Exactly Sparse Delayed-State Filters

Ryan M. Eustice* and Hanumant Singh†
* MIT/WHOI Joint Program in Applied Ocean Science and Engineering
† Department of Applied Ocean Physics and Engineering
Woods Hole Oceanographic Institution
Woods Hole, MA, USA
{ryan,hanu}@whoi.edu

John J. Leonard
Department of Ocean Engineering
Massachusetts Institute of Technology
Cambridge, MA, USA
jleonard@mit.edu

Abstract— **This paper presents the novel insight that the SLAM information matrix is *exactly* sparse in a delayed-state framework. Such a framework is used in view-based representations of the environment which rely upon scan-matching raw sensor data. Scan-matching raw data results in virtual observations of robot motion with respect to a place its previously been. The exact sparseness of the delayed-state information matrix is in contrast to other recent feature-based SLAM information algorithms like Sparse Extended Information Filters or Thin Junction Tree Filters. These methods have to make approximations in order to force the feature-based SLAM information matrix to be sparse. The benefit of the exact sparseness of the delayed-state framework is that it allows one to take advantage of the information space parameterization without having to make any approximations. Therefore, it can produce equivalent results to the "full-covariance" solution.**

Index Terms— **Delayed states, EIF, SLAM.**

I. Introduction

Good navigation is often a prerequisite for many of the tasks assigned to mobile robots. For example, much of the deep-water science over the past two decades has relied upon Remotely Operated Vehicles (ROVs) to be the hands and eyes of humans. Scientists who use these vehicles often want to be able to co-locate the data they collect spatially and temporally such as in: studies of bio-diversity [1], coral reef health [2], and archeology [3]. Since GPS signals do not penetrate the ocean surface engineers mainly rely upon acoustic-beacon networks to obtain bounded-error triangulated vehicle positions for large-area navigation [4]. The disadvantage of this method is that it requires the deployment, calibration, and recovery of the transponder net. This is often an acceptable trade-off for long-term deployments, but frequently it is the bane of short-term surveys. In more recent years underwater vehicles have seen advances in their dead-reckoning abilities. The advent of sensors such as the acoustic Doppler velocity log [5] and north seeking ring laser gyro [6] have improved the ability of near-seafloor vehicles to navigate with reported error bounds of less than 1% of distance traveled [7]. However, in the absence of an external reference, error continues to grow unbounded as a monotonic function of time.

Eustice, Pizarro and Singh [8] presented a simultaneous localization and mapping (SLAM) technique for near sea-floor navigation called Visually Augmented Navigation (VAN). This technique incorporates pairwise camera constraints from low-overlap imagery to constrain the vehicle position estimate and "reset" the accumulated navigation drift error. In this framework, the camera provides measurements of the 6 DOF relative coordinate transformation between poses modulo scale. The method recursively incorporates these relative pose constraints by estimating the global poses which are consistent with the camera measurements and navigation prior. These global poses correspond to samples from the robot's trajectory acquired at image acquisition and, therefore, unlike the typical feature-based SLAM estimation problem which keeps track of the current robot pose and an associated landmark map, the VAN state vector consists entirely of delayed vehicle states corresponding to the vehicle poses at the times the images were captured. This delayed-state approach corresponds to a view-based representation of the environment which can be traced back to a batch scan-matching method by Lu and Milios [9] using laser data and hybrid batch/recursive formulations by Fleischer [10] and McLauchlan [11] using camera images.

The VAN technique proposed the use of an Extended Kalman Filter (EKF) as the fusion framework for merging the navigation and camera sensor measurements. This is a well known approach whose application to SLAM was developed by Smith, Self, and Cheeseman [12] and Moutarlier and Chatila [13]. The EKF maintains the joint correlations over all elements in the state vector and therefore can "optimally" update estimates of all the elements involved in key events like loop closure. However, maintaining these joint correlations also represent a major computational burden since each measurement update requires quadratic complexity in the size of the state vector. This limits the online use of an EKF to relatively small maps — for the VAN approach this translates into an upper bound of approximately one hundred 6-element poses.

The quadratic computational complexity associated with the EKF has been a long recognized issue in the SLAM community. One popular method for reducing the computational burden is to decouple the estimation problem into a series of manageable submaps [14]–[16]. In this approach each map is limited to a fixed number of elements and therefore the computational burden of updating each map has an upper bound. However, the trade-off these techniques make is a reduced rate of convergence since measurements do not optimally update the entire network of submaps [16].

Two other approaches which have gained recent attention are the Sparse Extended Information Filter (SEIF) [17] and the Thin Junction-Tree Filter (TJTF) [18]. The authors of these algorithms make the important empirical observation that when the feature-based SLAM posterior is cast in the form of the Extended Information Filter (EIF), (*i.e.* the dual of the EKF), many of the off-diagonal elements in the information matrix are near zero when properly normalized. SEIFs and TJTFs approximate the posterior with a sparse representation which prevents these weak inter-landmark links from forming. Their approaches effectively limit the density of the information matrix and as each shows — allows for constant time updates. However, the delicate and nontrivial issue that must be dealt with is how to sparsify the information matrix.

In the remainder of this paper we will first illustrate why the feature-based SLAM information matrix is naturally dense and therefore, why SEIFs and TJTFs have to *approximate* the SLAM posterior with a sparse representation. We'll then proceed to show the novel insight that the information form is *exactly* sparse for a delayed-state representation. In turn, we show this sparse property allows us to cast the delayed-state framework in an efficient representation, but without any sparse approximation error.

II. Information Form

The information form is often called the canonical or natural representation of the Gaussian distribution. This notion of a "natural" representation stems from expanding the quadratic in the exponential of the Gaussian distribution as shown in (1). The result is that rather than parameterizing the normal distribution in terms of its mean and covariance as in $\mathcal{N}\big(\boldsymbol{\xi}_t; \boldsymbol{\mu}_t, \Sigma_t\big)$, it is instead parametrized in terms of its information vector and information matrix, $\mathcal{N}^{-1}\big(\boldsymbol{\xi}_t; \boldsymbol{\eta}_t, \Lambda_t\big)$ [19]. Equation (2) shows how the two forms are related.

$$\begin{aligned}
p\left(\boldsymbol{\xi}_t\right) &= \mathcal{N}\big(\boldsymbol{\xi}_t; \boldsymbol{\mu}_t, \Sigma_t\big) \\
&\propto \exp\big\{-\tfrac{1}{2}(\boldsymbol{\xi}_t - \boldsymbol{\mu}_t)^\top \Sigma_t^{-1} (\boldsymbol{\xi}_t - \boldsymbol{\mu}_t)\big\} \\
&= \exp\big\{-\tfrac{1}{2}(\boldsymbol{\xi}_t^\top \Sigma_t^{-1} \boldsymbol{\xi}_t - 2\boldsymbol{\mu}_t^\top \Sigma_t^{-1} \boldsymbol{\xi}_t + \boldsymbol{\mu}_t^\top \Sigma_t^{-1} \boldsymbol{\mu}_t)\big\} \\
&\propto \exp\big\{-\tfrac{1}{2}\boldsymbol{\xi}_t^\top \Sigma_t^{-1} \boldsymbol{\xi}_t + \boldsymbol{\mu}_t^\top \Sigma_t^{-1} \boldsymbol{\xi}_t\big\} \\
&= \exp\big\{-\tfrac{1}{2}\boldsymbol{\xi}_t^\top \Lambda_t \boldsymbol{\xi}_t + \boldsymbol{\eta}_t^\top \boldsymbol{\xi}_t\big\} \\
&\propto \mathcal{N}^{-1}\big(\boldsymbol{\xi}_t; \boldsymbol{\eta}_t, \Lambda_t\big) \qquad (1)
\end{aligned}$$

where the proportionality sign accounts for the normalization constant and

$$\Lambda_t = \Sigma_t^{-1} \qquad \boldsymbol{\eta}_t = \Lambda_t \boldsymbol{\mu}_t \qquad (2)$$

A. Marginalization and Conditioning

The covariance and information representations lead to very different computational characteristics with respect to the fundamental probabilistic operations of marginalization and conditioning. This is important because these two operations appear at the core of any SLAM algorithm, for example motion prediction and measurement updates. Table I summarizes these operations on a Gaussian distribution where we see that the covariance and information representations exhibit a dual relationship with respect to marginalization and conditioning. For example, marginalization is easy in covariance form since it corresponds to extracting the appropriate sub-block from the covariance matrix while in information form it is hard because it involves calculating the Schur complement over the variables we wish to keep. Note that the opposite relation holds true for conditioning which is easy in the information form and hard in covariance form.

TABLE I

Summary of Marginalization and Conditioning Operations on a Gaussian Distribution Expressed in Covariance and Information Form

$$p\left(\boldsymbol{\alpha}, \boldsymbol{\beta}\right) = \mathcal{N}\Big(\Big[\begin{smallmatrix}\boldsymbol{\mu}_\alpha \\ \boldsymbol{\mu}_\beta\end{smallmatrix}\Big], \Big[\begin{smallmatrix}\Sigma_{\alpha\alpha} & \Sigma_{\alpha\beta} \\ \Sigma_{\beta\alpha} & \Sigma_{\beta\beta}\end{smallmatrix}\Big]\Big) = \mathcal{N}^{-1}\Big(\Big[\begin{smallmatrix}\boldsymbol{\eta}_\alpha \\ \boldsymbol{\eta}_\beta\end{smallmatrix}\Big], \Big[\begin{smallmatrix}\Lambda_{\alpha\alpha} & \Lambda_{\alpha\beta} \\ \Lambda_{\beta\alpha} & \Lambda_{\beta\beta}\end{smallmatrix}\Big]\Big)$$

	Marginalization	Conditioning
	$p\left(\boldsymbol{\alpha}\right) = \int p\left(\boldsymbol{\alpha}, \boldsymbol{\beta}\right) d\boldsymbol{\beta}$	$p\left(\boldsymbol{\alpha} \mid \boldsymbol{\beta}\right) = p\left(\boldsymbol{\alpha}, \boldsymbol{\beta}\right) / p\left(\boldsymbol{\beta}\right)$
Cov. Form	$\boldsymbol{\mu} = \boldsymbol{\mu}_\alpha$ $\Sigma = \Sigma_{\alpha\alpha}$	$\boldsymbol{\mu}' = \boldsymbol{\mu}_\alpha + \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)$ $\Sigma' = \Sigma_{\alpha\alpha} - \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta\alpha}$
Info. Form	$\boldsymbol{\eta} = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\boldsymbol{\eta}_\beta$ $\Lambda = \Lambda_{\alpha\alpha} - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\Lambda_{\beta\alpha}$	$\boldsymbol{\eta}' = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\boldsymbol{\beta}$ $\Lambda' = \Lambda_{\alpha\alpha}$

B. Information Form and Feature-Based SLAM

Most SLAM approaches are feature-based which assumes that the robot can extract an abstract representation of features in the environment from its sensor data and then use re-observation of these features for localization [12]. In this approach a landmark map is explicitly built and maintained. The process of concurrently performing localization *and* feature map building are inherently coupled. The robot must then represent a joint-distribution over landmarks and current pose. Using the EKF to represent these coupled errors requires maintaining the cross-correlations in the covariance matrix — in which there are quadratically many. Updating the joint correlations over map and robot leads to an $\mathcal{O}(N^2)$ complexity per update, with N being the number of landmarks in the map.

Recently, some substantial papers have been published in which the authors explore re-parameterizing the feature-based SLAM posterior in the information form [17], [18], [20]. For example, Thrun *et al.* [17] make the observation that when the EIF is used for inference, measurement updates are additive and efficient. The downside of the EIF is that motion prediction is *generally* $\mathcal{O}(N^3)$. However, in [17], the authors show that if the information matrix obeys a certain *sparse* structure, the EIF motion prediction can be performed in *constant* time. To obtain this sparse structure, the authors make an important empirical observation regarding the architecture of the feature-based SLAM information matrix. They show that when properly normalized, many of the inter-landmark constraints in the information matrix are redundant and weak. Based upon this insight, the methods presented in [17] and [18] try to approximate the information matrix with a sparse representation in which these weak inter-landmark constraints have been eliminated allowing for the development of efficient SLAM algorithms like SEIFs and TJTFs.

The delicate issue that must be dealt with in these methods, however, is how to perform the necessary approximation step which is required to keep the information matrix sparse. In fact, the sparsification step is an important issue not to be glossed over because the feature-based SLAM information matrix associated with the joint-posterior over robot pose $\mathbf{x}_t$ and landmark map $\mathbf{M}$ given sensor measurements $\mathbf{z}^t$ and control inputs $\mathbf{u}^t$ (i.e., $p\left(\mathbf{x}_t, \mathbf{M} \mid \mathbf{z}^t, \mathbf{u}^t\right)$) is naturally fully *dense*. As we show next, this density arises from marginalizing out past robot poses.

To see this consider the diagram shown in Fig. 1. We begin with the schematic shown to the upper left which represents the robot $\mathbf{x}_t$ at time t connected to three landmarks $\mathbf{L}_1$, $\mathbf{L}_2$ and $\mathbf{L}_3$ in the context of a non-directed Bayes-Net [21]. This diagram is a graphical representation of the conditional independencies in the distribution $p\left(\mathbf{x}_t, \mathbf{L}_{1:3} \mid \mathbf{z}^t, \mathbf{u}^t\right)$ and shows that the only constraints which exists are between robot and landmark, no inter-landmark constraints appear. This lack of inter-landmark constraints should be correctly interpreted to mean that each landmark is conditionally independent given the robot pose as described in [22], [23]. The intuition behind this comes from viewing the noise of each sensor reading as being independent, and therefore, determining each landmark position is an independent estimation problem given the known location of the sensor.

Shown directly below each Bayes-Net in Fig. 1 is an illustration of the corresponding information matrix. Here we see that the nonzero off-diagonal elements encode the robot/landmark constraints while the zeros in the information matrix encode the lack of direct inter-landmark constraints [18]. The diagram shown in the middle of Fig. 1 represents the intermediate distribution $p\left(\mathbf{x}_{t+1}, \mathbf{x}_t, \mathbf{L}_{1:4} \mid \mathbf{z}^t, \mathbf{u}^t\right)$. This distribution represents a time-propagation of the previous distribution by augmenting the state vector to include the term $\mathbf{x}_{t+1}$ which represents the new robot pose at time $t+1$, re-observation of feature $\mathbf{L}_3$, and observation of a new landmark $\mathbf{L}_4$. Because the robot state evolves according to a Markov process, we see that the new robot state $\mathbf{x}_{t+1}$ is only linked to the previous robot state $\mathbf{x}_t$, and that observation of the landmarks $\mathbf{L}_3$ and $\mathbf{L}_4$ add two additional constraints to $\mathbf{x}_{t+1}$. In the typical feature-based SLAM approach only the current robot pose is estimated and not the complete trajectory. Therefore, we always marginalize out the previous robot pose $\mathbf{x}_t$ during our time-projection step to give us the distribution over current pose and map $p\left(\mathbf{x}_{t+1}, \mathbf{L}_{1:4} \mid \mathbf{z}^t, \mathbf{u}^t\right) = \int p\left(\mathbf{x}_{t+1}, \mathbf{x}_t, \mathbf{L}_{1:4} \mid \mathbf{z}^t, \mathbf{u}^t\right) d\mathbf{x}_t$. Recalling the formula for marginalization applied to a Gaussian in the information form (see Table I), we note that it is the the outer product of $\Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\Lambda_{\alpha\beta}^{\top}$[1] that causes the information matrix to fill in and become dense as shown in the illustration to the far right of Fig. 1.

Intuitively, the landmarks $\mathbf{L}_1$, $\mathbf{L}_2$, $\mathbf{L}_3$, which used to be indirectly connected via a direct relationship with $\mathbf{x}_t$, must now represent that indirect relationship directly by creating new links between each other. Therefore, the penalty for a feature-based SLAM representation which always marginalizes out the robot trajectory is that the landmark Bayes-Net becomes fully connected and the associated information matrix becomes fully *dense* (though as previously mentioned , [17] makes the empirical observation that many of the off-diagonal elements are relatively small).

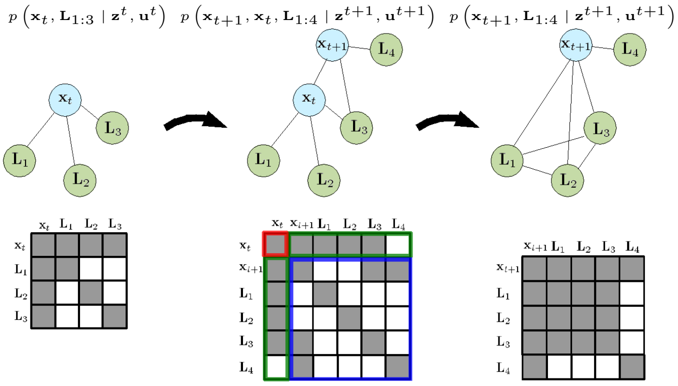
$p\left(\mathbf{x}_t, \mathbf{L}_{1:3} \mid \mathbf{z}^t, \mathbf{u}^t\right)$
$p\left(\mathbf{x}_{t+1}, \mathbf{x}_t, \mathbf{L}_{1:4} \mid \mathbf{z}^{t+1}, \mathbf{u}^{t+1}\right)$
$p\left(\mathbf{x}_{t+1}, \mathbf{L}_{1:4} \mid \mathbf{z}^{t+1}, \mathbf{u}^{t+1}\right)$
$\mathbf{x}_t$
$\mathbf{L}_1$
$\mathbf{L}_2$
$\mathbf{L}_3$
$\mathbf{x}_{t+1}$
$\mathbf{L}_4$
x_t L_1 L_2 L_3
x_t x_{t+1} L_1 L_2 L_3 L_4
x_{t+1} L_1 L_2 L_3 L_4

Fig. 1. A graphical explanation of why the feature-based SLAM information matrix is naturally fully dense. Refer to the text of §II-B for a detailed discussion. (left) The posterior over robot pose $\mathbf{x}_t$ and landmarks $\mathbf{L}_{1:3}$ given sensor measurements $\mathbf{z}^t$ and control inputs $\mathbf{u}^t$ is represented as a non-directed Bayes-Net. The corresponding information matrix is shown directly below and encodes the Bayes-Net link structure within the non-zero off-diagonal elements. (middle) The time-propagation of the posterior is now shown where the state vector has been augmented to include the robot pose at time $t+1$ (i.e., $\mathbf{x}_{t+1}$), re-observation of landmark $\mathbf{L}_3$, and observation of a new landmark $\mathbf{L}_4$. Appropriate sub-blocks of the information matrix have been colored to indicate the different portions involved in marginalizing out the past robot pose $\mathbf{x}_t$. Referring to Table I, $\Lambda_{\alpha\alpha}$ is blue, $\Lambda_{\beta\beta}$ is red, and $\Lambda_{\alpha\beta} = \Lambda_{\beta\alpha}^{\top}$ is green. (right) This posterior reflects the effect of marginalizing out the past robot state $\mathbf{x}_t$ and its consequent "fill in" of the information matrix.

III. Exactly Sparse Delayed-State Filters

An alternative formulation of the SLAM problem is to use a view-based representation rather than a feature-based approach [9]. View-based representations such as [8], [9] do not explicitly model landmark features in the environment, instead the estimation problem consists of keeping track of the current robot pose as well as a sampling of key past poses from it's trajectory. The associated posterior in this case is defined over a collection of delayed-states [8]–[10], [24]. In the the view-based representation, raw sensor data is registered to provide virtual observations of pose displacements. For example, in [9] these virtual observation come from scan-matching laser range data while in our application [8] a camera is used to provide pairwise constraints.

[1] Here $\boldsymbol{\alpha} = \{\mathbf{x}_{t+1}, \mathbf{L}_{1:4}\}$ and $\boldsymbol{\beta} = \mathbf{x}_t$

A. State Augmentation

We begin by describing the method of state augmentation which is how we "grow" the state vector to contain a new delayed-state; this operation occurs whenever we have a new key robot pose that we wish to store. For example, in the VAN framework we add a delayed-state for each acquired image of the environment that we wish to be able to revisit at a later time. Assume for the moment that our estimate at time t is described by the distribution in (3) expressed in both covariance and information form.

$$\begin{aligned} p\left(\mathbf{x}_t, \mathbf{M} \mid \mathbf{z}^t, \mathbf{u}^t\right) &= \mathcal{N}\Big(\begin{bmatrix} \boldsymbol{\mu}_{x_t} \\ \boldsymbol{\mu}_M \end{bmatrix}, \begin{bmatrix} \Sigma_{x_t x_t} & \Sigma_{x_t M} \\ \Sigma_{M x_t} & \Sigma_{MM} \end{bmatrix}\Big) \\ &= \mathcal{N}^{-1}\Big(\begin{bmatrix} \boldsymbol{\eta}_{x_t} \\ \boldsymbol{\eta}_M \end{bmatrix}, \begin{bmatrix} \Lambda_{x_t x_t} & \Lambda_{x_t M} \\ \Lambda_{M x_t} & \Lambda_{MM} \end{bmatrix}\Big) \end{aligned} \quad (3)$$

This distribution represents a map and current robot state, $\mathbf{M}$ and $\mathbf{x}_t$ respectively, given all measurements $\mathbf{z}^t$ and control inputs $\mathbf{u}^t$. Here the map variable $\mathbf{M}$ is used in a general sense, for example it could represent a collection of delayed-states or a set of landmark features in the environment. For now we don't care, because we want to show what happens when we augment our representation to include the new robot state $\mathbf{x}_{t+1}$ obtaining the distribution $p\left(\mathbf{x}_{t+1}, \mathbf{x}_t, \mathbf{M} \mid \mathbf{z}^t, \mathbf{u}^{t+1}\right)$ which can be factored as

$$\begin{aligned} & p\left(\mathbf{x}_{t+1}, \mathbf{x}_t, \mathbf{M} \mid \mathbf{z}^t, \mathbf{u}^{t+1}\right) \\ &\quad = \; p\left(\mathbf{x}_{t+1} \mid \mathbf{x}_t, \mathbf{M}, \mathbf{z}^t, \mathbf{u}^{t+1}\right) p\left(\mathbf{x}_t, \mathbf{M} \mid \mathbf{z}^t, \mathbf{u}^{t+1}\right) \\ &\quad \overset{\text{Markov}}{=} p\left(\mathbf{x}_{t+1} \mid \mathbf{x}_t, \mathbf{u}_{t+1}\right) p\left(\mathbf{x}_t, \mathbf{M} \mid \mathbf{z}^t, \mathbf{u}^t\right) \end{aligned} \quad (4)$$

In (4) we factored the posterior into the product of a probabilistic state-transition multiplied by our prior using the common assumption that the robot state evolves according to a Markov process. Equation (5) describes the general nonlinear Markov robot motion model we assume and (6) its first-order linearized form where F is the Jacobian evaluated at $\boldsymbol{\mu}_{x_t}$ and $\mathbf{w}_t \sim \mathcal{N}\big(\mathbf{0}, \mathrm{Q}\big)$ is the white process noise. Note that in general our state description $\mathbf{x}_t$ consists of pose (i.e., position, orientation) and kinematic components (e.g., body-frame velocities, angular rates).

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{f}(\mathbf{x}_t, \mathbf{u}_{t+1}) + \mathbf{w}_t && (5) \\ &\approx \mathbf{f}(\boldsymbol{\mu}_{x_t}, \mathbf{u}_{t+1}) + \mathrm{F}(\mathbf{x}_t - \boldsymbol{\mu}_{x_t}) + \mathbf{w}_t && (6) \end{aligned}$$

Under the linearized approximation in (6) the augmented state distribution given in (4) is also Gaussian and in covariance form its result is given by (7) [12].

$$\begin{aligned} p\left(\mathbf{x}_{t+1}, \mathbf{x}_t, \mathbf{M} \mid \mathbf{z}^t, \mathbf{u}^{t+1}\right) &= \mathcal{N}\big(\boldsymbol{\mu}'_{t+1}, \Sigma'_{t+1}\big) \\ \boldsymbol{\mu}'_{t+1} &= \begin{bmatrix} \mathbf{f}(\boldsymbol{\mu}_{x_t}, \mathbf{u}_{t+1}) \\ \boldsymbol{\mu}_{x_t} \\ \boldsymbol{\mu}_M \end{bmatrix} \\ \Sigma'_{t+1} &= \begin{bmatrix} (\mathrm{F}\Sigma_{x_t x_t}\mathrm{F}^\top + \mathrm{Q}) & \mathrm{F}\Sigma_{x_t x_t} & \mathrm{F}\Sigma_{x_t M} \\ \Sigma_{x_t x_t}\mathrm{F}^\top & \Sigma_{x_t x_t} & \Sigma_{x_t M} \\ \Sigma_{M x_t}\mathrm{F}^\top & \Sigma_{M x_t} & \Sigma_{MM} \end{bmatrix} \end{aligned} \quad (7)$$

The lower-right 2×2 sub-block of Σ'_{t+1} corresponds to the covariance between the delayed robot state element $\mathbf{x}_t$ and the map $\mathbf{M}$ and has remained unchanged from the prior. Meanwhile, the first row and column contain the cross-covariances associated with the time propagated robot state $\mathbf{x}_{t+1}$ which includes the effect of the process model.

Having obtained the delayed-state distribution in covariance form, we can now transform (7) to its information form (8) via (2). This requires inversion of the 3×3 block covariance matrix Σ'_{t+1} whose derivation we omit here due to space limitation.

$$\begin{aligned} p\left(\mathbf{x}_{t+1}, \mathbf{x}_t, \mathbf{M} \mid \mathbf{z}^t, \mathbf{u}^{t+1}\right) &= \mathcal{N}^{-1}\big(\boldsymbol{\eta}'_{t+1}, \Lambda'_{t+1}\big) \\ \boldsymbol{\eta}'_{t+1} &= \begin{bmatrix} \mathrm{Q}^{-1}\big(\mathbf{f}(\boldsymbol{\mu}_{x_t}, \mathbf{u}_{t+1}) - \mathrm{F}\boldsymbol{\mu}_{x_t}\big) \\ \boldsymbol{\eta}_{x_t} - \mathrm{F}^\top\mathrm{Q}^{-1}\big(\mathbf{f}(\boldsymbol{\mu}_{x_t}, \mathbf{u}_{t+1}) - \mathrm{F}\boldsymbol{\mu}_{x_t}\big) \\ \boldsymbol{\eta}_M \end{bmatrix} \\ \Lambda'_{t+1} &= \begin{bmatrix} \mathrm{Q}^{-1} & -\mathrm{Q}^{-1}\mathrm{F} & 0 \\ -\mathrm{F}^\top\mathrm{Q}^{-1} & (\Lambda_{x_t x_t} + \mathrm{F}^\top\mathrm{Q}^{-1}\mathrm{F}) & \Lambda_{x_t M} \\ 0 & \Lambda_{M x_t} & \Lambda_{MM} \end{bmatrix} \end{aligned} \quad (8)$$

Equation (8) yields a key insight into the structure of the information matrix regarding delayed-states. We see that augmenting our state vector to include the new robot state $\mathbf{x}_{t+1}$ introduces shared information between the new state $\mathbf{x}_{t+1}$ and the previous robot state $\mathbf{x}_t$. Moreover, we see that the shared information between $\mathbf{x}_{t+1}$ and the map $\mathbf{M}$ is *always* zero irrespective of what $\mathbf{M}$ abstractly represents (i.e., it doesn't matter whether $\mathbf{M}$ represents a set of feature landmarks or a collection of delayed-states the result will always be zero). This sparsity in the augmented state information matrix is a consequence of the Markov property associated with the state transition probability $p\left(\mathbf{x}_{t+1} \mid \mathbf{x}_t, \mathbf{u}_{t+1}\right)$. In terms of a Bayes diagram, $\mathbf{x}_{t+1}$ is only serially connected to its parent node $\mathbf{x}_t$ and therefore, is conditionally independent of $\mathbf{M}$.

By induction, a key property of state augmentation in the information form is that if we continue to augment our state vector with additional delayed-states we see that the information matrix will exhibit a block tridiagonal structure linking each delayed-state with the post and previous states as shown in (9). Hence, the view-based delayed-state SLAM information matrix is *naturally* sparse without having to make any sparseness approximations.

$$\begin{bmatrix} \Lambda_{x_{t+1}x_{t+1}} & \Lambda_{x_{t+1}x_t} & & & \\ \Lambda^\top_{x_{t+1}x_t} & \Lambda_{x_t x_t} & \Lambda_{x_t x_{t-1}} & & \\ & \Lambda^\top_{x_t x_{t-1}} & \Lambda_{x_{t-1}x_{t-1}} & \Lambda_{x_{t-1}x_{t-2}} & \\ & & \ddots & \ddots & \ddots \end{bmatrix} \quad (9)$$

B. Relative Pose Measurements

One of the very attractive properties of the information form is that measurement updates are additive in the EIF [17]. Assume the following general nonlinear measurement function (10) and its first order linearized form (11)

$$\begin{aligned} \mathbf{z}_t &= \mathbf{h}(\boldsymbol{\xi}_t) + \mathbf{v}_t && (10) \\ &\approx \mathbf{h}(\bar{\boldsymbol{\mu}}_t) + \mathrm{H}(\boldsymbol{\xi}_t - \bar{\boldsymbol{\mu}}_t) + \mathbf{v}_t && (11) \end{aligned}$$

where $\boldsymbol{\xi}_t$ is the predicted state vector distributed according to $\boldsymbol{\xi}_t \sim \mathcal{N}\big(\bar{\boldsymbol{\mu}}_t, \bar{\Sigma}_t\big) = \mathcal{N}^{-1}\big(\bar{\boldsymbol{\eta}}_t, \bar{\Lambda}_t\big)$, $\mathbf{v}_t$ is the white measurement noise $\mathbf{v}_t \sim \mathcal{N}\big(\mathbf{0}, \mathrm{R}\big)$, and H is the Jacobian evaluated at $\bar{\boldsymbol{\mu}}_t$. Thrun *et al* [17] show that the corresponding EIF update is given by (12).

$$\begin{aligned} \Lambda_t &= \bar{\Lambda}_t + \mathrm{H}^\top \mathrm{R}^{-1} \mathrm{H} \\ \boldsymbol{\eta}_t &= \bar{\boldsymbol{\eta}}_t + \mathrm{H}^\top \mathrm{R}^{-1}\big(\mathbf{z}_t - h(\bar{\boldsymbol{\mu}}_t) + \mathrm{H}\bar{\boldsymbol{\mu}}_t\big) \end{aligned} \tag{12}$$

Equation (12) shows that the information matrix is additively updated by the outer product term $\mathrm{H}^\top \mathrm{R}^{-1} \mathrm{H}$. In general, this outer product modifies all elements of the predicted information matrix $\bar{\Lambda}_t$, however a key observation is that the SLAM Jacobian H is always sparse [17]. For example, in the VAN framework [8] pairwise registration of images I_i and I_j provides a relative pose measurement between states $\mathbf{x}_i$ and $\mathbf{x}_j$ resulting in a sparse Jacobian of the form

$$\mathrm{H} = \begin{bmatrix} 0 \cdots & \frac{\partial \mathbf{h}}{\partial \mathbf{x}_i} & \cdots 0 \cdots & \frac{\partial \mathbf{h}}{\partial \mathbf{x}_j} & \cdots 0 \end{bmatrix} \tag{13}$$

As a result only the four-block elements corresponding to $\mathbf{x}_i$ and $\mathbf{x}_j$ of the information matrix need to be modified. In particular, the information in the diagonal blocks $\bar{\Lambda}_{x_i x_i}$ and $\bar{\Lambda}_{x_j x_j}$ is increased, while new information appears at $\bar{\Lambda}_{x_i x_j}$ and its symmetric counterpart $\bar{\Lambda}_{x_j x_i}$. This new off-diagonal information reflects the addition of a new constraint into the corresponding non-directed Bayes-Net connecting the nodes $\mathbf{x}_i$ and $\mathbf{x}_j$. An important consequence of this result is that given the natural sparse tridiagonal structure of the delayed-state information matrix as shown in (9), we see that the number of nonzero off-diagonal elements in the information matrix will be *linear* in the number of relative pose constraints (i.e., camera measurements).

It is worth pointing out that (10) assumes that the measurements are corrupted by independent noise. Since scan-matching methods rely upon registering raw data, this criteria may be violated if data is reused. In our VAN application, relative pose measurements are generated by pairwise registration of images with common overlap. We use a feature-based image registration approach built upon extracting interest points with the Harris [25] and SIFT [26] interest operators. To ensure measurement independence we only use a set of image feature correspondences once so that camera measurements remain uncorrelated.

C. Motion Prediction

Motion prediction corresponds to a time propagation of the robot's state from time t to time $t+1$. In (8) we derived an expression in information form for the joint-distribution between the time predicted robot pose $\mathbf{x}_{t+1}$ and its previous state $\mathbf{x}_t$ (i.e., $p\left(\mathbf{x}_{t+1}, \mathbf{x}_t, \mathbf{M} \mid \mathbf{z}^t, \mathbf{u}^{t+1}\right)$). To derive the time propagated distribution $p\left(\mathbf{x}_{t+1}, \mathbf{M} \mid \mathbf{z}^t, \mathbf{u}^{t+1}\right)$, we simply marginalize out the previous state $\mathbf{x}_t$ from the joint-distribution in (8). Referring to Table I for marginalization of a Gaussian in information form we have[2]

$$\begin{aligned} p\left(\mathbf{x}_{t+1}, \mathbf{M} \mid \mathbf{z}^t, \mathbf{u}^{t+1}\right) &= \mathcal{N}^{-1}\big(\bar{\boldsymbol{\eta}}_{t+1}, \bar{\Lambda}_{t+1}\big) \\ &= \int p\left(\mathbf{x}_{t+1}, \mathbf{x}_t, \mathbf{M} \mid \mathbf{z}^t, \mathbf{u}^{t+1}\right) d\mathbf{x}_t \end{aligned}$$

$$\begin{aligned} \bar{\boldsymbol{\eta}}_{t+1} &= \begin{bmatrix} \mathrm{Q}^{-1}\big(\mathbf{f}(\boldsymbol{\mu}_{x_t}, \mathbf{u}_{t+1}) - \mathrm{F}\boldsymbol{\mu}_{x_t}\big) \\ \boldsymbol{\eta}_M \end{bmatrix} - \\ &\quad \begin{bmatrix} -\mathrm{Q}^{-1}\mathrm{F} \\ \Lambda_{Mx_t} \end{bmatrix} \Omega^{-1}\Big(\boldsymbol{\eta}_{x_t} - \mathrm{F}^\top \mathrm{Q}^{-1}\big(\mathbf{f}(\boldsymbol{\mu}_{x_t}, \mathbf{u}_{t+1}) - \mathrm{F}\boldsymbol{\mu}_{x_t}\big)\Big) \\ &= \begin{bmatrix} \mathrm{Q}^{-1}\mathrm{F}\Omega^{-1}\boldsymbol{\eta}_{x_t} + \Psi\big(\mathbf{f}(\boldsymbol{\mu}_{x_t}, \mathbf{u}_{t+1}) - \mathrm{F}\boldsymbol{\mu}_{x_t}\big) \\ \boldsymbol{\eta}_M - \Lambda_{Mx_t}\Omega^{-1}\Big(\boldsymbol{\eta}_{x_t} - \mathrm{F}^\top \mathrm{Q}^{-1}\big(\mathbf{f}(\boldsymbol{\mu}_{x_t}, \mathbf{u}_{t+1}) - \mathrm{F}\boldsymbol{\mu}_{x_t}\big)\Big) \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \bar{\Lambda}_{t+1} &= \begin{bmatrix} \mathrm{Q}^{-1} & 0 \\ 0 & \Lambda_{MM} \end{bmatrix} - \begin{bmatrix} -\mathrm{Q}^{-1}\mathrm{F} \\ \Lambda_{Mx_t} \end{bmatrix} \Omega^{-1} \begin{bmatrix} -\mathrm{F}^\top \mathrm{Q}^{-1} & \Lambda_{x_t M} \end{bmatrix} \\ &= \begin{bmatrix} \Psi & \mathrm{Q}^{-1}\mathrm{F}\Omega^{-1}\Lambda_{x_t M} \\ \Lambda_{Mx_t}\Omega^{-1}\mathrm{F}^\top \mathrm{Q}^{-1} & \Lambda_{MM} - \Lambda_{Mx_t}\Omega^{-1}\Lambda_{x_t M} \end{bmatrix} \end{aligned} \tag{14}$$

where

$$\begin{aligned} \Psi &= \mathrm{Q}^{-1} - \mathrm{Q}^{-1}\mathrm{F}(\Lambda_{x_t x_t} + \mathrm{F}^\top \mathrm{Q}^{-1}\mathrm{F})^{-1}\mathrm{F}^\top \mathrm{Q}^{-1} \\ &= (\mathrm{Q} + \mathrm{F}\Lambda_{x_t x_t}^{-1}\mathrm{F}^\top)^{-1} \end{aligned} \tag{15}$$

$$\Omega = (\Lambda_{x_t x_t} + \mathrm{F}^\top \mathrm{Q}^{-1}\mathrm{F}) \tag{16}$$

An important consequence of the delayed-state framework is that (14) can be implemented in constant time. To see this we refer to Fig. 2 which illustrates the effect of motion prediction for a collection of delayed-states. We begin with the Bayes diagram at the top of the figure which shows a segregated collection of delayed-states. Our view-based "map" corresponds to the set of states $\mathbf{M} = \{\mathbf{x}_{t-4}, \mathbf{x}_{t-3}, \mathbf{x}_{t-2}, \mathbf{x}_{t-1}\}$ which have an interconnected dependence due to camera measurements while the states $\mathbf{x}_t$ and $\mathbf{x}_{t+1}$ are serially connected and corresponds to the previous and predicted robot states respectively. Referring back to Table I we see that the bottommost portion of Fig. 2 graphically illustrates the effect of marginalization on the information matrix. We note that since $\mathbf{x}_t$ is only serially connected to $\mathbf{x}_{t+1}$ and $\mathbf{x}_{t-1}$, marginalizing it out only requires modifying the information blocks associated with these elements (i.e., $\Lambda'_{x_{t+1}x_{t+1}}$ and $\Lambda'_{x_{t-1}x_{t-1}}$ shown with cross-hairs and the symmetric blocks $\Lambda'_{x_{t+1}x_{t-1}} = \Lambda'^\top_{x_{t-1}x_{t+1}}$ shown with black dots). Therefore, since only a fixed portion of the information matrix is ever involved in the calculation of (14) motion prediction can be performed in constant time. This is an important result since in practice fusion of asynchronous navigation sensor measurements (e.g., odometry, compass) means that prediction is typically a high bandwidth operation (e.g., $\mathcal{O}(10\mathrm{Hz})$).

D. State Recovery

As discussed in §II, the information form of the Gaussian is parameterized by its information vector and information

[2] Equation (15) employs the matrix inversion lemma $(\mathrm{A} + \mathrm{BCB}^\top)^{-1} = \mathrm{A}^{-1} - \mathrm{A}^{-1}\mathrm{B}(\mathrm{B}^\top \mathrm{A}^{-1}\mathrm{B} + \mathrm{C}^{-1})^{-1}\mathrm{B}^\top \mathrm{A}^{-1}$

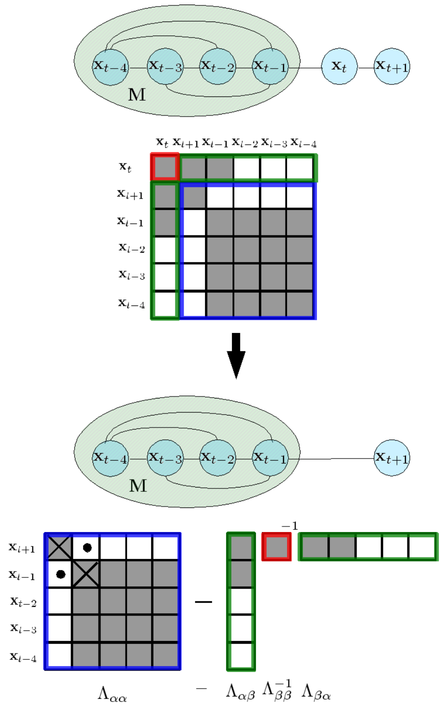

Fig. 2. Graphical illustration of the effect of motion prediction within a delayed-state framework. (top) A non-directed Bayes diagram for a segregated collection of delayed-states. The view-based "map" $\mathbf{M}$ is composed of the set $\mathbf{M} = \{\mathbf{x}_{t-4}, \mathbf{x}_{t-3}, \mathbf{x}_{t-2}, \mathbf{x}_{t-1}\}$ which is a collection of delayed-states that are interlinked by camera constraints. The previous and predicted robot states, $\mathbf{x}_t$ and $\mathbf{x}_{t+1}$ respectively, are serially linked to the map. Below the Bayes diagram is a schematic showing the nonzero structure (colored in gray) of the associated information matrix. (bottom) Recalling from Table I the expression for marginalization of a Gaussian in information form we see that the bottommost schematic illustrates this operation graphically. The end result is that only the states which were linked to $\mathbf{x}_t$ (i.e., $\mathbf{x}_{t-1}$ and $\mathbf{x}_{t+1}$) are effected by the marginalization operation as indicated by the cross-hairs and black dots superimposed on $\Lambda_{\alpha\alpha}$.

matrix, $\boldsymbol{\eta}_t$ and Λ_t respectively. However, the expressions for motion prediction (14) and measurement update (12) additionally require sub-elements from the state mean vector $\boldsymbol{\mu}_t$ so that the nonlinear models (5) and (10) can be linearized. Therefore, in order for the information form to be a computationally efficient parameterization for delayed-states we also need to be able to easily recover portions of the state mean vector $\boldsymbol{\mu}_t$. Fortunately this is the case due to the sparse structure of the information matrix Λ_t.

Recovery of the state mean vector $\boldsymbol{\mu}_t$ corresponds to solving the sparse, symmetric, positive-definite, linear system of equations

$$\Lambda_t \boldsymbol{\mu}_t = \boldsymbol{\eta}_t \tag{17}$$

Such systems can be iteratively solved via the classic method of conjugate gradients [27]. In addition, the recently proposed multigrid SLAM methods by Konolige [28] and Frese *et al* [29] appear capable of solving this system very efficiently incurring only linear asymptotic complexity; moreover, an important observation regarding the expressions for motion prediction (14) and measurement updates (12) is that they only require knowing *subsets* of the state mean $\boldsymbol{\mu}_t$. In light of this we note that rather than solving for the entire state mean vector $\boldsymbol{\mu}_t$, we can partition (17) into two sets of coupled equations as shown in (18).

$$\begin{bmatrix} \Lambda_{\ell\ell} & \Lambda_{\ell b} \\ \Lambda_{b\ell} & \Lambda_{bb} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu}_\ell \\ \boldsymbol{\mu}_b \end{bmatrix} = \begin{bmatrix} \boldsymbol{\eta}_\ell \\ \boldsymbol{\eta}_b \end{bmatrix} \tag{18}$$

This partitioning of the full state vector $\boldsymbol{\mu}_t$ into what we call the "local portion" of the map $\boldsymbol{\mu}_\ell$ and the "benign portion" $\boldsymbol{\mu}_b$, allows us to sub-optimally solve for the local portion of the map we are interested in. By holding our current estimate for the benign portion of the map fixed, we can solve (18) for an estimate of the local portion $\hat{\boldsymbol{\mu}}_\ell$ as

$$\hat{\boldsymbol{\mu}}_\ell = \Lambda_{\ell\ell}^{-1} \left(\boldsymbol{\eta}_\ell - \Lambda_{\ell b} \hat{\boldsymbol{\mu}}_b \right) \tag{19}$$

Equation (19) gives us a method for recovering an estimate of the local map $\hat{\boldsymbol{\mu}}_\ell$ provided that our estimate for the benign portion $\hat{\boldsymbol{\mu}}_b$ is a decent approximation to the actual mean $\boldsymbol{\mu}_b$. Furthermore, note that only a *subset* of $\hat{\boldsymbol{\mu}}_b$ is actually required in the calculation of $\hat{\boldsymbol{\mu}}_\ell$ corresponding to the nonzero elements in the sparse matrix $\Lambda_{\ell b}$. Adopting [17]'s notation, this active subset, denoted $\boldsymbol{\mu}_b^+$, represents the Markov blanket of $\boldsymbol{\mu}_\ell$ and corresponds to elements which are directly connected to $\boldsymbol{\mu}_\ell$ in the associated Bayes constraint network. Therefore, calculation of the local map $\hat{\boldsymbol{\mu}}_\ell$ only requires an estimate of the *locally* connected delayed-state network $\hat{\boldsymbol{\mu}}_b^+$ and does not depend upon passive elements in the benign portion of the map.

In particular, (19) provides an accurate and constant-time approximation for recovering the robot mean during incorporation of high bandwidth navigation sensor measurements since the robot state is only serially connected to the map as discussed in section §III-C. Note though, that (19) will only provide a good approximation as long as the mean estimates $\hat{\mathbf{u}}_b^+$ are accurate. In the case that they are not (e.g., as a result of loop closure) the true full mean $\boldsymbol{\mu}_t$ should be solved for via (17).

IV. Discussion

A. Connection to Lu-Milios

The concept of a view-based map representation has strong roots going back to a seminal paper by Lu and Milios [9]. Their approach sidestepped difficulties associated with feature segmentation and representation by doing away with an explicit feature-based representation of the environment. Rather, their technique indirectly represented a physical map via a collection of global robot poses and raw scan data. To determine the global poses, they formulated the nonlinear optimization problem as one of estimating a set of global robot poses consistent with the relative pose constraints obtained by scan matching and odometry. They then solved this sparse nonlinear optimization problem in an batch-iterative fashion. Our Exactly Sparse Delayed State Filter (ESDSF) framework essentially attempts to recursively solve the same problem. Note though that in the ESDSF framework the nonlinear relative pose constraints are only linearized *once* about the current state when the measurement is incorporated via (12) while in the noncausal Lu-Milios batch formulation they are *relinearized* around the current best estimate of the state at *each* iteration of the nonlinear optimization. This implies that while the ESDSF solution can be performed recursively, it will be more prone to linearization error.

B. Connection to Feature-Based SLAM

Another interesting theoretical connection involves relating the delayed-state SLAM framework to feature-based SLAM. In §II-B we saw that the feature-based SLAM information matrix is naturally *dense* as a result of marginalizing out the robot's trajectory. On a similar train-of-thought, we can view the off-diagonal elements appearing in the delayed-state SLAM information matrix as being a result of marginalizing out the *landmarks*. Since landmarks are only ever locally observed, they only create links to spatially close robot states. Therefore, each time we eliminate a landmark, it only introduces new off-diagonal elements into the information matrix linking all robot states which observed that landmark. Interestingly, this same type of constraint phenomenon also appears in photogrammetry and in particular in large scale bundle adjustment techniques [30]. These techniques are based upon a partitioned Levenberg-Marquardt algorithm which takes advantage of the inherent sparsity between camera and 3D feature constraints in the reconstruction problem. Their central component is based upon eliminating 3D-structure equations to yield a coupled set of equations over camera poses which they solve and then back-substitute to recover the associated 3D-structure. Therefore, loosely speaking the delayed-state information framework represents a recursive linearized formulation of this same problem.

V. Experimental Result

This section presents an initial experimental result validating the sparseness of the delayed-state information filter framework. Our application consists of a single camera 6 DOF SLAM implementation which uses a view-based representation of the environment. This representation is based upon storing a collection of images of the environment and estimating their associated robot states. Details of our implementation can be found in [8], [31].

The experimental setup consisted of a downward-looking digital still camera mounted to a moving underwater pose instrumented ROV at the Johns Hopkins University Hydrodynamic Test Facility. Their vehicle is instrumented with a typical suite of oceanographic dead-reckoning navigation sensors capable of measuring heading, attitude, XYZ bottom-referenced Doppler velocities, and a pressure sensor for depth. The vehicle and test facility are also equipped with a high frequency acoustic LBL system which provides centimeter-level bounded error XY vehicle positions used for validation purposes only. A simulated seafloor environment was created by placing textured carpet, riverbed rocks, and landscaping boulders on the tank floor and was appropriately scaled to match a rugged seafloor environment with considerable 3D scene relief.

Fig. 3 shows the result of estimating the ROV delayed-states associated with a 100 image sequence using a full covariance EKF and sparse EIF. To benchmark compare the results of these two filters, we employed the full state recovery technique of (17) to obtain the state mean used for linearizing (12) and (14). For this experiment the vehicle started near the top-left corner of the plot at (-2.5,2.75) and then drove a course consisting of two grid-based surveys, one oriented SW to NE and the other W to E. The top plot shows the spatial XY pose topology, 99.9% confidence bounds, and link network of camera constraints. Green links correspond to temporally consecutive images which were successfully registered while red links correspond to spatially registered image pairs. The bottommost plot in this figure compares the densities associated with the full EKF covariance matrix versus the EIF information matrix. Note that while the EKF correlation matrix is dense, the information matrix exhibits a sparse tridiagonal structure with the number of off-diagonal elements being linear in the number of camera measurements. In all there are 305 camera constraints (90 temporal / 215 spatial) and each delayed-state is a 12 vector consisting of 6 pose and 6 kinematic components. Therefore, 100 delayed-states results in a 1200×1200 information matrix containing $12^2(100 + 2 \cdot 99) + 6^2(2 \cdot 215) = 58,392$ nonzero elements as shown. We found the EKF and EIF solutions to be numerically equivalent, and furthermore that the EIF only required 1/6 the storage of the EKF for this experiment.

VI. Conclusions

In conclusion, we showed that the feature-based SLAM information matrix is naturally dense and illustrated that this is a consequence of marginalizing out the robot's trajectory. This forces algorithms like SEIFs or TJTFs to make approximations to keep the feature-based SLAM information matrix sparse. We then showed that the delayed-state SLAM information matrix is *exactly* sparse with a tridiagonal block structure due to the Markov property of the process model. Furthermore, the number of nonzero off-diagonal elements in the information matrix is *linear* in the number of measured relative pose constraints. This sparse information matrix structure allows for efficient full state recovery via recently proposed multigrid SLAM methods. We also presented a technique for sub-optimal partial state recovery which allows motion prediction and navigation measurement updates to be performed in constant time.

A remaining open issue with the delayed-state information framework is to provide efficient access to probabilities, (i.e., the covariance matrix), for tasks such as proposing link hypotheses and detecting loop closure. This issue is currently under investigation.

VII. Acknowledgments

We graciously thank our hosts Louis Whitcomb and James Kinsey for their collaboration in collecting the JHU data set. This work was funded in part by the CenSSIS ERC of the NSF under grant EEC-9986821 and in part by WHOI through a grant from the Penzance Foundation. This paper is WHOI contribution number 11294.

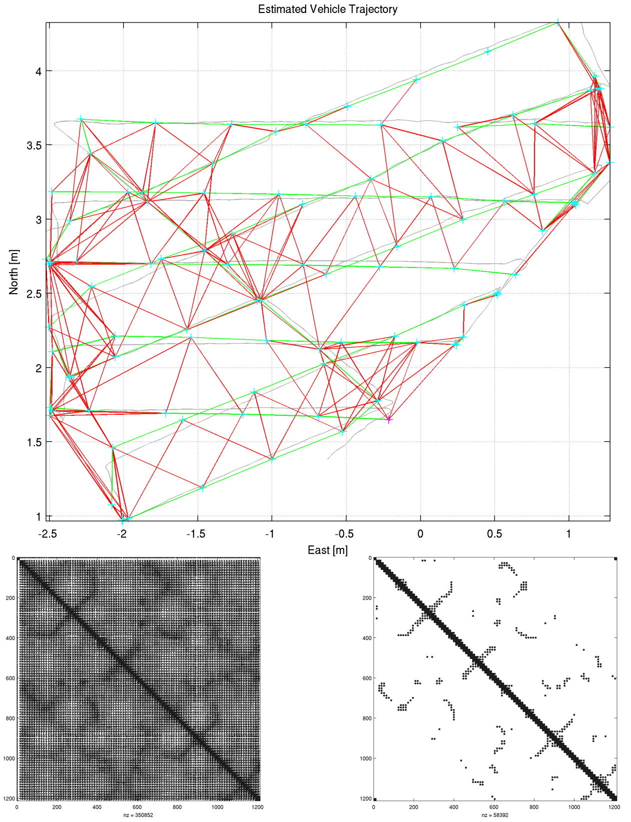

Fig. 3. This figure contrasts the exact sparsity of the delayed-state information matrix versus the density of the full covariance matrix; refer to §V for details. (top) Spatial topology of a 100 image sequence of underwater images collected from the JHU ROV. In all there are 305 camera constraints. (bottom) The covariance and information matrices associated with the delayed-state topology, left and right respectively. For the covariance matrix, all elements above a normalized correlation score of 0.1% are shown. The thresholded covariance matrix has 350,852 nonzero elements while the information matrix has only 58,392. The covariance matrix and information matrix are numerically equivalent, however the information matrix is exactly sparse.

References

[1] J. Reynolds, R. Highsmith, B. Konar, C. Wheat, and D. Doudna, "Fisheries and Fisheries Habitat Investigations Using Undersea Technology," in *MTS/IEEE OCEANS 2001*, vol. 2, Honolulu, HI, USA, November 2001, pp. 812–820.

[2] H. Singh, R. Armstrong, F. Gilbes, R. Eustice, C. Roman, O. Pizarro, and J. Torres, "Imaging Coral I: Imaging Coral Habitats with the SeaBED AUV," *The Journal for Subsurface Sensing Technologies and Applications*, vol. 5, no. 1, pp. 25–42, January 2004.

[3] H. Singh, J. Adams, B. Foley, and D. Mindell, "Imaging Underwater for Archeology," *American Journal of Field Archaeology*, Accepted, To Appear.

[4] M. Hunt, W. Marquet, D. Moller, K. Peal, W. Smith, and R. Spindel, "An Acoustic Navigation System," Woods Hole Oceanographic Institution, Tech. Rep. WHOI-74-6, December 1974.

[5] "Acoustic Doppler Current Profiler: Principles of Operation a Practical Primer," RD Instruments, Tech. Rep., 1996.

[6] T. Gaiffe, "U-Phins: A Fog-Based Inertial Navigation System Developed Specifically for AUV Navigation and Control," in *International Conference on Underwater Intervention*, New Orleans, LA, February 2002.

[7] L. Whitcomb, D. Yoerger, and H. Singh, "Advances in Doppler-Based Navigation of Underwater Robotic Vehicles," in *Proceedings of the 1999 International Conference on Robotics and Automation*, vol. 1, 1999, pp. 399–406.

[8] R. Eustice, O. Pizarro, and H. Singh, "Visually Augmented Navigation in an Unstructured Environment Using a Delayed State History," in *Proceedings of the 2004 IEEE International Conference on Robotics and Automation*, vol. 1, New Orleans, USA, April 2004, pp. 25–32.

[9] F. Lu and E. Milios, "Globally Constent Range Scan Alignment for Environment Mapping," *Autonomous Robots*, vol. 4, pp. 333–349, April 1997.

[10] S. Fleischer, "Bounded-Error Vision-Based Navigation of Autonomous Underwater Vehicles," Ph.D. Thesis, Stanford University, 2000.

[11] P. F. McLauchlan, "A Batch/Recursive Algorithm for 3D Scene Reconstruction," in *Int. Conf. Computer Vision and Pattern Recognition*, vol. 2, Hilton Head, SC, USA, 2000, pp. 738–743.

[12] R. Smith, M. Self, and P. Cheeseman, "A Stochastic Map for Uncertain Spatial Relationships," in *Autonomous Mobile Robots*, S. Lyengar and A. Elfes, Eds. IEEE Computer Society Press, Date 1991, pp. 323–330.

[13] P. Moutarlier and R. Chatila, "An Experimental System for Incremental Environment Modeling by an Autonomous Mobile Robot," in *Proceedings of the 1st International Symposium on Experimental Robotics*, Montreal, Canada, June 1989.

[14] J. Leonard and H. Feder, "Decoupled Stochastic Mapping," *IEEE Journal of Oceanic Engineering*, vol. 26, no. 4, pp. 561–571, 2001.

[15] M. Bosse, "Atlas, a Framework for Scalable Mapping," Ph.D. Thesis, Massachusetts Institute of Technology, 2004.

[16] J. Leonard and P. Newman, "Consistent, Convergent, and Constant-Time Slam," in *International Joint Conference on Artificial Intelligence*, Acapulco, Mexico, August 2003.

[17] S. Thrun, Y. Liu, D. Koller, A. Ng, Z. Ghahramani, and H. Durrant-Whyte, "Simultaneous Localization and Mapping with Sparse Extended Information Filters," *International Journal of Robotics Research*, Accepted, To Appear.

[18] M. Paskin, "Thin Junction Tree Filters for Simultaneous Localization and Mapping," in *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, San Francisco, CA, 2003, pp. 1157–1164.

[19] Y. Bar-Shalom, X. Rong Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York: John Wiley & Sons, Inc., Date 2001.

[20] U. Frese and G. Hirzinger, "Simultaneous Localization and Mapping - a Discussion," in *Proceedings of the IJCAI Workshop Reasoning with Uncertainty in Robotics*, Seattle, WA, 2001, pp. 17–26.

[21] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Prentice Hall, Date 2003.

[22] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "Fastslam: A Factored Solution to the Simultaneous Localization and Mapping Problem," in *Proceedings of the AAAI National Conference on Artificial Intelligence*. Edmonton, Canada: AAAI, 2002.

[23] K. Murphy, "Bayesian Map Learning in Dynamic Environments," in *Advances in Neural Information Processing Systems*. MIT Press, 1999.

[24] J. Leonard and R. Rikoski, "Incorporation of Delayed Decision Making into Stochastic Mapping," in *Experimental Robotics VII*, ser. Lecture Notes in Control and Information Sciences, vol. 271. Springer-Verlag, 2001, pp. 533–542.

[25] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," in *Proceedings of the 4th Alvey Vision Conference*, Manchester, U.K., 1988, pp. 147–151.

[26] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[27] J. Shewchuk, "An Introduction to the Conjugate Gradient Method without the Agonizing Pain," Carnegie Mellon Univerisity, Technical Report CMU-CS-94-125, August 1994.

[28] K. Konolige, "Large-Scale Map-Making," in *Proceedings of the AAAI*, San Jose, CA, 2004, pp. 457–463.

[29] U. Frese, P. Larsson, and T. Duckett, "A Multigrid Algorithm for Simultaneous Localization and Mapping," *IEEE Transactions on Robotics*, Accepted, To Appear.

[30] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle Adjustment – a Modern Synthesis," in *Vision Algorithms: Theory and Practice*, ser. Lncs, W. Triggs, A. Zisserman, and R. Szeliski, Eds. Springer-Verlag, Date 2000, pp. 298–375.

[31] O. Pizarro, R. Eustice, and H. Singh, "Relative Pose Estimation for Instrumented, Calibrated Imaging Platforms," in *Proceedings of Digital Image Computing Techniques and Applications*, Sydney, Australia, December 2003, pp. 601–612.