2 = \frac{1}{s_2}\mathbf{p}_{12}\mathbf{p}_{12}^T\left(\mathbf{I}_3 - \frac{\mathbf{p}_{11}\mathbf{p}_{11}^T}{\mathbf{p}_{11}^T\mathbf{p}_{11}}\right), \;\; \mathbf{F}_3 = \frac{1}{s_3}\mathbf{p}_{22}\mathbf{p}_{22}^T\left(\mathbf{I}_3 - \frac{\mathbf{p}_{21}\mathbf{p}_{21}^T}{\mathbf{p}_{21}^T\mathbf{p}_{21}}\right) \tag{133}$$

Note that in the above expressions we have divided by the terms s_{i+1}, $i = 0, 1, 2$. To prove that these terms are nonzero, note that

$$s_{i+1} = 0 \;\Leftrightarrow\; \mathbf{p}_{i2}^T\mathbf{p}_{i2}\,\mathbf{p}_{i1}^T\mathbf{p}_{i1} - (\mathbf{p}_{i2}^T\mathbf{p}_{i1})^2 = 0 \;\Leftrightarrow\; ||\mathbf{p}_{i2}||_2^2||\mathbf{p}_{i1}||_2^2 - (\mathbf{p}_{i2}^T\mathbf{p}_{i1})^2 = 0 \tag{134}$$

By the Cauchy-Schwartz inequality, we know that the above condition is met only when $\mathbf{p}_{i1} = c\mathbf{p}_{i2}$, i.e., only when the vectors $\mathbf{p}_{i1}$ and $\mathbf{p}_{i2}$ are parallel. This is not possible however, since in that case the two features would be projected to the same point on the camera.

From (131), we conclude that:

$$\mathrm{rank}(\mathbf{\Gamma}) = 6 + \mathrm{rank}(\mathbf{B}) \tag{135}$$

with

$$\mathbf{B} = \begin{bmatrix} (\mathbf{p}_{01}\times\mathbf{p}_{02})^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ (\mathbf{p}_{11}\times\mathbf{p}_{12})^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ (\mathbf{p}_{21}\times\mathbf{p}_{22})^T & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} & \mathbf{0}_{1\times3} \\ \mathbf{F}_1 & \mathbf{I}_3 & \Delta t_0\mathbf{I}_3 & \Delta {t_0}^2\mathbf{I}_3 \\ \mathbf{F}_3 & \mathbf{I}_3 & \Delta t_1\mathbf{I}_3 & \Delta {t_1}^2\mathbf{I}_3 \\ \mathbf{F}_2 & \mathbf{I}_3 & \Delta t_2\mathbf{I}_3 & \Delta {t_2}^2\mathbf{I}_3 \end{bmatrix} \tag{136}$$

Re-arranging the columns and rows of $\mathbf{B}$, we can write:

$$\mathbf{B} \sim \begin{bmatrix} \mathbf{I}_3 & \Delta t_0 \mathbf{I}_3 & \Delta {t_0}^2 \mathbf{I}_3 & \mathbf{F}_1 \\ \mathbf{I}_3 & \Delta t_1 \mathbf{I}_3 & \Delta {t_1}^2 \mathbf{I}_3 & \mathbf{F}_2 \\ \mathbf{I}_3 & \Delta t_2 \mathbf{I}_3 & \Delta {t_2}^2 \mathbf{I}_3 & \mathbf{F}_3 \\ \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & (\mathbf{p}_{01} \times \mathbf{p}_{02})^T \\ \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & (\mathbf{p}_{11} \times \mathbf{p}_{12})^T \\ \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & (\mathbf{p}_{21} \times \mathbf{p}_{22})^T \end{bmatrix} \tag{137}$$

Subtract the first block row from the second and third block rows

$$\sim \begin{bmatrix} \mathbf{I}_3 & \Delta t_0 \mathbf{I}_3 & \Delta {t_0}^2 \mathbf{I}_3 & \mathbf{F}_1 \\ \mathbf{0}_{3\times 3} & (\Delta t_1 - \Delta t_0)\mathbf{I}_3 & ({\Delta t_1}^2 - {\Delta t_0}^2)\mathbf{I}_3 & \mathbf{F}_2 - \mathbf{F}_1 \\ \mathbf{0}_{3\times 3} & (\Delta t_2 - \Delta t_0)\mathbf{I}_3 & ({\Delta t_2}^2 - {\Delta t_0}^2)\mathbf{I}_3 & \mathbf{F}_3 - \mathbf{F}_1 \\ \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & (\mathbf{p}_{01} \times \mathbf{p}_{02})^T \\ \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & (\mathbf{p}_{11} \times \mathbf{p}_{12})^T \\ \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & (\mathbf{p}_{21} \times \mathbf{p}_{22})^T \end{bmatrix} \tag{138}$$

Divide the second and third block rows by $(\Delta t_1 - \Delta t_0)$ and $(\Delta t_2 - \Delta t_0)$, respectively

$$\sim \begin{bmatrix} \mathbf{I}_3 & \Delta t_0 \mathbf{I}_3 & \Delta {t_0}^2 \mathbf{I}_3 & \mathbf{F}_1 \\ \mathbf{0}_{3\times 3} & \mathbf{I}_3 & (\Delta t_1 + \Delta t_0)\mathbf{I}_3 & \frac{1}{\Delta t_1 - \Delta t_0}(\mathbf{F}_2 - \mathbf{F}_1) \\ \mathbf{0}_{3\times 3} & \mathbf{I}_3 & (\Delta t_2 + \Delta t_0)\mathbf{I}_3 & \frac{1}{\Delta t_2 - \Delta t_0}(\mathbf{F}_3 - \mathbf{F}_1) \\ \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & (\mathbf{p}_{01} \times \mathbf{p}_{02})^T \\ \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & (\mathbf{p}_{11} \times \mathbf{p}_{12})^T \\ \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & (\mathbf{p}_{21} \times \mathbf{p}_{22})^T \end{bmatrix} \tag{139}$$

Subtract the second block row from the third block row

$$\sim \left[\begin{array}{ccc|c} \mathbf{I}_3 & \Delta t_0 \mathbf{I}_3 & \Delta {t_0}^2 \mathbf{I}_3 & \mathbf{F}_1 \\ \mathbf{0}_{3\times 3} & \mathbf{I}_3 & (\Delta t_1 + \Delta t_0)\mathbf{I}_3 & \frac{1}{\Delta t_1 - \Delta t_0}(\mathbf{F}_2 - \mathbf{F}_1) \\ \mathbf{0}_{3\times 3} & \mathbf{0}_{3\times 3} & (\Delta t_2 - \Delta t_1)\mathbf{I}_3 & \frac{1}{\Delta t_2 - \Delta t_0}(\mathbf{F}_3 - \mathbf{F}_1) - \frac{1}{\Delta t_1 - \Delta t_0}(\mathbf{F}_2 - \mathbf{F}_1) \\ \hline \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & (\mathbf{p}_{01} \times \mathbf{p}_{02})^T \\ \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & (\mathbf{p}_{11} \times \mathbf{p}_{12})^T \\ \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & \mathbf{0}_{1\times 3} & (\mathbf{p}_{21} \times \mathbf{p}_{22})^T \end{array}\right] \tag{140}$$

The top-left submatrix above is full rank, and therefore we can write

$$\text{rank}(\mathbf{B}) = 9 + \text{rank}(\boldsymbol{\Theta}) \tag{141}$$

where

$$\boldsymbol{\Theta} = \begin{bmatrix} (\mathbf{p}_{01} \times \mathbf{p}_{02})^T \\ (\mathbf{p}_{11} \times \mathbf{p}_{12})^T \\ (\mathbf{p}_{21} \times \mathbf{p}_{22})^T \end{bmatrix} \tag{142}$$

Using Eq. (127) we see that, for $i = \{0, 1, 2\}$, it is:

$$\begin{aligned} \mathbf{p}_{i1} \times \mathbf{p}_{i2} &= ({}^{B_0}\mathbf{p}_{C_i} - {}^{B_0}\mathbf{p}_1) \times ({}^{B_0}\mathbf{p}_{C_i} - {}^{B_0}\mathbf{p}_2) \\ &= {}^{B_0}\mathbf{p}_1 \times {}^{B_0}\mathbf{p}_2 + {}^{B_0}\mathbf{p}_{C_i} \times ({}^{B_0}\mathbf{p}_1 - {}^{B_0}\mathbf{p}_2) \end{aligned}$$

and using the fact that ${}^{B_0}\mathbf{p}_1 \times {}^{B_0}\mathbf{p}_2 \perp ({}^{B_0}\mathbf{p}_1 - {}^{B_0}\mathbf{p}_2)$ and ${}^{B_0}\mathbf{p}_{C_i} \times ({}^{B_0}\mathbf{p}_1 - {}^{B_0}\mathbf{p}_2) \perp ({}^{B_0}\mathbf{p}_1 - {}^{B_0}\mathbf{p}_2)$, we obtain

$$(\mathbf{p}_{i1} \times \mathbf{p}_{i2})^T ({}^{B_0}\mathbf{p}_1 - {}^{B_0}\mathbf{p}_2) = 0 \tag{143}$$

Therefore, the vectors $(\mathbf{p}_{01} \times \mathbf{p}_{02})$, $(\mathbf{p}_{11} \times \mathbf{p}_{12})$, and $(\mathbf{p}_{21} \times \mathbf{p}_{22})$ lie in a subspace of dimension two (this subspace is the plane with normal vector $({}^{B_0}\mathbf{p}_1 - {}^{B_0}\mathbf{p}_2)$). More precisely, we have proven that the vectors lie in a subspace of dimension *at most* two. They will lie in a subspace of dimension one only when $\mathbf{p}_{i1} \times \mathbf{p}_{i2} = c_i\mathbf{d}$ for some $\mathbf{d}$, which means that all the camera positions and all the features are in the same plane. Unless this is the case, then the vectors $(\mathbf{p}_{01} \times \mathbf{p}_{02}), (\mathbf{p}_{11} \times \mathbf{p}_{12}), (\mathbf{p}_{21} \times \mathbf{p}_{22})$ span a subspace of dimension two. As a result,

$$\operatorname{rank}(\boldsymbol{\Theta}) = \operatorname{rank}\left(\begin{bmatrix} (\mathbf{p}_{01} \times \mathbf{p}_{02})^T \\ (\mathbf{p}_{11} \times \mathbf{p}_{12})^T \\ (\mathbf{p}_{21} \times \mathbf{p}_{22})^T \end{bmatrix}\right) = 2$$

Using this result together with (135) and (141), we conclude that:

$$\operatorname{rank}(\boldsymbol{\Gamma}) = 6 + 9 + 2 = 17 \tag{144}$$

From the rank-nullity theorem, this result implies that the solution space $\mathbf{y}$ of Eq. (126) has dimension of 1. This means that the solution space of the equation $\mathbf{A}\mathbf{x} = \mathbf{0}$ has dimension of 1. This leads to the conclusion that when two features are observed in three images, $\mathbf{A}$ is rank-deficient by one. We can easily verify that the nullspace of the matrix $\mathbf{A}$ has the following vector as its basis:

$$\mathbf{n} = \begin{bmatrix} {}^{B_0}\mathbf{p}_1 - {}^{B_0}\mathbf{p}_{C_0} \\ {}^{B_0}\mathbf{p}_2 - {}^{B_0}\mathbf{p}_{C_0} \\ \frac{\Delta t_2({}^{B_0}\mathbf{p}_{C_1} - {}^{B_0}\mathbf{p}_{C_0})}{\Delta t_1(\Delta t_2 - \Delta t_1)} - \frac{\Delta t_1({}^{B_0}\mathbf{p}_{C_2} - {}^{B_0}\mathbf{p}_{C_0})}{\Delta t_2(\Delta t_2 - \Delta t_1)} \\ \frac{-2}{\Delta t_2 - \Delta t_1}\left(\frac{{}^{B_0}\mathbf{p}_{C_1} - {}^{B_0}\mathbf{p}_{C_0}}{\Delta t_1} - \frac{{}^{B_0}\mathbf{p}_{C_2} - {}^{B_0}\mathbf{p}_{C_0}}{\Delta t_2}\right) \end{bmatrix} \tag{145}$$

B.2.4 Four images, one feature

In this section, we consider the minimal case where one feature is observed and tracked in four images. Proceeding similarly to Appendix B.2.3, we obtain a similar system of equations:

$$\begin{cases} \mathbf{a}_1 - \Delta t_0 \mathbf{a}_v - \frac{1}{2}\Delta t_0^2 \mathbf{a}_g + k_{01}\mathbf{p}_{01} &= \mathbf{0} \\ \mathbf{a}_1 - \Delta t_1 \mathbf{a}_v - \frac{1}{2}\Delta t_1^2 \mathbf{a}_g + k_{11}\mathbf{p}_{11} &= \mathbf{0} \\ \mathbf{a}_1 - \Delta t_2 \mathbf{a}_v - \frac{1}{2}\Delta t_2^2 \mathbf{a}_g + k_{21}\mathbf{p}_{21} &= \mathbf{0} \\ \mathbf{a}_1 - \Delta t_3 \mathbf{a}_v - \frac{1}{2}\Delta t_3^2 \mathbf{a}_g + k_{31}\mathbf{p}_{31} &= \mathbf{0} \end{cases} \tag{146}$$

and in linear-system form:

$$\underbrace{\begin{bmatrix} \mathbf{I}_3 & \Delta t_0 \mathbf{I}_3 & \Delta t_0{}^2 \mathbf{I}_3 & \mathbf{p}_{01} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{I}_3 & \Delta t_1 \mathbf{I}_3 & \Delta t_1{}^2 \mathbf{I}_3 & \mathbf{0}_{3\times1} & \mathbf{p}_{11} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{I}_3 & \Delta t_2 \mathbf{I}_3 & \Delta t_2{}^2 \mathbf{I}_3 & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{21} & \mathbf{0}_{3\times1} \\ \mathbf{I}_3 & \Delta t_3 \mathbf{I}_3 & \Delta t_3{}^2 \mathbf{I}_3 & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{31} \end{bmatrix}}_{\boldsymbol{\Gamma}} \underbrace{\begin{bmatrix} \mathbf{a}_1 \\ -\mathbf{a}_v \\ -\frac{1}{2}\mathbf{a}_g \\ k_{01} \\ k_{11} \\ k_{21} \\ k_{31} \end{bmatrix}}_{\mathbf{y}} = \mathbf{0} \tag{147}$$

where the vector $\mathbf{y}$ contains all the unknowns in the equations, and $\mathbf{p}_{ij}$ is defined as in Eq (127). We next identify the conditions under which the homogeneous equation $\boldsymbol{\Gamma}\mathbf{y} = \mathbf{0}$ has a non-zero solution by analyzing the rank of the 12×13 matrix $\boldsymbol{\Gamma}$. We perform the following elementary row and column operations on $\boldsymbol{\Gamma}$:

$$\boldsymbol{\Gamma} = \begin{bmatrix} \mathbf{I}_3 & \Delta t_0 \mathbf{I}_3 & \Delta t_0{}^2 \mathbf{I}_3 & \mathbf{p}_{01} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{I}_3 & \Delta t_1 \mathbf{I}_3 & \Delta t_1{}^2 \mathbf{I}_3 & \mathbf{0}_{3\times1} & \mathbf{p}_{11} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{I}_3 & \Delta t_2 \mathbf{I}_3 & \Delta t_2{}^2 \mathbf{I}_3 & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{21} & \mathbf{0}_{3\times1} \\ \mathbf{I}_3 & \Delta t_3 \mathbf{I}_3 & \Delta t_3{}^2 \mathbf{I}_3 & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{31} \end{bmatrix} \tag{148}$$

Subtract the first block row from other block rows, noting that $\Delta t_0 = t_0 - t_0 = 0$

$$\sim \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{p}_{01} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times3} & \Delta t_1 \mathbf{I}_3 & \Delta t_1{}^2 \mathbf{I}_3 & -\mathbf{p}_{01} & \mathbf{p}_{11} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times3} & \Delta t_2 \mathbf{I}_3 & \Delta t_2{}^2 \mathbf{I}_3 & -\mathbf{p}_{01} & \mathbf{0}_{3\times1} & \mathbf{p}_{21} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times3} & \Delta t_3 \mathbf{I}_3 & \Delta t_3{}^2 \mathbf{I}_3 & -\mathbf{p}_{01} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{p}_{31} \end{bmatrix} \tag{149}$$

Divide each block row by the corresponding Δt_i

$$\sim \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{p}_{01} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_1 \mathbf{I}_3 & -\frac{1}{\Delta t_1}\mathbf{p}_{01} & \frac{1}{\Delta t_1}\mathbf{p}_{11} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_2 \mathbf{I}_3 & -\frac{1}{\Delta t_2}\mathbf{p}_{01} & \mathbf{0}_{3\times1} & \frac{1}{\Delta t_2}\mathbf{p}_{21} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_3 \mathbf{I}_3 & -\frac{1}{\Delta t_3}\mathbf{p}_{01} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \frac{1}{\Delta t_3}\mathbf{p}_{31} \end{bmatrix} \tag{150}$$

Subtract the second block row from the lower block rows

$$\sim \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{p}_{01} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_1 \mathbf{I}_3 & -\frac{1}{\Delta t_1}\mathbf{p}_{01} & \frac{1}{\Delta t_1}\mathbf{p}_{11} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & (\Delta t_2 - \Delta t_1)\mathbf{I}_3 & (\frac{1}{\Delta t_1} - \frac{1}{\Delta t_2})\mathbf{p}_{01} & -\frac{1}{\Delta t_1}\mathbf{p}_{11} & \frac{1}{\Delta t_2}\mathbf{p}_{21} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & (\Delta t_3 - \Delta t_1)\mathbf{I}_3 & (\frac{1}{\Delta t_1} - \frac{1}{\Delta t_3})\mathbf{p}_{01} & -\frac{1}{\Delta t_1}\mathbf{p}_{11} & \mathbf{0}_{3\times1} & \frac{1}{\Delta t_3}\mathbf{p}_{31} \end{bmatrix} \tag{151}$$

Divide the last two block rows by the corresponding $(\Delta t_i - \Delta t_1)$

$$\sim \begin{bmatrix} \mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{p}_{01} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_1 \mathbf{I}_3 & -\frac{1}{\Delta t_1}\mathbf{p}_{01} & \frac{1}{\Delta t_1}\mathbf{p}_{11} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \frac{1}{\Delta t_2 \Delta t_1}\mathbf{p}_{01} & -\frac{1}{(\Delta t_2 - \Delta t_1)\Delta t_1}\mathbf{p}_{11} & \frac{1}{(\Delta t_2 - \Delta t_1)\Delta t_2}\mathbf{p}_{21} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \frac{1}{\Delta t_3 \Delta t_1}\mathbf{p}_{01} & -\frac{1}{(\Delta t_3 - \Delta t_1)\Delta t_1}\mathbf{p}_{11} & \mathbf{0}_{3\times1} & \frac{1}{(\Delta t_3 - \Delta t_1)\Delta t_3}\mathbf{p}_{31} \end{bmatrix} \tag{152}$$

Subtract the third block row from the fourth block row

$$\sim \left[\begin{array}{ccc|cccc} \mathbf{I}_3 & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{p}_{01} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times3} & \mathbf{I}_3 & \Delta t_1 \mathbf{I}_3 & -\frac{1}{\Delta t_1}\mathbf{p}_{01} & \frac{1}{\Delta t_1}\mathbf{p}_{11} & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} \\ \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_3 & \frac{1}{\Delta t_2 \Delta t_1}\mathbf{p}_{01} & -\frac{1}{(\Delta t_2 - \Delta t_1)\Delta t_1}\mathbf{p}_{11} & \frac{1}{(\Delta t_2 - \Delta t_1)\Delta t_2}\mathbf{p}_{21} & \mathbf{0}_{3\times1} \\ \hline \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & -\frac{\Delta t_3 - \Delta t_2}{\Delta t_3 \Delta t_2 \Delta t_1}\mathbf{p}_{01} & \frac{\Delta t_3 - \Delta t_2}{(\Delta t_3 - \Delta t_1)(\Delta t_2 - \Delta t_1)\Delta t_1}\mathbf{p}_{11} & -\frac{1}{(\Delta t_2 - \Delta t_1)\Delta t_2}\mathbf{p}_{21} & \frac{1}{(\Delta t_3 - \Delta t_1)\Delta t_3}\mathbf{p}_{31} \end{array}\right] \tag{153}$$

From Eq. (153), we have that:

$$\mathrm{rank}(\mathbf{\Gamma}) = 9 + \mathrm{rank}(\mathbf{B}) \tag{154}$$

with

$$\mathbf{B} = \begin{bmatrix} -\frac{\Delta t_3 - \Delta t_2}{\Delta t_3 \Delta t_2 \Delta t_1}\mathbf{p}_{01} & \frac{\Delta t_3 - \Delta t_2}{(\Delta t_3 - \Delta t_1)(\Delta t_2 - \Delta t_1)\Delta t_1}\mathbf{p}_{11} & -\frac{1}{(\Delta t_2 - \Delta t_1)\Delta t_2}\mathbf{p}_{21} & \frac{1}{(\Delta t_3 - \Delta t_1)\Delta t_3}\mathbf{p}_{31} \end{bmatrix} \tag{155}$$

$$\sim \begin{bmatrix} \mathbf{p}_{01} & \mathbf{p}_{11} & \mathbf{p}_{21} & \mathbf{p}_{31} \end{bmatrix} \tag{156}$$

In general, when the camera is moving randomly in 3D space, then the four 3×1 vectors $\{\mathbf{p}_{01}, \mathbf{p}_{11}, \mathbf{p}_{21}, \mathbf{p}_{31}\}$ span a subspace of dimension three, and $\mathrm{rank}(\mathbf{B}) = 3$. They span a subspace of dimension two when all those vectors lie in the same 2D plane, which is the case when all the camera positions and the feature are in the same plane. In that case, $\mathrm{rank}(\mathbf{B}) = 2$.

From Eq. (154), we conclude that, in general,

$$\mathrm{rank}(\mathbf{\Gamma}) = 9 + 3 = 12 \tag{157}$$

Based on the rank-nullity theorem, this implies that the solution space $\mathbf{y}$ of Eq. (147) has dimension of one. This means that the solution space of the equation $\mathbf{A}\mathbf{x} = \mathbf{0}$ has dimension of 1. This leads to the conclusion that when one feature is observed in four images, $\mathbf{A}$ is rank-deficient by one.

References

[1] E. Jones and S. Soatto, “Visual–inertial navigation, mapping and localization: A scalable real-time causal approach,” *The International Journal of Robotics Research*, 2011.

[2] J. Kelly and G. Sukhatme, “Visual-Inertial Sensor Fusion: Localization, Mapping and Sensor-to-Sensor Self-calibration,” *The International Journal of Robotics Research*, vol. 30, no. 1, p. 56, 2011.

[3] A. I. Mourikis and S. I. Roumeliotis, “A multi-state constraint Kalman filter for vision-aided inertial navigation,” in *Proceedings of the IEEE International Conference on Robotics and Automation*, Rome, Italy, Apr. 2007, pp. 3565–3572.

[4] M. Bryson and S. Sukkarieh, "Observability analysis and active control for airborne SLAM," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 44, no. 1, pp. 261 – 280, 2008.

[5] T. Lupton and S. Sukkarieh, "Efficient integration of inertial observations into visual slam without initialization," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009, pp. 1547–1552.

[6] D. D. Diel, P. DeBitetto, and S. Teller, "Epipolar constraints for vision-aided inertial navigation," in *Proceedings of the IEEE Workshop on Motion and Video Computing*, Jan. 2005, pp. 221–228.

[7] A. Martinelli, "Closed-Form Solution for Attitude and Speed Determination by Fusing Monocular Vision and Inertial Sensor Measurements," in *Proceedings of the IEEE International Conference on Robotics and Automation*, Shanghai, China, May 2011.

[8] H. Stewenius, C. Engels, and D. Nister, "Recent developments on direct relative orientation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 60, pp. 284–294, June 2006.

[9] O. Faugeras and S. Maybank, "Motion from point matches: Multiplicity of solutions," *International Journal of Computer Vision*, vol. 4, no. 3, pp. 225–246, July-Aug. 1990.

[10] F. Fraundorfer, P. Tanskanen, and M. Pollefeys, "A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles," in *Proceedings of the European conference on Computer vision*, 2010, pp. 269–282.

[11] M. Kalantari, A. Hashemi, F. Jung, and J.-P. Guedon, "A new solution to the relative orientation problem using only 3 points and the vertical direction," *Journal of Mathematical Imaging and Vision*, Nov. 2010.

[12] J. Shi and C. Tomasi, "Good features to track," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1994, pp. 593–600.

[13] A. B. Chatfield, *Fundamentals of High Accuracy Inertial Navigation*. American Institute of Aeronautics and Astronautics, Inc., 1997.

[14] J. A. Farrell, *Aided Navigation: GPS and High Rate Sensors*. McGraw-Hill, 2008.

[15] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. SIAM, 2000.

[16] J. Chou and M. Kamel, "Quaternions Approach to Solve the Kinematic Equation of Rotation AaAx= AxAb of a Sensor Mounted Robotic Manipulator," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 1988.

[17] ——, "Finding the position and orientation of a sensor on a robot manipulator using quaternions," *International Journal of Robotics Research*, vol. 10, no. 3, pp. 240–254, 1991.

[18] N. Trawny and S. Roumeliotis, "Indirect Kalman filter for 6D pose estimation," *University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep*, vol. 2, 2005.

[19] W. Gander, G. Golub, and U. Matt, "A constrained eigenvalue problem," *Linear Algebra and its applications*, vol. 114, pp. 815–839, 1989.

[20] E. Spjøtvoll, "A Note on a Theorem of Forsythe and Golub," *SIAM Journal on Applied Mathematics*, vol. 23, no. 3, pp. 307–311, 1972.

[21] B. Triggs, P. McLauchlan, R. Hartley, and Fitzgibbon, "Bundle adjustment – a modern synthesis," in *Vision Algorithms: Theory and Practice*. Springer Verlag, 2000, pp. 298–375.