Research Notes:
Covariance Intersection
Delayed Feature Initialization

Patrick Geneva

March 23, 2022

1 Problem Statement

In an extended Kalman filter [5], we wish initialize previously unseen state variables using feature measurements. This process is called *delayed initialization* since typically the process delayed to collect enough measurement observations to fully recover the to-be initialized state variable. In what follows we will first introduce two methods for performing delayed initialization, after which we will introduce the covariance intersection (CI) [2] update. We then re-derive the delayed initialization procedure when covariance intersection is leveraged.

1.1 Nonlinear Measurement Model

Consider the following nonlinear measurement function:

$$\mathbf{z}_m = h(\mathbf{x}_k) + \mathbf{n}_m \tag{1}$$

where we have the measurement noise $\mathbf{n}_m \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_m)$. For the standard EKF update, one linearizes the above equation at the current state estimate. In our case, as in the indirect EKF [4], we linearize (1) with respect to the current zero-mean error state (i.e. $\tilde{\mathbf{x}} = \mathbf{x} \boxminus \hat{\mathbf{x}} \sim \mathcal{N}(\mathbf{0}, \mathbf{P})$):

$$\mathbf{z}_{m,k} = h(\hat{\mathbf{x}}_k) + \mathbf{H}_k \tilde{\mathbf{x}}_k + \mathbf{n}_{m,k} \tag{2}$$

$$\Rightarrow \ \mathbf{r} = \mathbf{H}_k \tilde{\mathbf{x}}_k + \mathbf{n}_m \tag{3}$$

where $\mathbf{H}_k$ is the measurement Jacobian. Throughout this paper $\hat{\mathbf{x}}$ is used to denote the estimate of a random variable $\mathbf{x}$, while $\tilde{\mathbf{x}} = \mathbf{x} \boxminus \hat{\mathbf{x}}$ is the error in this estimate. The updated estimate from a correction $\boldsymbol{\delta}\mathbf{x}$ is $\hat{\mathbf{x}}^{\oplus} = \hat{\mathbf{x}} \boxplus \boldsymbol{\delta}\mathbf{x}$. Using this linearized measurement model, we can now perform the following standard EKF update to ensure the updated states remain on-manifold:

$$\hat{\mathbf{x}}_k^{\oplus} = \hat{\mathbf{x}}_k + \mathbf{K}_k \ \mathbf{r} \tag{4}$$

$$= \hat{\mathbf{x}}_k + \mathbf{K}_k(\mathbf{z}_m - h(\hat{\mathbf{x}}_k)) \tag{5}$$

$$\mathbf{P}_k^{\oplus} = \mathbf{P}_k - \mathbf{K}_k \mathbf{H}_k \mathbf{P}_k \tag{6}$$

$$\mathbf{K}_k = \mathbf{P}_k \mathbf{H}_k^{\top} \mathbf{S}_k^{-1} \tag{7}$$

$$\mathbf{S}_k^{-1} = (\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^{\top} + \mathbf{R}_m)^{-1} \tag{8}$$

1.2 Feature Bearing Observation Model

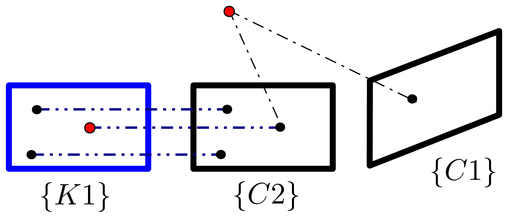

Figure 1: Illustration of the considered visual feature observation scenario. In this case, a historical keyframe $\{K1\}$ has been matched to an actively tracked feature (red). We wish to initialize this feature estimate into our state using all three measurements from the keyframe and poses $\{C1\}$ and $\{C2\}$.

We now consider a bit more concrete measurement model as shown in Figure 1. We consider we observe a 3d environmental feature with a camera from two camera clone poses, $\{C1\}$ and $\{C2\}$, along with a loop-closure measurement from a historical keyframe state $\{K1\}$. In general we have the following state vector:

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{x}_A^\top & \mathbf{x}_K^\top & {}^G\mathbf{p}_f^\top \end{bmatrix}^\top \tag{9}$$

$$\mathbf{x}_A = \begin{bmatrix} \mathbf{x}_{I_k}^\top & \mathbf{x}_C^\top \end{bmatrix}^\top \tag{10}$$

$$\mathbf{x}_K = \begin{bmatrix} \mathbf{x}_{T_1}^\top & \cdots & \mathbf{x}_{T_n}^\top \end{bmatrix}^\top \tag{11}$$

where we have:

$$\mathbf{x}_{I_k} = \begin{bmatrix} {}^{I_k}_G\bar{q}^\top & \mathbf{b}_{\omega_k}^\top & {}^G\mathbf{v}_{I_k}^\top & \mathbf{b}_{a_k}^\top & {}^G\mathbf{p}_{I_k}^\top \end{bmatrix}^\top \tag{12}$$

$$\mathbf{x}_C = \begin{bmatrix} \mathbf{x}_{T_{k-1}}^\top & \cdots & \mathbf{x}_{T_{k-c}}^\top \end{bmatrix}^\top \tag{13}$$

$$\mathbf{x}_{T_i} = \begin{bmatrix} {}^{I_i}_G\bar{q}^\top & {}^G\mathbf{p}_{I_i}^\top \end{bmatrix}^\top \tag{14}$$

where we define the "active" state $\mathbf{x}_A$ and map of n keyframe poses $\mathbf{x}_K$. The clone state $\mathbf{x}_C$ contains c historical IMU poses. ${}^I_G\bar{q}$ is the unit quaternion parameterizing the rotation ${}^I_G\mathbf{R}$ from the global frame of reference $\{G\}$ to the IMU local frame $\{I\}$ [6], $\mathbf{b}_\omega$ and $\mathbf{b}_a$ are the gyroscope and accelerometer biases, and ${}^G\mathbf{v}_I$ and ${}^G\mathbf{p}_I$ are the velocity and position of the IMU expressed in the global frame, respectively. In the case of delayed initialization we do not have the feature estimate ${}^G\mathbf{p}_f$ mean, its uncertainty, and its correlation with the rest of the state yet.

We consider a bearing measurement $\mathbf{z}$ seen at timestep i can be related to the state by the following (simplified for presentation, model in [1] is used):

$$\mathbf{z}_i = \mathbf{h}(\mathbf{x}_{T_i}, {}^G\mathbf{p}_f) + \mathbf{n}_i \tag{15}$$

$$= \boldsymbol{\Lambda}({}^{C_i}\mathbf{p}_f) + \mathbf{n}_i \tag{16}$$

$$\boldsymbol{\Lambda}([x\ y\ z]^\top) = \begin{bmatrix} x/z & y/z \end{bmatrix}^\top \tag{17}$$

$${}^{C_i}\mathbf{p}_f = {}^C_I\mathbf{R}\,{}^{I_i}_G\mathbf{R}({}^G\mathbf{p}_f - {}^G\mathbf{p}_{I_i}) + {}^C\mathbf{p}_I \tag{18}$$

where $\mathbf{n}_i$ is the white Gaussian noise with covariance $\mathbf{R}_i = \sigma_{pix}^2\mathbf{I}$. We can now linearize this measurement model and obtain the following residual:

$$\mathbf{r}_i = \mathbf{z}_i - \mathbf{h}(\hat{\mathbf{x}}_{T_i}, {}^G\hat{\mathbf{p}}_f) \tag{19}$$

$$\simeq \mathbf{H}_{T_i}\tilde{\mathbf{x}}_{T_i} + \mathbf{H}_{f_i}{}^G\tilde{\mathbf{p}}_f + \mathbf{n}_i \tag{20}$$

where $\mathbf{H}_{T_i}$ and $\mathbf{H}_{f_i}$ are the measurement Jacobians, and $\tilde{\mathbf{x}}_{T_i}$ and ${}^G\tilde{\mathbf{p}}_f$ are the error states for the observation pose and feature, respectively. After sufficient observations of the feature, we can "stack" them as:

$$\begin{aligned}
\mathbf{r} &= \mathbf{H}_T\tilde{\mathbf{x}}_{T_{1..c}} + \mathbf{H}_{T_k}\tilde{\mathbf{x}}_{T_k} + \mathbf{H}_f{}^G\tilde{\mathbf{p}}_f + \mathbf{n} && (21)\\
&= \mathbf{H}_a\tilde{\mathbf{x}}_A + \mathbf{H}_k\tilde{\mathbf{x}}_K + \mathbf{H}_f{}^G\tilde{\mathbf{p}}_f + \mathbf{n} && (22)\\
&= \underbrace{\begin{bmatrix}\mathbf{H}_a & \mathbf{H}_k\end{bmatrix}}_{\mathbf{H}_x}\begin{bmatrix}\tilde{\mathbf{x}}_A\\ \tilde{\mathbf{x}}_K\end{bmatrix} + \mathbf{H}_f{}^G\tilde{\mathbf{p}}_f + \mathbf{n} && (23)
\end{aligned}$$

where the measurement is a function of c clone poses, $\tilde{\mathbf{x}}_{T_{1..c}} = [\tilde{\mathbf{x}}_{T_1}^\top \cdots \tilde{\mathbf{x}}_{T_c}^\top]^\top$, corresponding to each non-keyframe observation time the feature was seen, and the stacked measurement noise is $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ where $\mathbf{R} = \sigma_{pix}^2\mathbf{I}$. Additionally, the loop-closure measurement from the keyframe introduces the Jacobian in respect to $\mathbf{x}_{T_k} \in \mathbf{x}_K$.

2 EKF-based Delayed Initialization

2.1 Method 1: Two System Invertible

Based on the stacked linearized measurement equation, Eq. (21), we aim to optimally compute the initial estimate of a new state variable and its covariance and correlations with the existing state variables. As derived by Mingyang Li [3] we first perform QR decomposition (e.g., using computationally efficient in-place Givens rotations) to separate the linear system into two subsystems: (i) one that depends on the new state (i.e., ${}^G\mathbf{p}_f$), and (ii) the other that does not.

$$\begin{aligned}
\mathbf{r} &= \begin{bmatrix}\mathbf{H}_x & \mathbf{H}_f\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{x}}_k\\ {}^G\tilde{\mathbf{p}}_f\end{bmatrix} + \mathbf{n} && (24)\\
\Rightarrow \begin{bmatrix}\mathbf{r}_1\\ \mathbf{r}_2\end{bmatrix} &= \begin{bmatrix}\mathbf{H}_{x1} & \mathbf{H}_{f1}\\ \mathbf{H}_{x2} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{x}}_k\\ {}^G\tilde{\mathbf{p}}_f\end{bmatrix} + \begin{bmatrix}\mathbf{n}_{f1}\\ \mathbf{n}_{f2}\end{bmatrix} && (25)
\end{aligned}$$

where $\mathbf{n}_{fi} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{fi})$, $i \in \{1, 2\}$. Note that in the above expression $\mathbf{r}_1$ and $\mathbf{r}_2$ are orthonormally transformed measurement residuals, not the direct partitions of $\mathbf{r}$. With the *top* transformed linearized measurement residual $\mathbf{r}_1$ in Eq. (25), we now initialize the state estimate of ${}^G\mathbf{p}_f$ and its covariance and correlations to $\mathbf{x}_k$ [see Eq. (5)], which will then be augmented to the current state and covariance matrix.

$$\begin{aligned}
{}^G\tilde{\mathbf{p}}_f &= \mathbf{H}_{f1}^{-1}(\mathbf{r}_1 - \mathbf{n}_1 - \mathbf{H}_x\tilde{\mathbf{x}}) && (26)\\
\Rightarrow\ \mathbb{E}[{}^G\tilde{\mathbf{p}}_f] &= \mathbf{H}_{f1}^{-1}(\mathbf{r}_1) && (27)
\end{aligned}$$

$$\begin{aligned}
\mathbf{P}_{ff} &= \mathbb{E}\Big[({}^G\tilde{\mathbf{p}}_f - \mathbb{E}[{}^G\tilde{\mathbf{p}}_f])({}^G\tilde{\mathbf{p}}_f - \mathbb{E}[{}^G\tilde{\mathbf{p}}_f])^\top\Big] && (28)\\
&= \mathbb{E}\Big[(\mathbf{H}_{f1}^{-1}(-\mathbf{n}_1 - \mathbf{H}_{x1}\tilde{\mathbf{x}}))(\mathbf{H}_{f1}^{-1}(-\mathbf{n}_1 - \mathbf{H}_{x1}\tilde{\mathbf{x}}))^\top\Big] && (29)\\
&= \mathbf{H}_{f1}^{-1}(\mathbf{H}_{x1}\mathbf{P}_{xx}\mathbf{H}_{x1}^\top + \mathbf{R}_1)\mathbf{H}_{f1}^{-\top} && (30)
\end{aligned}$$

$$\begin{aligned}
\mathbf{P}_{xf} &= \mathbb{E}\Big[(\tilde{\mathbf{x}})({}^G\tilde{\mathbf{p}}_f - \mathbb{E}[{}^G\tilde{\mathbf{p}}_f])^\top\Big] && (31)\\
&= \mathbb{E}\Big[(\tilde{\mathbf{x}})(\mathbf{H}_{f1}^{-1}(-\mathbf{n}_1 - \mathbf{H}_{x1}\tilde{\mathbf{x}}))^\top\Big] && (32)
\end{aligned}$$

$$= -\mathbf{P}_{xx}\mathbf{H}_{x1}^{\top}\mathbf{H}_{f1}^{-\top} \tag{33}$$

where $\mathbb{E}[\cdot]$ is the expectation operator. These derivations can be summarized as follows:

$$^{G}\mathbf{p}_f^{\oplus} = {}^{G}\mathbf{p}_f + \mathbf{H}_{f1}^{-1}\mathbf{r}_1 \tag{34}$$

$$\mathbf{P}_{xx}^{\oplus} = \mathbf{P}_{xx} \tag{35}$$
$$\mathbf{P}_{ff}^{\oplus} = \mathbf{H}_{f1}^{-1}(\mathbf{H}_{x1}\mathbf{P}_{xx}\mathbf{H}_{x1}^{\top} + \mathbf{R}_{f1})\mathbf{H}_{f1}^{-\top} \tag{36}$$
$$\mathbf{P}_{xf}^{\oplus} = -\mathbf{P}_{xx}\mathbf{H}_{x1}^{\top}\mathbf{H}_{f1}^{-\top} \tag{37}$$
$$\mathbf{P}_{fx}^{\oplus} = (\mathbf{P}_{xf}^{\oplus})^{\top} \tag{38}$$

It should be noted that a full-rank $\mathbf{H}_{f1}$ is needed to perform the above initialization, which normally is the case if enough measurements are collected (i.e., delayed initialization). Note also that to utilize all available measurement information, we also perform EKF update using the *bottom* measurement residual $\mathbf{r}_2$ in Eq. (25).

2.2 Method 2: Infinite Uncertainty with Update

We now look at an alternate formulation for delayed initialization. In this method, we consider the case were the state already has a prior covariance of the state-to-be-initialized but its uncertainty is at infinity and has not been correlated with the current state through a measurement yet. More concretely, we define the following covariance:

$$\mathbf{P}_k = \begin{bmatrix} \mathbf{P}_{xx} & \mathbf{0} \\ \mathbf{0} & \mu\mathbf{I} \end{bmatrix} \tag{39}$$

where we have defined $\mathbf{P}_{ff} = \mu\mathbf{I}$ with $\mu \to \infty$ since we have no prior knowledge of the feature's state. We now wish to perform an EKF update, see Eq. (6), using the measurement information collected. We define the stacked measurements as:

$$\mathbf{r} = \mathbf{H}_k\tilde{\mathbf{x}}_k + \mathbf{n} \tag{40}$$
$$= \begin{bmatrix} \mathbf{H}_x & \mathbf{H}_f \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_A \\ {}^{G}\tilde{\mathbf{p}}_f \end{bmatrix} + \mathbf{n} \tag{41}$$

This gives us the following update equations:

$$\mathbf{x}_A^{\oplus} = \mathbf{x}_A + \mathbf{K}_x\mathbf{r} \tag{42}$$
$$^{G}\mathbf{p}_f^{\oplus} = {}^{G}\mathbf{p}_f + \mathbf{K}_f\mathbf{r} \tag{43}$$

$$\mathbf{P}_k^{\oplus} = \mathbf{P}_k - \begin{bmatrix} \mathbf{K}_x\mathbf{S}_k\mathbf{K}_x^{\top} & \mathbf{K}_x\mathbf{H}_k \begin{bmatrix} \mathbf{P}_{xf} \\ \mathbf{P}_{ff} \end{bmatrix} \\ \begin{bmatrix} \mathbf{P}_{xf} \\ \mathbf{P}_{ff} \end{bmatrix}^{\top} \mathbf{H}_k{}^{\top}\mathbf{K}_x^{\top} & \mathbf{K}_f\mathbf{S}_k\mathbf{K}_f^{\top} \end{bmatrix} \tag{44}$$

$$= \begin{bmatrix} \mathbf{P}_{xx} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{ff} \end{bmatrix} - \begin{bmatrix} \mathbf{K}_x\mathbf{S}_k\mathbf{K}_x^{\top} & \mathbf{K}_x\mathbf{H}_f\mathbf{P}_{ff} \\ \mathbf{P}_{ff}\mathbf{H}_f{}^{\top}\mathbf{K}_x^{\top} & \mathbf{K}_f\mathbf{S}_k\mathbf{K}_f^{\top} \end{bmatrix} \tag{45}$$

where we have used that the initial feature is uncorrelated with the state (i.e., $\mathbf{P}_{xf} = \mathbf{P}_{xf}^\top = \mathbf{0}$) and we have defined the following Kalman gains:

$$\mathbf{K}_k = \begin{bmatrix} \mathbf{K}_x \\ \mathbf{K}_f \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{xx}\mathbf{H}_x^\top + \mathbf{P}_{xf}\mathbf{H}_f^\top \\ \mathbf{P}_{fx}\mathbf{H}_x^\top + \mathbf{P}_{ff}\mathbf{H}_f^\top \end{bmatrix} \mathbf{S}_k^{-1} := \begin{bmatrix} \mathbf{P}_{xx}\mathbf{H}_x^\top \\ \mathbf{P}_{ff}\mathbf{H}_f^\top \end{bmatrix} \mathbf{S}_k^{-1} \tag{46}$$

We now first look at how to calculate the measurement innovation term. It is as follows:

$$\begin{aligned}
\mathbf{S}_k^{-1} &= \left(\mathbf{H}_k\mathbf{P}_k\mathbf{H}_k^\top + \mathbf{R}_m\right)^{-1} && (47)\\
&= \left(\mathbf{H}_x\mathbf{P}_{xx}\mathbf{H}_x^\top + \mathbf{H}_f\mathbf{P}_{ff}\mathbf{H}_f^\top + \mathbf{R}_m\right)^{-1} && (48)\\
&= \left(\mathbf{A} + \mathbf{H}_f\mathbf{P}_{ff}\mathbf{H}_f^\top\right)^{-1} && (49)\\
&= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{H}_f\left(\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f + \cancel{\mathbf{P}_{ff}^{-1}}\right)^{-1}\mathbf{H}_f^\top\mathbf{A}^{-1} && (50)\\
&= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{H}_f\left(\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f\right)^{-1}\mathbf{H}_f^\top\mathbf{A}^{-1} && (51)
\end{aligned}$$

where we have defined $\mathbf{A} = \mathbf{H}_x\mathbf{P}_{xx}\mathbf{H}_x^\top + \mathbf{R}_m$, and $\mathbf{P}_{ff}^{-1} = (\mu\mathbf{I})^{-1} \to \mathbf{0}$ when $\mu \to \infty$, and the matrix inversion lemma as:

$$(\mathbf{A} + \mathbf{U}\mathbf{C}\mathbf{V})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}\left(\mathbf{V}\mathbf{A}^{-1}\mathbf{U} + \mathbf{C}^{-1}\right)^{-1}\mathbf{V}\mathbf{A}^{-1} \tag{52}$$

This leads the following conclusion for $\mathbf{P}_{xx}$:

$$\begin{aligned}
\mathbf{P}_{xx}^{\oplus} &= \mathbf{P}_{xx} - \mathbf{P}_{xx}\mathbf{H}_x^\top\mathbf{S}_k^{-1}\mathbf{H}_x\mathbf{P}_{xx} && (53)\\
&= \mathbf{P}_{xx} - \mathbf{P}_{xx}\mathbf{H}_x^\top\left(\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{H}_f\left(\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f\right)^{-1}\mathbf{H}_f^\top\mathbf{A}^{-1}\right)\mathbf{H}_x\mathbf{P}_{xx} && (54)
\end{aligned}$$

Now we look at how to compute updated feature uncertainty $\mathbf{P}_{ff}$. We have the following:

$$\begin{aligned}
\mathbf{P}_{ff}^{\oplus} &= \mathbf{P}_{ff} - \mathbf{K}_f\mathbf{S}_k\mathbf{K}_f^\top && (55)\\
&= \mathbf{P}_{ff} - \mathbf{P}_{ff}\mathbf{H}_f^\top\mathbf{S}_k^{-1}\mathbf{H}_f\mathbf{P}_{ff} && (56)\\
&= \mathbf{P}_{ff} + \mathbf{P}_{ff}\mathbf{H}_f^\top(-\mathbf{S}_k^{-1})\mathbf{H}_f\mathbf{P}_{ff} && (57)\\
&= \left(\cancel{\mathbf{P}_{ff}^{-1}} - \cancel{\mathbf{P}_{ff}^{-1}\mathbf{P}_{ff}}\mathbf{H}_f^\top\left((-\mathbf{S}_k) + \mathbf{H}_f\mathbf{P}_{ff}\cancel{\mathbf{P}_{ff}^{-1}\mathbf{P}_{ff}}\mathbf{H}_f^\top\right)^{-1}\mathbf{H}_f\cancel{\mathbf{P}_{ff}\mathbf{P}_{ff}^{-1}}\right)^{-1} && (58)\\
&= \left(-\mathbf{H}_f^\top\left((-\mathbf{S}_k) + \mathbf{H}_f\mathbf{P}_{ff}\mathbf{H}_f^\top\right)^{-1}\mathbf{H}_f\right)^{-1} && (59)\\
&= \left(-\mathbf{H}_f^\top\left((-\mathbf{A} - \mathbf{H}_f\mathbf{P}_{ff}\mathbf{H}_f^\top) + \mathbf{H}_f\mathbf{P}_{ff}\mathbf{H}_f^\top\right)^{-1}\mathbf{H}_f\right)^{-1} && (60)\\
&= \left(\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f\right)^{-1} && (61)\\
& && (62)
\end{aligned}$$

This leads the following conclusion for $\mathbf{P}_{ff}$:

$$\mathbf{P}_{ff}^{\oplus} = \left(\mathbf{H}_f^\top \mathbf{A}^{-1}\mathbf{H}_f\right)^{-1} \tag{63}$$

Now we look at how to compute feature's correlation with the state $\mathbf{P}_{xf}$. We have the following:

$$\mathbf{P}_{xf}^{\oplus} = \cancel{\mathbf{P}_{xf}} - \mathbf{K}_x\mathbf{H}_f\mathbf{P}_{ff} \tag{64}$$

$$= -\mathbf{P}_{xx}\mathbf{H}_x^\top\mathbf{S}_k^{-1}\mathbf{H}_f\mathbf{P}_{ff} \tag{65}$$

Looking at the last three terms and substituting in the equality from Eq. (50) (the only part that is a function of $\mathbf{P}_{ff}$) we have:

$$\mathbf{S}_k^{-1}\mathbf{H}_f\mathbf{P}_{ff} = \left(\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{H}_f\left(\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f + \mathbf{P}_{ff}^{-1}\right)^{-1}\mathbf{H}_f^\top\mathbf{A}^{-1}\right)\mathbf{H}_f\mathbf{P}_{ff} \tag{66}$$

$$= \mathbf{A}^{-1}\mathbf{H}_f\left(\mathbf{I} - \left(\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f + \mathbf{P}_{ff}^{-1}\right)^{-1}\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f\right)\mathbf{P}_{ff} \tag{67}$$

$$= \mathbf{A}^{-1}\mathbf{H}_f\left(\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f + \mathbf{P}_{ff}^{-1}\right)^{-1}\left[\left(\cancel{\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f} + \mathbf{P}_{ff}^{-1}\right) - \cancel{\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f}\right]\mathbf{P}_{ff} \tag{68}$$

$$= \mathbf{A}^{-1}\mathbf{H}_f\left(\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f + \cancel{\mathbf{P}_{ff}^{-1}}\right)^{-1} \tag{69}$$

$$= \mathbf{A}^{-1}\mathbf{H}_f\left(\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f\right)^{-1} \tag{70}$$

This leads the following conclusion for $\mathbf{P}_{xf}$:

$$\mathbf{P}_{xf}^{\oplus} = -\mathbf{P}_{xx}\mathbf{H}_x^\top\mathbf{A}^{-1}\mathbf{H}_f\left(\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f\right)^{-1} \tag{71}$$

The state means can be updated similarly.

2.3 Method Equivalence

A natural question is the equivalence between these two methods. Just by looking at Method's 1 Eq. (35) and Method's 2 Eq. (54) one can see that Method 2 updates the original covariance while the first method does not! At first glance this would mean that the two methods are not doing the exact same thing, and that one is better than the other. There is a subtle difference between the two: Method 1 first initializes with a sub-system of the full measurement, while the Method 2 initializes the prior information with all measurements. We can show that Method 2 is exactly the same as the first by considering we have a square measurement Jacobian that is invertible $\mathbf{H}_f\mathbf{H}_f^{-1} = \mathbf{H}_f^{-1}\mathbf{H}_f = \mathbf{I}$ (thus there is no second update using $\mathbf{r}_2$ in method 1). We get:

$$\mathbf{P}_{xx}^{\oplus} = \mathbf{P}_{xx} - \mathbf{P}_{xx}\mathbf{H}_x^\top\left(\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{H}_f\left(\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f\right)^{-1}\mathbf{H}_f^\top\mathbf{A}^{-1}\right)\mathbf{H}_x\mathbf{P}_{xx} \tag{72}$$

$$= \mathbf{P}_{xx} - \mathbf{P}_{xx}\mathbf{H}_x^\top\left(\mathbf{A}^{-1} - \mathbf{A}^{-1}\cancel{\mathbf{H}_f\mathbf{H}_f^{-1}}\mathbf{A}\cancel{\mathbf{H}_f^{-\top}\mathbf{H}_f^\top}\mathbf{A}^{-1}\right)\mathbf{H}_x\mathbf{P}_{xx} \tag{73}$$

$$
\begin{aligned}
&= \mathbf{P}_{xx} - \mathbf{P}_{xx}\mathbf{H}_x^\top \left(\mathbf{A}^{-1} - \cancel{\mathbf{A}^{-1}\mathbf{A}}\mathbf{A}^{-1}\right)\mathbf{H}_x\mathbf{P}_{xx} && (74)\\
&= \mathbf{P}_{xx} - \cancel{\mathbf{P}_{xx}\mathbf{H}_x^\top \left(\mathbf{A}^{-1} - \mathbf{A}^{-1}\right)\mathbf{H}_x\mathbf{P}_{xx}} && (75)\\
&= \mathbf{P}_{xx} && (76)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{P}_{ff}^{\oplus} &= \left(\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f\right)^{-1} && (77)\\
&= \mathbf{H}_f^{-1}\left(\mathbf{H}_x\mathbf{P}_{xx}\mathbf{H}_x^\top + \mathbf{R}_m\right)\mathbf{H}_f^{-\top} && (78)
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{P}_{xf}^{\oplus} &= -\mathbf{P}_{xx}\mathbf{H}_x^\top\mathbf{A}^{-1}\mathbf{H}_f\left(\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f\right)^{-1} && (79)\\
&= -\mathbf{P}_{xx}\mathbf{H}_x^\top\mathbf{A}^{-1}\cancel{\mathbf{H}_f\mathbf{H}_f^{-1}}\mathbf{A}\mathbf{H}_f^{-\top} && (80)\\
&= -\mathbf{P}_{xx}\mathbf{H}_x^\top\cancel{\mathbf{A}^{-1}}\mathbf{A}\mathbf{H}_f^{-\top} && (81)\\
&= -\mathbf{P}_{xx}\mathbf{H}_x^\top\mathbf{H}_f^{-\top} && (82)
\end{aligned}
$$

To explain it in an intuitive way, if you have a square matrix, there is enough measurement information to recover the state you wish to initialize. But just having this information does not allow you to improve your state estimate (decrease $\mathbf{P}_{xx}$). Once you have more measurements then what is required to initialize it then you can improve it (e.g., non empty secondary system in Method 1).

3 Covariance Intersection-based Delayed Initialization

3.1 Covariance Intersection State Update

To guarantee consistency when updating with this measurement, we adopt the CI-EKF update [2] to construct a prior covariance such that:

$$
\mathbf{Diag}\left(\frac{1}{\omega_a}\mathbf{P}_{aa}, \frac{1}{\omega_1}\mathbf{P}_1, \cdots, \frac{1}{\omega_n}\mathbf{P}_n\right) \geq \mathbf{P}_k \qquad (83)
$$

where the left side is the CI covariance with zero off-diagonal elements and the right hand side is the unknown true covariance of the state with cross-covariances. The weights $\omega_l > 0$ and $\sum_l \omega_l = 1$, for $l \in \{a, 1...n\}$, can be found optimally [2]. The first weight correspond to the "active" covariance, while the remainder correspond to each keyframe forwhich we only keep their marginal covariance and do not track their correlations with the active state elements.

Substituting Eq. (83) into the standard EKF equations and only selecting the portion that updates active state yields (that is, we do not update keyframe states in the prior map):

$$
\begin{aligned}
\mathbf{x}_A^{\oplus} &= \mathbf{x}_A + \frac{1}{\omega_a}\mathbf{P}_{aa}\mathbf{H}_a^\top\mathbf{S}_k^{-1}\mathbf{r} && (84)\\
\mathbf{P}_{aa}^{\oplus} &= \frac{1}{\omega_a}\mathbf{P}_{aa} - \frac{1}{\omega_a^2}\mathbf{P}_{aa}\mathbf{H}_a^\top\mathbf{S}_k^{-1}\mathbf{H}_a\mathbf{P}_{aa} && (85)\\
\mathbf{S}_k &= \sum_{o\in\{a,1...n\}}\left(\frac{1}{\omega_o}\mathbf{H}_o\mathbf{P}_{oo}\mathbf{H}_o^\top\right) + \mathbf{R}_m && (86)
\end{aligned}
$$

3.2 Delayed Initialization

We can follow the logic presented in the previous segments. Specifically, we can start with the following covariance matrix:

$$\mathbf{P}_k = \begin{bmatrix} \frac{1}{\omega_a}\mathbf{P}_{aa} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \frac{1}{\omega_o}\mathbf{P}_{oo} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mu\mathbf{I} \end{bmatrix} \tag{87}$$

We can see that the CI variables only show up in the $\mathbf{S}_k$ term and can be grouped into the value $\mathbf{A}$ as before (see Eq. (48)).

$$\mathbf{A} = \frac{1}{\omega_a}\mathbf{H}_a\mathbf{P}_{aa}\mathbf{H}_a^\top + \sum_{o\in\{1\dots n\}}\left(\frac{1}{\omega_o}\mathbf{H}_o\mathbf{P}_{oo}\mathbf{H}_o^\top\right) + \mathbf{R}_m \tag{88}$$

We can then perform an update using Eq. (84)-(86) to get the following:

$$\mathbf{P}_{xx}^\oplus = \frac{1}{\omega_a}\mathbf{P}_{aa} - \frac{1}{\omega_a^2}\mathbf{P}_{aa}\mathbf{H}_a^\top\left(\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{H}_f\left(\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f\right)^{-1}\mathbf{H}_f^\top\mathbf{A}^{-1}\right)\mathbf{H}_a\mathbf{P}_{aa} \tag{89}$$

$$\mathbf{P}_{ff}^\oplus = \left(\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f\right)^{-1} \tag{90}$$

$$\mathbf{P}_{af}^\oplus = -\frac{1}{\omega_a}\mathbf{P}_{aa}\mathbf{H}_a^\top\mathbf{A}^{-1}\mathbf{H}_f\left(\mathbf{H}_f^\top\mathbf{A}^{-1}\mathbf{H}_f\right)^{-1} \tag{91}$$

We can then equate this result to Method's 1 structure to get:

$${}^G\mathbf{p}_f^\oplus = {}^G\mathbf{p}_f + \mathbf{H}_f^{-1}\mathbf{r}_1 \tag{92}$$

$$\mathbf{P}_{aa}^\oplus = \frac{1}{\omega_a}\mathbf{P}_{aa} \tag{93}$$

$$\mathbf{P}_{ff}^\oplus = \mathbf{H}_f^{-1}\left[\frac{1}{\omega_a}\mathbf{H}_a\mathbf{P}_{aa}\mathbf{H}_a^\top + \sum_{o\in\{1\dots n\}}\left(\frac{1}{\omega_o}\mathbf{H}_o\mathbf{P}_{oo}\mathbf{H}_o^\top\right) + \mathbf{R}_m\right]\mathbf{H}_f^{-\top} \tag{94}$$

$$\mathbf{P}_{af}^\oplus = -\frac{1}{\omega_a}\mathbf{P}_{aa}\mathbf{H}_a^\top\mathbf{H}_f^{-\top} \tag{95}$$

$$\mathbf{P}_{fa}^\oplus = (\mathbf{P}_{af}^\oplus)^\top \tag{96}$$

References

[1] Patrick Geneva et al. "OpenVINS: A Research Platform for Visual-Inertial Estimation". In: *Proc. of the IEEE International Conference on Robotics and Automation*. Paris, France, 2020. URL: https://github.com/rpng/open_vins.

[2] SJ Julier and Jeffrey K Uhlmann. "General decentralized data fusion with covariance intersection". In: *Handbook of multisensor data fusion: theory and practice* (2009), pp. 319–344.

[3] Mingyang Li. "Visual-Inertial Odometry on Resource-Constrained Systems". PhD thesis. UC Riverside, 2014.

[4] Peter S. Maybeck. *Stochastic Models, Estimation, and Control*. Vol. 1. London: Academic Press, 1979.

[5] Anastasios I Mourikis and Stergios I Roumeliotis. "A multi-state constraint Kalman filter for vision-aided inertial navigation". In: *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE. 2007, pp. 3565–3572.

[6] Nikolas Trawny and Stergios I. Roumeliotis. *Indirect Kalman Filter for 3D Attitude Estimation*. Tech. rep. University of Minnesota, Dept. of Comp. Sci. & Eng., Mar. 2005.