2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

C-KLAM: Constrained Keyframe-Based Localization and Mapping

Esha D. Nerurkar[†], Kejian J. Wu[‡], and Stergios I. Roumeliotis[†]

***Abstract*— In this paper, we present C-KLAM, a Maximum A Posteriori (MAP) estimator-based keyframe approach for SLAM. Instead of discarding information from non-keyframes for reducing the computational complexity, the proposed C-KLAM presents a novel, elegant, and computationally-efficient technique for incorporating most of this information in a consistent manner, resulting in improved estimation accuracy. To achieve this, C-KLAM projects both proprioceptive and exteroceptive information from the non-keyframes to the keyframes, using marginalization, while *maintaining the sparse structure of the associated information matrix*, resulting in fast and efficient solutions. The performance of C-KLAM has been tested in experiments, using visual and inertial measurements, to demonstrate that it achieves performance comparable to that of the computationally-intensive batch MAP-based 3D SLAM, that uses all available measurement information.**

I. INTRODUCTION AND RELATED WORK

One of the main challenges in designing an estimation algorithm for large-scale Simultaneous Localization and Mapping (SLAM) is its inherently high computational complexity. For example, the computational complexity of the Minimum Mean Squared Error (MMSE) estimator for SLAM, i.e., the Extended Kalman filter [1], is $O(N^2)$ at each time step, where N is the number of landmarks in the map. Similarly, for the batch Maximum A Posteriori (MAP) estimator-based SLAM (smoothing and mapping) [2], the worst-case computational complexity is $O([K+N]^3)$, where K is the number of robot poses in the trajectory. While existing batch MAP-based SLAM approaches such as the $\sqrt{SAM}$ [2], g^2o [3], and SPA [4] generate efficient solutions by exploiting the sparsity of the information matrix, for large-scale SLAM with frequent loop closures, this cost eventually prohibits real-time operation.

The approximate solutions developed to reduce MAP-based SLAM's computational complexity can be classified into three main categories. The first category of approaches such as iSAM [5] and iSAM2 [6] *incrementally* optimize over all robot poses and landmarks, using *all* available measurement information. However, for trajectories with frequent loop closures, (i) nonzero fill-ins into the information matrix are generated between periodic batch updates for iSAM, when the number of constraints is greater than five times the number of robot poses [5], and (ii) many nodes in the Bayes tree used by iSAM2 have to be relinearized, hence degrading the performance of these approaches. The graphical SLAM approach of [7] provides efficient solutions by employing block coordinate descent-based minimization and by postponing relinearization. Besides the approximation used for minimizing the cost function, this method's accuracy also suffers due to the accumulation of linearization errors when frequent loop closures occur.

The second category includes fixed-lag smoothing approaches such as [8], [9] that consider a constant-size, sliding-window of recent robot poses and landmarks, along with measurements only in that time window. Here, old robot poses and landmarks are *marginalized* and the corresponding measurements are discarded. However, marginalization destroys the sparsity of the information matrix, and the cost of this approach becomes $O(R^3)$, hence limiting the number of poses, R, in the sliding window. Moreover, this approach is unable to close loops for long trajectories.

The third category consists of *keyframe*-based approaches, such as PTAM [10]. PTAM processes measurement information from only a *subset* of all available views, hence information from non-keyframes is *discarded* (as opposed to marginalized) in order to retain the sparsity of the information matrix. Keyframe-based pose-graph approaches [11], [12], [13], [14], [15], [16], on the other hand, make use of all information from both key and non-key frames, but measurements from each frame are used multiple times to generate relative pose-to-pose constraints, especially in loop closure events. Re-using information results in inconsistent estimates, hence degrading the estimation accuracy.

In this paper, we present the Constrained Keyframe-based Localization and Mapping (C-KLAM), an approximate batch MAP-based algorithm, which estimates only keyframes (key robot poses) and key landmarks while also exploiting information (e.g., visual observations and odometry measurements) available to the non-keyframes. In particular, this information is projected onto the keyframes, by generating consistent pose constraints between them. Our main contributions are as follows:

- C-KLAM utilizes both proprioceptive [e.g., inertial measurement unit (IMU)] and exteroceptive (e.g., camera) measurements from non-keyframes to generate pose constraints between the keyframes in a consistent manner. This is achieved by marginalizing the non-keyframes along with the landmarks observed from them.
- In contrast to sliding-window approaches, C-KLAM incorporates information from marginalized frames and landmarks *without* destroying the sparsity of the information matrix, and hence generates fast and efficient

[†]E. D. Nerurkar, and S. I. Roumeliotis are with the Department of Computer Science and Engineering, Univ. of Minnesota, Minneapolis, USA {nerurkar,stergios}@cs.umn.edu

[‡]K. J. Wu is with the Department of Electrical and Computer Engineering, Univ. of Minnesota, Minneapolis, USA kejian@cs.umn.edu

This work was supported by the University of Minnesota through the Digital Technology Center (DTC), and AFOSR (FA9550-10-1-0567).

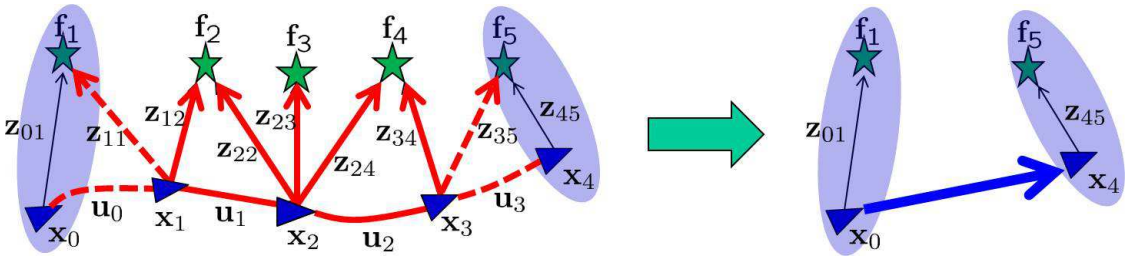

Fig. 1: An example of the exploration epoch before (left) and after (right) the approximation employed in C-KLAM. $\mathbf{x}_0$, $\mathbf{x}_4$ are the keyframes to be retained, and $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ are the non-keyframes to be marginalized. Similarly, $\mathbf{f}_1$, $\mathbf{f}_5$ are key landmarks (observed from the keyframes) to be retained, while $\mathbf{f}_2$, $\mathbf{f}_3$, and $\mathbf{f}_4$ are non-key landmarks (observed exclusively from the non-keyframes) to be marginalized. In the left figure, the arrows denote the measurements between different states. In the right figure, the blue arrow represents the pose constraint generated between the keyframes using C-KLAM.

solutions.

- The cost of marginalization in C-KLAM is cubic, $\mathcal{O}(M_r^3)$, only in the number of non-keyframes, M_r, between consecutive keyframes, and *linear* in the number of landmarks, M_f, observed exclusively from the M_r non-keyframes, where $M_r \ll M_f$.
- The keyframes and the associated landmark-map are maintained over the entire robot trajectory, and thus C-KLAM enables efficient loop closures, necessary for ensuring accurate and consistent long-term navigation.

II. ALGORITHM DESCRIPTION

In this section, we first present a brief overview of batch MAP-based SLAM, followed by the details of the proposed C-KLAM algorithm. Moreover, to facilitate the description of these estimation algorithms, we will use the specific example scenario depicted in Fig. 1. Note, however, that C-KLAM is a general approach that can be used for any number of key and non-key poses[1] and landmarks.

A. Batch MAP-based SLAM

Consider a robot, equipped with proprioceptive (e.g., IMU) and exteroceptive (e.g., camera) sensors, navigating in a 3D environment. The motion model for the robot is given by:

$$\mathbf{x}_{i+1} = \mathbf{f}(\mathbf{x}_i, \mathbf{u}_i - \mathbf{w}_i) \tag{1}$$

where $\mathbf{f}$ is a general nonlinear function[2], $\mathbf{x}_i$ and $\mathbf{x}_{i+1}$ denote the robot poses at time-steps i and $i+1$, respectively, $\mathbf{u}_i = \mathbf{u}_{i_t} + \mathbf{w}_i$, is the measured control input (linear acceleration and rotational velocity), where $\mathbf{u}_{i_t}$ denotes the true control input, and $\mathbf{w}_i$ is the zero-mean, white Gaussian measurement noise with covariance $\mathbf{Q}_i$. The measurement model for the robot at time-step i, obtaining an observation, $\mathbf{z}_{ij}$, to landmark $\mathbf{f}_j$ is given by:

$$\mathbf{z}_{ij} = \mathbf{h}(\mathbf{x}_i, \mathbf{f}_j) + \mathbf{v}_{ij} \tag{2}$$

where $\mathbf{h}$ is a general nonlinear measurement function[2] and $\mathbf{v}_{ij}$ is the zero-mean, white Gaussian measurement noise with covariance $\mathbf{R}_{ij}$.

[1]The terms key poses and keyframes are used interchangeably in this paper.

[2]The details of the IMU motion model as well as the camera measurement model can be found in [8].

Consider the current exploration epoch shown in Fig. 1, consisting of five robot poses, $\mathbf{x}_i$, $i = 0, 1, \ldots, 4$, and of five point landmarks, $\mathbf{f}_j$, $j = 1, 2, \ldots, 5$, observed from these poses. The batch MAP estimates, $\hat{\mathbf{x}}_{0:4}^{MAP}, \hat{\mathbf{f}}_{1:5}^{MAP}$, of *all* robot poses, $\mathbf{x}_{0:4}$, and *all* landmark positions, $\mathbf{f}_{1:5}$, using *all* available proprioceptive, $\mathbf{u}_{0:3}$, and exteroceptive, $\mathcal{Z}_{0:4}$, measurements are given by:

$$\hat{\mathbf{x}}_{0:4}^{MAP}, \hat{\mathbf{f}}_{1:5}^{MAP} \triangleq \arg\max_{\mathbf{x}_{0:4}, \mathbf{f}_{1:5}} p(\mathbf{x}_{0:4}, \mathbf{f}_{1:5} | \mathcal{Z}_{0:4}, \mathbf{u}_{0:3}) \tag{3}$$

where $\mathcal{Z}_i$ denotes the set of all exteroceptive measurements obtained at robot pose $\mathbf{x}_i$, $i = 0, 1, \ldots, 4$. Under the Gaussian and independent noise assumptions, (3) is equivalent to minimizing the following nonlinear least-squares cost function:

$$\begin{aligned}
&\mathcal{C}(\mathbf{x}_{0:4}, \mathbf{f}_{1:5}; \mathcal{Z}_{0:4}, \mathbf{u}_{0:3}) \\
&= \frac{1}{2}||\mathbf{x}_0 - \hat{\mathbf{x}}_{0|0}||^2_{\mathbf{P}_{0|0}} + \sum_{i=0}^{3} \frac{1}{2}||\mathbf{x}_{i+1} - \mathbf{f}(\mathbf{x}_i, \mathbf{u}_i)||^2_{\mathbf{Q}'_i} \\
&\quad + \sum_{\mathbf{z}_{ij} \in \mathcal{Z}_{0:4}} \frac{1}{2}||\mathbf{z}_{ij} - \mathbf{h}(\mathbf{x}_i, \mathbf{f}_j)||^2_{\mathbf{R}_{ij}} \\
&\triangleq \mathcal{C}_P(\mathbf{x}_0; \hat{\mathbf{x}}_{0|0}) + \sum_{i=0}^{3} \mathcal{C}_M(\mathbf{x}_{i+1}, \mathbf{x}_i; \mathbf{u}_i) \\
&\quad + \sum_{\mathbf{z}_{ij} \in \mathcal{Z}_{0:4}} \mathcal{C}_O(\mathbf{x}_i, \mathbf{f}_j; \mathbf{z}_{ij})
\end{aligned} \tag{4}$$

where $\mathbf{x}_0 \sim \mathcal{N}(\hat{\mathbf{x}}_{0|0}, \mathbf{P}_{0|0})$ denotes the prior for the robot pose, $\mathbf{Q}'_i = \mathbf{G}_i \mathbf{Q}_i \mathbf{G}_i^T$, and $\mathbf{G}_i$ is the Jacobian of $\mathbf{f}$ with respect to the noise $\mathbf{w}_i$. In what follows, we denote the cost terms arising from the prior, the robot motion, and the landmark observations by $\mathcal{C}_P$, $\mathcal{C}_M$, and $\mathcal{C}_O$, respectively.

A standard approach for minimizing (4) is to employ the Gauss-Newton iterative minimization algorithm [17] with computational complexity up to $O([K+N]^3)$, where K and N denote the number of robot poses and landmarks, respectively. Note that, as the robot explores the environment and observes new landmarks, the size of the optimization problem (both K and N) in (4) continuously increases. Therefore, for long trajectories with many features and frequent loop closures, the cost of solving (4) may prohibit real-time operation.

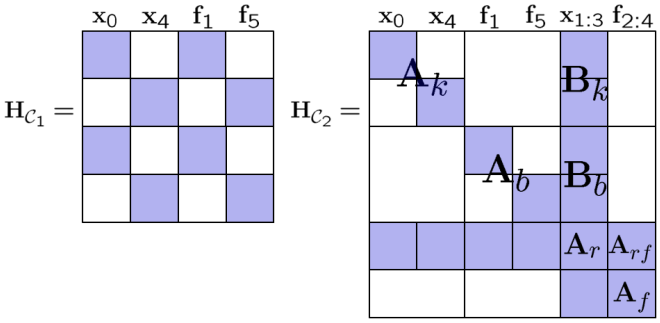

Fig. 2: Structure of the Hessian matrices, $\mathbf{H}_{\mathcal{C}_1}$ and $\mathbf{H}_{\mathcal{C}_2}$, corresponding to the cost functions $\mathcal{C}_1$ and $\mathcal{C}_2$ [see (5)], respectively. The colored blocks denote non-zero elements. Specifically, for $\mathbf{H}_{\mathcal{C}_2}$, associated with the measurements denoted by red arrows in Fig. 1, the block-diagonal sub-matrices $\mathbf{A}_k$ and $\mathbf{A}_b$ correspond to key poses and key landmarks, respectively. $\mathbf{A}_r$ and $\mathbf{A}_f$ correspond to non-key poses and non-key landmarks to be marginalized, respectively. Here $\mathbf{A}_k$ and $\mathbf{A}_r$ are, in general, block tri-diagonal, while $\mathbf{A}_b$ and $\mathbf{A}_f$ are block diagonal.

B. C-KLAM Algorithm

1) Problem Formulation: In order to reduce the computational complexity of MAP-based SLAM and ensure accurate and real-time navigation over long time durations, the proposed C-KLAM approach (i) builds a sparse map of the environment consisting of *only* the key robot poses and the distinctive landmarks observed from these key poses, and (ii) uses measurement information from non-key poses to create constraints between the key poses, in order to improve estimation accuracy.

Specifically, for the example in Fig. 1, let us assume that we retain: (i) $\mathbf{x}_0$ and $\mathbf{x}_4$ as key poses, and (ii) landmarks, $\mathbf{f}_1$ and $\mathbf{f}_5$, observed from these key poses as key landmarks[3]. In this case, (4) can be split into two parts as follows:

$$\mathcal{C} = \underbrace{\mathcal{C}_P(\mathbf{x}_0;\hat{\mathbf{x}}_{0|0}) + \mathcal{C}_O(\mathbf{x}_0,\mathbf{f}_1;\mathbf{z}_{01}) + \mathcal{C}_O(\mathbf{x}_4,\mathbf{f}_5;\mathbf{z}_{45})}_{\mathcal{C}_1(\mathbf{x}_0,\mathbf{x}_4,\mathbf{f}_1,\mathbf{f}_5;\hat{\mathbf{x}}_{0|0},\mathbf{z}_{01},\mathbf{z}_{45})} + \underbrace{\sum_{i=0}^{3}\mathcal{C}_M(\mathbf{x}_{i+1},\mathbf{x}_i;\mathbf{u}_i) + \sum_{\mathbf{z}_{ij}\in\mathcal{Z}_{1:3}}\mathcal{C}_O(\mathbf{x}_i,\mathbf{f}_j;\mathbf{z}_{ij})}_{\mathcal{C}_2(\mathbf{x}_{1:3},\mathbf{f}_{2:4},\mathbf{x}_0,\mathbf{x}_4,\mathbf{f}_1,\mathbf{f}_5;\mathcal{Z}_{1:3},\mathbf{u}_{0:3})} \quad (5)$$

The first part of the cost function, $\mathcal{C}_1$, depends only upon the key poses, key landmarks, and the measurements between them (denoted by thin black arrows in Fig. 1). This part consists of cost terms arising from the prior term and from the two exteroceptive measurements, $\mathbf{z}_{01}$ and $\mathbf{z}_{45}$, obtained at the key poses $\mathbf{x}_0$ and $\mathbf{x}_4$, respectively. The second part of the cost function, $\mathcal{C}_2$, contains all cost terms that involve non-key poses and non-key landmarks. Specifically, these correspond to two types of cost terms: (i) terms that involve *only* non-key poses and non-key landmarks (corresponding to measurements denoted by solid red lines in Fig. 1), e.g., $\mathcal{C}_O(\mathbf{x}_1,\mathbf{f}_2;\mathbf{z}_{12})$, and (ii) terms that involve *both* key and non-key elements (corresponding to measurements denoted by dashed red lines in Fig. 1), e.g., $\mathcal{C}_O(\mathbf{x}_1,\mathbf{f}_1;\mathbf{z}_{11})$ and $\mathcal{C}_M(\mathbf{x}_1,\mathbf{x}_0;\mathbf{u}_0)$.

2) Marginalization and Naïve Approximation: Before we proceed, we note that some keyframe-based approaches such as PTAM [10] optimize only over $\mathcal{C}_1$ in order to reduce the computational complexity, i.e., the cost terms in $\mathcal{C}_2$ and the corresponding measurements are discarded, resulting in significant information loss. An alternative approach to retain a part of the information in $\mathcal{C}_2$, is to *marginalize* the non-key poses and landmarks, $\mathbf{x}_{1:3}$ and $\mathbf{f}_{2:4}$, respectively. Mathematically, this is equivalent to approximating the cost function $\mathcal{C}$ by $\mathcal{C}'$ as follows (see Fig. 2):

$$\begin{aligned}\mathcal{C} &\simeq \mathcal{C}'(\mathbf{x}_0,\mathbf{x}_4,\mathbf{f}_1,\mathbf{f}_5;\hat{\mathbf{x}}_{0|0},\mathbf{z}_{01},\mathbf{z}_{45},\hat{\mathbf{x}}_0,\hat{\mathbf{x}}_4,\hat{\mathbf{f}}_1,\hat{\mathbf{f}}_5)\\ &= \mathcal{C}_1 + \mathcal{C}'_2(\mathbf{x}_0,\mathbf{x}_4,\mathbf{f}_1,\mathbf{f}_5;\hat{\mathbf{x}}_0,\hat{\mathbf{x}}_4,\hat{\mathbf{f}}_1,\hat{\mathbf{f}}_5) \quad (6)\end{aligned}$$

where,

$$\mathcal{C}'_2 = \alpha' + \mathbf{g}_{\mathcal{C}'_2}^T \begin{bmatrix}\mathbf{x}_0-\hat{\mathbf{x}}_0\\ \mathbf{x}_4-\hat{\mathbf{x}}_4\\ \mathbf{f}_1-\hat{\mathbf{f}}_1\\ \mathbf{f}_5-\hat{\mathbf{f}}_5\end{bmatrix} + \frac{1}{2}\begin{bmatrix}\mathbf{x}_0-\hat{\mathbf{x}}_0\\ \mathbf{x}_4-\hat{\mathbf{x}}_4\\ \mathbf{f}_1-\hat{\mathbf{f}}_1\\ \mathbf{f}_5-\hat{\mathbf{f}}_5\end{bmatrix}^T \mathbf{H}_{\mathcal{C}'_2}\begin{bmatrix}\mathbf{x}_0-\hat{\mathbf{x}}_0\\ \mathbf{x}_4-\hat{\mathbf{x}}_4\\ \mathbf{f}_1-\hat{\mathbf{f}}_1\\ \mathbf{f}_5-\hat{\mathbf{f}}_5\end{bmatrix} \quad (7)$$

with,

$$\mathbf{H}_{\mathcal{C}'_2} = \begin{bmatrix}\mathbf{A}_k & \mathbf{0}\\ \mathbf{0} & \mathbf{A}_b\end{bmatrix} - \begin{bmatrix}\mathbf{B}_k & \mathbf{0}\\ \mathbf{B}_b & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{A}_r & \mathbf{A}_{rf}\\ \mathbf{A}_{fr} & \mathbf{A}_f\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{B}_k^T & \mathbf{B}_b^T\\ \mathbf{0} & \mathbf{0}\end{bmatrix} \quad (8)$$

$$\mathbf{g}_{\mathcal{C}'_2} = \begin{bmatrix}\mathbf{g}_k\\ \mathbf{g}_b\end{bmatrix} - \begin{bmatrix}\mathbf{B}_k & \mathbf{0}\\ \mathbf{B}_b & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{A}_r & \mathbf{A}_{rf}\\ \mathbf{A}_{fr} & \mathbf{A}_f\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{g}_r\\ \mathbf{g}_f\end{bmatrix} \triangleq \begin{bmatrix}\mathbf{g}_{\mathcal{C}'_2,k}\\ \mathbf{g}_{\mathcal{C}'_2,b}\end{bmatrix}. \quad (9)$$

Here, $\hat{\mathbf{x}}_0$, $\hat{\mathbf{x}}_4$, $\hat{\mathbf{f}}_1$, and $\hat{\mathbf{f}}_5$ are the estimates of $\mathbf{x}_0$, $\mathbf{x}_4$, $\mathbf{f}_1$, and $\mathbf{f}_5$, respectively, at the time of marginalization, α' is a constant term independent of the optimization variables, and $\mathbf{g}_k, \mathbf{g}_b, \mathbf{g}_r$, and $\mathbf{g}_f$ are the gradient vectors of $\mathcal{C}_2$ with respect to $\{\mathbf{x}_0,\mathbf{x}_4\}$, $\{\mathbf{f}_1,\mathbf{f}_5\}$, $\{\mathbf{x}_{1:3}\}$, and $\{\mathbf{f}_{2:4}\}$, respectively. Also, $\mathbf{g}_{\mathcal{C}'_2}$ and $\mathbf{H}_{\mathcal{C}'_2}$ denote the Jacobian and Hessian matrix, respectively. Lastly, we note that $\mathbf{H}_{\mathcal{C}'_2}$, as expected, is the Schur complement of the diagonal block, corresponding to non-key poses and non-key landmarks, of the Hessian, $\mathbf{H}_{\mathcal{C}_2}$, of the original cost function, $\mathcal{C}_2$ (see Fig. 2).

As expected, however, this marginalization of non-key elements creates additional constraints between the key poses and the key landmarks, which directly translates into fill-ins in the reduced Hessian matrix, $\mathbf{H}_{\mathcal{C}'_2}$. This destroys the sparse structure of the Hessian matrix, $\mathbf{H}_{\mathcal{C}'} = \mathbf{H}_{\mathcal{C}_1} + \mathbf{H}_{\mathcal{C}'_2}$, that corresponds to the cost function $\mathcal{C}'$ [see (6)], and substantially increases the computational cost of obtaining a solution to the minimization problem. By studying the relationship between the measurement graph corresponding to Fig. 1 and the sparsity pattern of the resulting Hessian matrix, we note that the exteroceptive measurements from

[3]Note that we retain only two key poses/landmarks in this example, in order to simplify the explanation. However, C-KLAM can be used to retain any number of key poses/landmarks. The key poses are selected based on certain criteria, e.g., distance traveled between two key poses, poses that observe points of interest, uniqueness of a image, etc. Furthermore, for the example in Fig. 1, we assume that the depth to the features is available (e.g., from an RGB-D camera), in order to reduce the number of measurements and poses required. However, if a regular camera is used, at least two observations of a key feature and the corresponding poses will need to be retained.

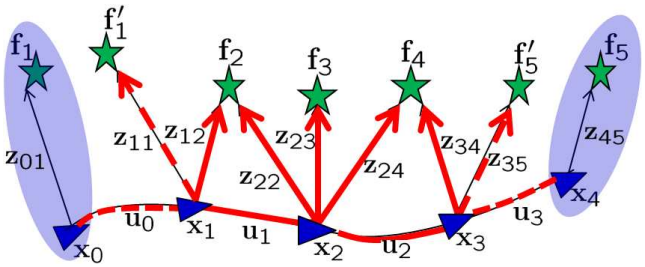

Fig. 3: Pictorial depiction of the approximation carried out by C-KLAM in order to ensure sparsity of the Hessian matrix. Instead of associating the measurements $\mathbf{z}_{11}$ and $\mathbf{z}_{35}$, to the key features $\mathbf{f}_1$ and $\mathbf{f}_5$ (see Fig. 1), respectively, C-KLAM assumes that these are measurements to different landmarks $\mathbf{f}_1'$ and $\mathbf{f}_5'$.

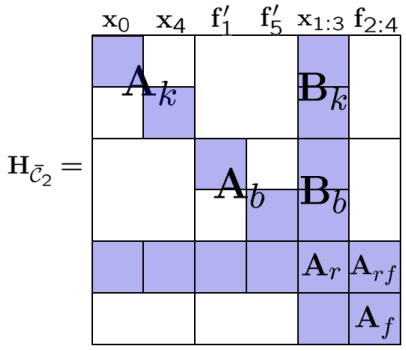

Fig. 4: Structure of the Hessian matrix, $\mathbf{H}_{\bar{\mathcal{C}}_2}$, corresponding to the cost function $\bar{\mathcal{C}}_2$ [see (11)]. The colored blocks denote non-zero elements. Note that this Hessian matrix does not have any entries corresponding to the key features $\mathbf{f}_1$ and $\mathbf{f}_5$. Instead, it has entries for the features $\mathbf{f}_1'$ and $\mathbf{f}_5'$.

non-key poses to key features, i.e., $\mathbf{z}_{11}$ and $\mathbf{z}_{35}$, are the ones responsible for generating fill-ins in the Hessian matrix, $\mathbf{H}_{\mathcal{C}'}$, after marginalization[4].

A straightforward solution to retain the sparsity of the Hessian matrix would be to first discard any exteroceptive measurements between non-key poses and key features (e.g., $\mathbf{z}_{11}$ and $\mathbf{z}_{35}$ in Fig. 1), and then proceed with the marginalization of non-key elements. However, in real-world scenarios, $\mathbf{f}_1$ and $\mathbf{f}_5$ are not single features, but they each correspond to a group of features. Hence, such an approximation would discard numerous measurements, resulting in substantial information loss.

3) C-KLAM Approximation: In order to address this problem and maintain the sparse structure of the Hessian (information) matrix while incorporating information from $\mathcal{C}_2$, C-KLAM carries out an additional approximation step, i.e., it further approximates $\mathcal{C}_2'$ in (6) by a quadratic cost term, $\mathcal{C}_2''(\mathbf{x}_0, \mathbf{x}_4; \hat{\mathbf{x}}_0, \hat{\mathbf{x}}_4)$ that constraints *only* the key poses $\mathbf{x}_0$ and $\mathbf{x}_4$.

Specifically, along with the non-key poses/landmarks, C-KLAM *marginalizes the key landmarks* $\mathbf{f}_1$ *and* $\mathbf{f}_5$, but *only* from $\mathcal{C}_2$; these key landmarks will still appear as optimization variables in $\mathcal{C}_1$ [see (5)]. Moreover, marginalizing $\mathbf{f}_1$ and $\mathbf{f}_5$ from $\mathcal{C}_2$, while retaining them in $\mathcal{C}_1$, implies that we ignore their data association[5] and treat them as different features (say $\mathbf{f}_1'$ and $\mathbf{f}_5'$) in $\mathcal{C}_2$. Mathematically, this process can be described by first considering the following equivalent optimization problems [see (4), (5), and Fig. 3]:

$$\begin{aligned} &\min \mathcal{C}(\mathbf{x}_{0:4}, \mathbf{f}_{1:5}; \mathcal{Z}_{0:4}, \mathbf{u}_{0:3}) \\ &\Leftrightarrow \min \bar{\mathcal{C}}(\mathbf{x}_{0:4}, \mathbf{f}_{1:5}, \mathbf{f}_1', \mathbf{f}_5'; \mathcal{Z}_{0:4}, \mathbf{u}_{0:3}) \\ &\quad \text{s.t. } \mathbf{f}_1 = \mathbf{f}_1', \mathbf{f}_5 = \mathbf{f}_5' \end{aligned} \tag{10}$$

where,

$$\begin{aligned} \bar{\mathcal{C}} = &\mathcal{C}_1(\mathbf{x}_0, \mathbf{x}_4, \mathbf{f}_1, \mathbf{f}_5; \hat{\mathbf{x}}_{0|0}, \mathbf{z}_{01}, \mathbf{z}_{45}) \\ &+ \bar{\mathcal{C}}_2(\mathbf{x}_{1:3}, \mathbf{f}_{2:4}, \mathbf{x}_0, \mathbf{x}_4, \mathbf{f}_1', \mathbf{f}_5'; \mathcal{Z}_{1:3}, \mathbf{u}_{0:3}) \end{aligned} \tag{11}$$

[4]Note that the proprioceptive measurements between key and non-key poses, i.e., $\mathbf{u}_0$ and $\mathbf{u}_3$, also generate fill-ins, but these fill-ins are desirable for our problem as they represent constraints between two consecutive key poses after marginalization.

[5]Besides the inability to relinearize marginalized states, ignoring this data association is the main information loss incurred by C-KLAM as compared to the batch MAP-based SLAM.

Note that minimizing the batch-MAP cost function in (4) is *exactly* equivalent to the constrained optimization problem presented in (10). Now, in order to maintain the sparsity of the Hessian matrix after marginalizing the non-key elements, C-KLAM *discards* the constraint in (10) and hence assumes that the features $\mathbf{f}_1'$ and $\mathbf{f}_5'$ are distinct from $\mathbf{f}_1$ and $\mathbf{f}_5$, respectively (see Fig. 3). Due to this relaxation, $\bar{\mathcal{C}}_2$ no longer depends on the key features $\mathbf{f}_1$ and $\mathbf{f}_5$, and hence has *no* cost terms corresponding to measurements between non-key poses and key features. Due to this approximation, C-KLAM can now marginalize the features $\mathbf{f}_1'$ and $\mathbf{f}_5'$, along with the non-key elements $\mathbf{x}_{1:3}$ and $\mathbf{f}_{2:4}$, from $\bar{\mathcal{C}}$ in (11), thus ensuring that the resulting Hessian matrix remains sparse. Specifically, C-KLAM approximates $\bar{\mathcal{C}}_2$ in (11) by [see Fig.s 2 and 4]:

$$\begin{aligned} \bar{\mathcal{C}}_2 &\simeq \mathcal{C}_2''(\mathbf{x}_0, \mathbf{x}_4; \hat{\mathbf{x}}_0, \hat{\mathbf{x}}_4) \\ &= \alpha'' + \mathbf{g}_{\mathcal{C}_2''}^T \begin{bmatrix} \mathbf{x}_0 - \hat{\mathbf{x}}_0 \\ \mathbf{x}_4 - \hat{\mathbf{x}}_4 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} \mathbf{x}_0 - \hat{\mathbf{x}}_0 \\ \mathbf{x}_4 - \hat{\mathbf{x}}_4 \end{bmatrix}^T \mathbf{H}_{\mathcal{C}_2''} \begin{bmatrix} \mathbf{x}_0 - \hat{\mathbf{x}}_0 \\ \mathbf{x}_4 - \hat{\mathbf{x}}_4 \end{bmatrix} \end{aligned} \tag{12}$$

with,

$$\mathbf{H}_{\mathcal{C}_2''} = \mathbf{A}_k - \mathbf{B}_k(\mathbf{D} - \mathbf{B}_b^T \mathbf{A}_b^{-1} \mathbf{B}_b)^{-1} \mathbf{B}_k^T \tag{13}$$

$$\begin{aligned} \mathbf{g}_{\mathcal{C}_2''} = &\mathbf{g}_{\mathcal{C}_2', k} + \mathbf{B}_k \mathbf{D}^{-1} \mathbf{B}_b^T \\ &\cdot (\mathbf{A}_b^{-1} + \mathbf{A}_b^{-1} \mathbf{B}_b (\mathbf{D} - \mathbf{B}_b^T \mathbf{A}_b^{-1} \mathbf{B}_b)^{-1} \mathbf{B}_b^T \mathbf{A}_b^{-1}) \mathbf{g}_{\mathcal{C}_2', b} \end{aligned} \tag{14}$$

and

$$\mathbf{D} = \mathbf{A}_r - \mathbf{A}_{rf} \mathbf{A}_f^{-1} \mathbf{A}_{fr}. \tag{15}$$

where α'' is a constant, independent of the optimization variables, and $\mathbf{g}_{\mathcal{C}_2''}$, $\mathbf{H}_{\mathcal{C}_2''}$ denote the Jacobian and Hessian matrix, respectively.

After this approximation, the final C-KLAM cost function becomes:

$$\begin{aligned} \mathcal{C}_{CKLAM} = &\mathcal{C}_1(\mathbf{x}_0, \mathbf{x}_4, \mathbf{f}_1, \mathbf{f}_5; \hat{\mathbf{x}}_{0|0}, \mathbf{z}_{01}, \mathbf{z}_{45}) \\ &+ \mathcal{C}_2''(\mathbf{x}_0, \mathbf{x}_4; \hat{\mathbf{x}}_0, \hat{\mathbf{x}}_4) \end{aligned} \tag{16}$$

whose corresponding Hessian would be the same as that of $\mathcal{C}_1$ (and thus sparse) plus an additional information (relative pose) constraint between $\mathbf{x}_0$ and $\mathbf{x}_4$ due to $\mathcal{C}_2''$. In summary, by approximating $\mathcal{C}_2$ by $\mathcal{C}_2''$, C-KLAM is able to incorporate most of the information from the non-key poses/landmarks,

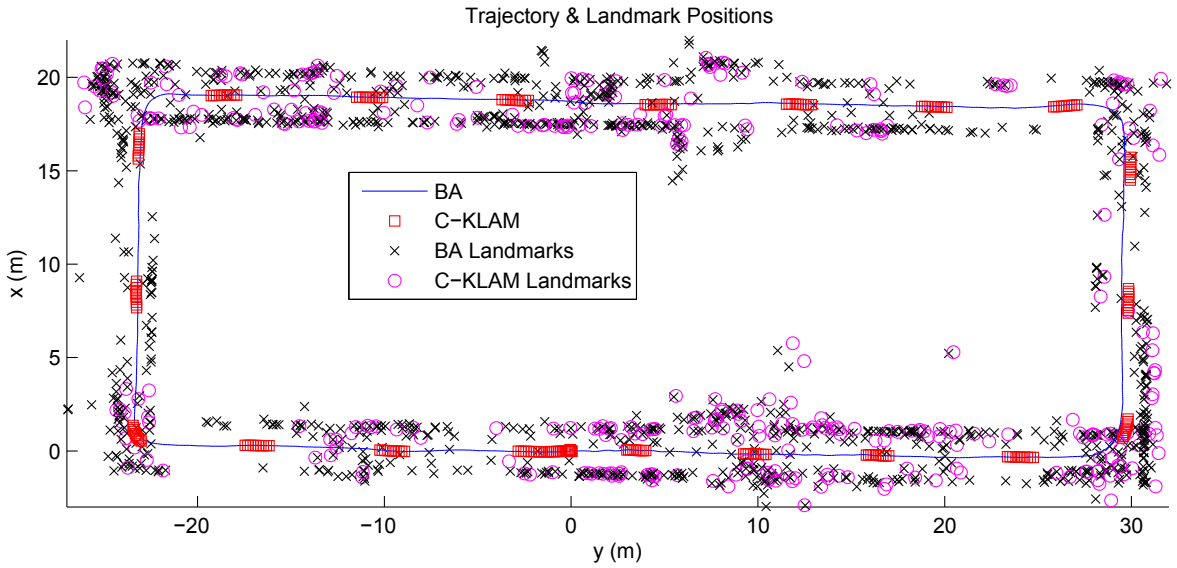

Fig. 5: Overhead $x-y$ view of the estimated 3D trajectory and landmark positions. The C-KLAM estimates only keyframes (marked with red squares) and key features (marked with magenta circles), while BA estimates the entire trajectory (marked by blue line) and all features (marked by black x-s).

while maintaining the sparsity of the Hessian matrix. Moreover, the part of the cost function, $\mathcal{C}_1$, corresponding to the key poses/landmarks, remains intact.

4) C-KLAM Computational Complexity: Lastly, we show that the approximation (marginalization) described above can be carried out with cost cubic in the number of marginalized non-key poses, and only linear in the number of marginalized non-key landmarks. For the complexity analysis, let us assume that we have M_r non-key poses and M_f non-key features to be marginalized, and M_b features that are observed from both key and non-key frames, where $M_f \gg M_r$ and $M_f \gg M_b$. The marginalization step involves the computation of the Hessian matrix, $\mathbf{H}_{\mathcal{C}_2''}$, and the Jacobian, $\mathbf{g}_{\mathcal{C}_2''}$, according to (13) - (15). For computing both the Hessian and the Jacobian, we first need to calculate $\mathbf{D}$ in (15). Since $\mathbf{A}_f$ is block-diagonal, $\mathbf{A}_f^{-1}$ in (15) can be computed with cost only $\mathcal{O}(M_f)$. Moreover, since the number of marginalized non-key features, M_f, far exceeds M_r and M_b, the cost of computing $\mathbf{D}$ remains $\mathcal{O}(M_f)$. To compute the Hessian [see (13)], note that $\mathbf{A}_b$ is also block-diagonal, hence obtaining $(\mathbf{D}-\mathbf{B}_b^T\mathbf{A}_b^{-1}\mathbf{B}_b)^{-1}$, which is the most computationally-intensive operation in (13), requires $\mathcal{O}(M_r^3)$ operations. The cost of calculating the remaining matrix multiplications and additions in (13) is significantly lower as compared to this cubic cost.

To compute the Jacobian, $\mathbf{g}_{\mathcal{C}_2''}$ [see (14)], we can reuse the values of $\mathbf{D}$, $(\mathbf{D}-\mathbf{B}_b^T\mathbf{A}_b^{-1}\mathbf{B}_b)^{-1}$, and $\mathbf{A}_b^{-1}$, which have already been calculated when computing the Hessian. In addition, we need to compute $\mathbf{D}^{-1}$, which can be found with complexity $\mathcal{O}(M_r^3)$. The rest of the computations involve only matrix-vector multiplications and vector additions at a negligible cost.

Hence, the overall cost of the marginalization step is cubic in the number of marginalized non-key poses, and only linear in the number of marginalized non-key landmarks. Since M_r is bounded (user defined), the marginalization in C-KLAM can be carried out with minimal computational overhead.

III. EXPERIMENTAL RESULTS

The experimental setup consists of a PointGrey Chameleon camera and a Navchip IMU, rigidly attached on a light-weight (100 g) platform. The IMU signals were sampled at a frequency of 100 Hz while camera images were acquired at 7.5 Hz. SIFT features [18] were detected in the camera images and matched using a vocabulary tree [19]. The experiment was conducted in an indoor environment where the sensor platform followed a 3D rectangular trajectory, of total length of 144 m, and returned back to the initial position in order to provide an estimate of the final position error.

In the C-KLAM implementation, the corresponding approximate batch-MAP optimization problem was solved every 20 incoming camera frames. The exploration epoch (see Fig. 1) was set to 60 camera frames, from which the first and last 10 consecutive camera frames were retained as keyframes, while the rest were marginalized using the C-KLAM algorithm. We compared the performance of C-KLAM to that of the computationally-intensive, batch MAP-based SLAM [bundle adjustment (BA)], which optimizes over all camera poses and landmarks, using all available measurements, to provide high-accuracy estimates as the comparison baseline. In the BA implementation, the batch-MAP optimization problem was solved every 20 incoming camera frames.

Fig. 5 shows the $x-y$ view of the estimated trajectory and landmark positions. As evident, the estimates of the robot trajectory and landmark positions generated by C-KLAM are almost identical to those of the BA. Loop closure was performed and the final position error was 7 cm for C-KLAM, only 5% more than that of the BA.

In terms of speed, C-KLAM took only 4% of the time required for the entire BA. At the end of this experiment, C-KLAM retained 238 keyframes and 349 key landmarks, while BA had 1038 camera frames and 1281 landmarks. This significant reduction in the number of estimated states in C-KLAM led to substantial improvement in efficiency. Moreover, by using information from non-keyframes to constrain the keyframes, C-KLAM was able to achieve estimation performance comparable to that of the BA. See the accompanying video for results of a more challenging experiment with a flying quadrotor.

IV. CONCLUSION

In this paper, we presented C-KLAM, an approximate MAP estimator-based SLAM algorithm. In order to reduce the computational complexity of the batch MAP-based SLAM, C-KLAM estimates only the keyframes and key landmarks, observed from these keyframes. However, instead of discarding the measurement information from non-keyframes and non-key landmarks, C-KLAM uses most of this information to generate consistent pose constraints between the keyframes, resulting in substantial information gain. Moreover, the approximations performed in C-KLAM retain the sparsity of the information matrix, and hence the resulting optimization problem can be solved efficiently. We presented experimental results for validating the performance of C-KLAM and compared it with that of the batch MAP-based SLAM (bundle adjustment). Our results demonstrated that C-KLAM not only obtains substantial speed-up, but also achieves estimation accuracy comparable to that of the batch MAP-based SLAM that uses all available measurement information.

REFERENCES

[1] R. Smith and P. Cheeseman, "On the representation and estimation of spatial uncertainty," *International Journal of Robotics Research*, vol. 5, no. 4, pp. 56–68, Dec. 1986.

[2] F. Dellaert and M. Kaess, "Square root SAM: Simultaneous Localization and Mapping via square root information smoothing," *International Journal of Robotics Research*, vol. 25, no. 12, pp. 1181–1203, Dec. 2006.

[3] R. Kummerle, G. Grisetti, H. Strasdat, K. Konolige, and W. Burgard, "g2o: A general framework for graph optimization," in *Proc. of the IEEE International Conference on Robotics and Automation*, Shanghai, China, May 9–13 2011, pp. 3607–3613.

[4] K. Konolige, G. Grisetti, R. Kummerle, W. Burgard, B. Limketkai, and R. Vincent, "Efficient Sparse Pose Adjustment for 2D mapping," in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, Taipei, Taiwan, Oct. 18–22 2010, pp. 22–29.

[5] M. Kaess, A. Ranganathan, and F. Dellaert, "iSAM: Incremental smoothing and mapping," *IEEE Transactions on Robotics*, vol. 24, no. 6, pp. 1365–1378, Dec. 2008.

[6] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the bayes tree," *International Journal of Robotics Research*, vol. 31, no. 2, pp. 216–235, Feb. 2012.

[7] J. Folkesson and H. Christensen, "Graphical slam - a self-correcting map," in *Proc. of the IEEE International Conference on Robotics and Automation*, New Orleans, LA, Apr. 26 – May 1, 2004, pp. 383–390.

[8] A. I. Mourikis, N. Trawny, S. I. Roumeliotis, A. Johnson, A. Ansar, and L. Matthies, "Vision-aided inertial navigation for spacecraft entry, descent, and landing," *IEEE Transactions on Robotics*, vol. 25, no. 2, pp. 264–280, Apr. 2009.

[9] G. Sibley, L. Matthies, and G. Sukhatme, "Sliding window filter with application to planetary landing," *Journal of Field Robotics*, vol. 27, no. 5, pp. 587–608, Sep./Oct. 2010.

[10] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," in *Proc. of the IEEE and ACM International Symposium on Mixed and Augmented Reality*, Nara, Japan, Nov. 13–16 2007, pp. 225–234.

[11] K. Konolige and M. Agrawal, "FrameSLAM: From bundle adjustment to real-time visual mapping," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 1066–1077, Oct. 2008.

[12] H. Strasdat, J. Montiel, and A. Davison, "Scale drift-aware large scale monocular slam," in *Proc. of Robotics: Science and Systems*, Zaragoza, Spain, Jun. 27–30 2010.

[13] H. Strasdat, A. Davison, J. Montiel, and K. Konolige, "Double window optimisation for constant time visual slam," in *Proc. of the IEEE International Conference on Computer Vision*, Barcelona, Spain, Nov. 6–13 2011, pp. 2352–2359.

[14] K. Konolige, J. Bowman, J. D. Chen, P. Mihelich, M. Calonder, V. Lepetit, and P. Fua, "View-based maps," *International Journal of Robotics Research*, vol. 29, no. 29, pp. 941–957, Jul. 2010.

[15] R. M. Eustice, H. Singh, and J. J. Leonard, "Exactly sparse delayed-state filters for view-based SLAM," *IEEE Transactions on Robotics*, vol. 22, no. 6, pp. 1100–1114, Dec. 2006.

[16] H. Johannsson, M. Kaess, M. Fallon, and J. Leonard, "Temporally scalable visual SLAM using a reduced pose graph," in *Proc. of the IEEE International Conference on Robotics and Automation*, Karlsruhe, Germany, May 6–10 2013, pp. 54–61.

[17] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle Adjustment - A Modern Synthesis," *Lecture Notes in Computer Science*, vol. 1883, pp. 298–372, Jan. 2000.

[18] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[19] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, NY, Jun. 17–22 2006, pp. 2161–2168.