A Proof for the Approximate Sparsity of SLAM Information Matrices*

Udo Frese
Bremen Institute of Safe Systems
Universität Bremen
D-28334 Bremen, Germany
ufrese@informatik.uni-bremen.de

***Abstract*— For the Simultaneous Localization and Mapping problem several efficient algorithms have been proposed that make use of a sparse information matrix representation (e.g. SEIF, TJTF, treemap). Since the exact SLAM information matrix is dense, these algorithm have to approximate it (*sparsification*). It has been empirically observed that this approximation is adequate because entries in the matrix corresponding to distant landmarks are extremely small.**

This paper provides a theoretical proof for this observation, specifically showing that the off-diagonal entries corresponding to two landmarks decay exponentially with the distance traveled between observation of first and second landmark.

***Index Terms*— SLAM, Information Matrix, Sparsification, SEIF, TJTF, treemap**

I. Introduction

Several efficient SLAM algorithms utilize a sparse information matrix representation. To the author's knowledge the first one was Consistent Pose Estimation by Lu & Milios [1]. Their approach directly builds a sparse linear equation system from measured metric relations between adjacent robot poses. More generally for Linearized Least Square (LLS) it is well known that if measurements involve only "local" variables the resulting information matrix A is sparse [2, §15.5] whereas the resulting covariance matrix A^{-1} not necessarily is. Lu & Milios did not exploit sparsity which was later achieved by Duckett et al. [3], [4]. One drawback of Lu & Milios approach is that all old robot poses need to be represented as variables making the representation size grow even when moving through an area already mapped. While removing old poses from representation is trivial in a covariance representation, removing them from an information matrix requires computing the so called Schur complement that destroys sparsity. So while the original matrix A is sparse, the result P' after eliminating old robot poses theoretically is dense (Fig. 1).

It has been conjectured by the author [5] and empirically observed by Thrun et al. [6] that the SLAM information matrix P' is *approximately sparse*, i.e. entries of distant landmarks are very small and thus can be replaced by a sparse approximation. They utilize this observation in their Sparse Extended Information Filter (SEIF). Later Paskin [7] proposed the Thin Junction Tree Filter (TJTF) and Frese the treemap algorithm [8] that are both based on the same observation.

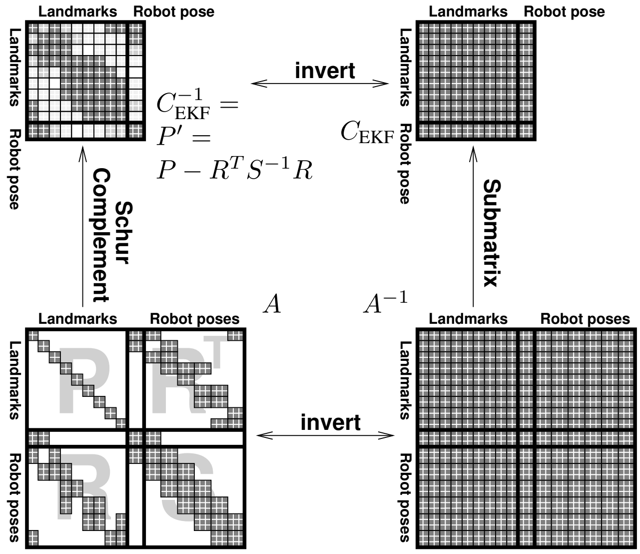

Fig. 1. Relation between the least squares information matrix $A = \begin{pmatrix} P & R^T \\ R & S \end{pmatrix}$ (lower left) which represents all robot poses and the covariance matrix C_{EKF} (upper right) used by EKF only representing the current robot pose. A^{-1} is the covariance matrix corresponding to A representing all robot poses (lower right). C is derived from A^{-1} as a submatrix. Accordingly C_{EKF}^{-1} (upper left) is the information matrix corresponding to C_{EKF} representing only the current robot pose. It is derived from A via Schur complement. So overall, taking a submatrix of a covariance matrix is equivalent to applying Schur complement to an information matrix.

The goal of this paper is to provide a theoretical foundation for these approaches by proofing that

> *in the SLAM information matrix P' off-diagonal entries corresponding to two landmarks decay exponentially with the distance traveled between observation of first and second landmark.*

This result is important both for computation and analysis. First, the approach of saving space and computation time by making the information matrix sparse is being confirmed. Second the result implies that the large scale uncertainty structure of a map estimate is generated by local uncertainties composed along the path the robot has been traveling. Thus, in contrast to the local uncertainty structure, it is rather simple and dominated by the map's geometry [9].

*This paper is based on the research conducted during the authors Ph.D. studies at the German Aerospace Center (DLR)

The outline of the paper is as follows: Section II compares the information matrix representation with the covariance representation used in the popular Extended Kalman Filter (EKF) [10]. Section III outlines the overall proof which is described in detail in sections IV to VI. Implications of the result are discussed in section VII.

II. Covariance vs. Information Matrices

Covariance and information matrices are complementary representations of uncertainty, since one is the inverse of the other. This duality extends to the operation of taking a submatrix, which is equivalent to applying Schur - complement in the inverse (Woodbury formula, [2][§2.7], [11]). Certainly this holds for any decomposition of A into 2×2 blocks $\left(\begin{smallmatrix} P & R^T \\ R & S \end{smallmatrix}\right)$. Of particular interest is the decomposition with rows and columns of the first block corresponding to landmarks (maybe including the current robot pose) and rows and columns of the second block corresponding to (old) robot poses (Fig. 1). In this case $P'^{-1} = \left(P - R^T S^{-1} R\right)^{-1}$ is the covariance matrix of all landmarks (and the current robot pose) as used by the EKF.

The Schur complement $P' = P - R^T S^{-1} R$ equals the corresponding submatrix P minus a correction term $R^T S^{-1} R$. This term can be thought of as somehow "transferring" the effect of S into the realm of P via a mapping provided by the off-diagonal block R^T.

Taking a submatrix of the information matrix or applying Schur - complement to the covariance matrix corresponds to random variables (landmark positions, robot poses) in the removed rows and columns being exactly known. Conversely taking a submatrix of the covariance matrix or applying Schur - complement to the information matrix corresponds to random variables in the removed rows and columns being unknown, i.e. all information about them is discarded.

The main difference between information and covariance matrix lies in the representation of indirect relations. Assume the robot is at pose P1 observing landmark L1 and moves to P2 observing L2. The measurements directly define relations P1-L1, P1-P2, P2-L2, indirectly constituting a relation L1-L2. The covariance matrix explicitly stores this relation in the off-diagonal entries corresponding to L1-L2, whereas the information matrix does not.

Thus the information matrix $A = \left(\begin{smallmatrix} P & R^T \\ R & S \end{smallmatrix}\right)$ used by LLS is sparse, having non-zero off-diagonal entries only for those pairs of random variables which are involved in a common measurement (Fig. 1). The inverse A^{-1} is the covariance matrix for the landmarks and *all* robot poses. A^{-1} represents all indirect relations explicitly and thus is not sparse. Removing the rows and columns corresponding to old robot poses yields the covariance matrix C of the EKF. Its inverse $C^{-1} = P'$ is the information matrix of all landmarks and the current robot pose. However, the inverse is not the corresponding submatrix of A, as eliminating all old robot poses from A requires computing their implicit effect on relations between the other random variables by Schur complement ($P' = P - R^T S^{-1} R$).

Although A is sparse the Schur complement $P - R^T S^{-1} R$ is dense, because S^{-1} is dense[1]. What is turning out is that it is *approximately sparse* with an off-diagonal entry $(P - R^T S^{-1} R)_{l_1 l_2}$ corresponding to two landmarks l_1, l_2 decaying exponentially with the distance traveled between observation of l_1 and l_2. This is the central result of this paper and will be shown in the following:

III. Proof Outline

The proof is essentially an analysis of information matrix A. It is a block matrix $A = \left(\begin{smallmatrix} P & R^T \\ R & S \end{smallmatrix}\right)$ with the first block row / column corresponding to the landmarks and the second corresponding to the different robot poses[2].

As discussed in the previous section, the diagonal blocks P and S are information matrices of two related subproblems: P is the information matrix of the mapping subproblem, describing the uncertainty of the landmarks, if the robot poses were known. Conversely, S is the information matrix of the localization subproblem, describing the uncertainty of the robot poses, if the landmarks were known. Both matrices are extremely sparse: P is block diagonal and S is block tridiagonal.

The matrix P' under investigation will be the information matrix of the landmarks only, i.e. without robot poses. It is $P - R^T S^{-1} R$ by Schur - complement. The role of R^T in this formula is to provide a mapping from robot poses to landmarks. It creates an off-diagonal entry between two landmarks, whose magnitude depends on the entry in S^{-1} corresponding to the two robot poses these landmarks have been observed from. S^{-1} is the covariance of all robot poses given the position of all landmarks. Hence the magnitude of an off-diagonal entry corresponding to two landmarks depends on the covariance the robot poses had if all landmark positions were known.

This covariance decays exponentially with the distance traveled. Intuitively the reason therefore is that in each localization step the pose estimate is replaced by a weighted sum of the old estimate and the measurements of observed landmarks. The covariance with a fixed old robot pose is reduced by a constant factor. Formally, this result is derived by bounding the eigenvalues of S (lemma 1 [12]) and applying a theorem on the decay of off-diagonal entries in the inverse of band matrices (theorem 1 [13]).

The proof is based on a close inspection of different parts of A affected by different measurements, being formally defined in the following subsection. Keeping in mind the structure of A as shown in Fig. 2 and the summary of definitions in table I will be sufficient to understand the argument of the proof.

IV. Sparsity Pattern of the Information Matrix

Each landmark is represented by 2 coordinates and each robot pose by 3. Thus P consists of 2×2 - blocks for each

[1] It is interesting to note, that if odometry is neglected, S and S^{-1} both become block diagonal and $P - R^T S^{-1} R$ is exactly sparse. All discussed algorithms could be used without sparsification in this case.

[2] To simplify the proof technically the current robot pose is included in S not P.

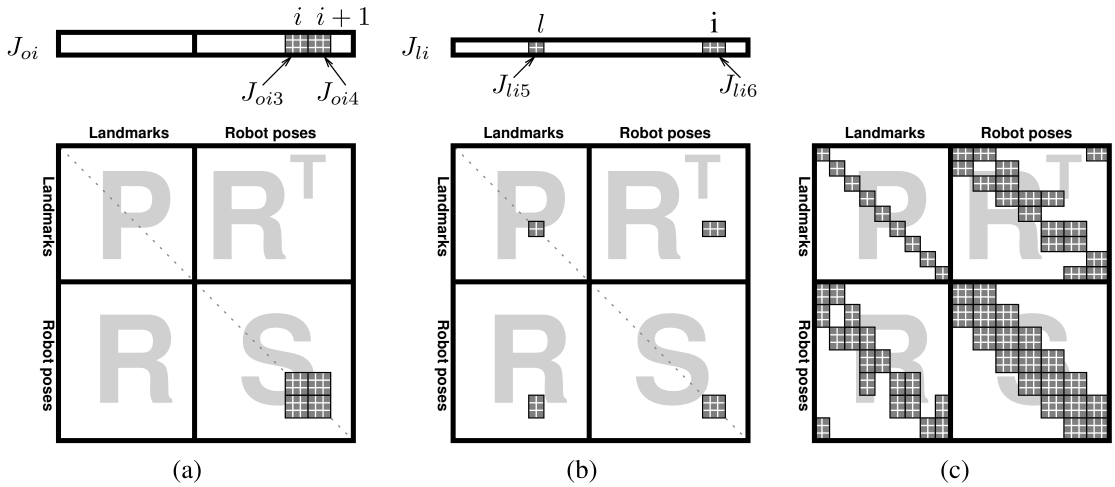

Fig. 2. Sparsity pattern of $A = \begin{pmatrix} P & R^T \\ R & S \end{pmatrix}$: (a) Jacobian J_{oi} for an odometry measurement; Non-zero blocks generated in S thereby (b) Jacobian J_{li} for a landmark measurement; Non-zero blocks generated in P, R and S thereby (c) example for a complete matrix.

landmark, S of 3×3 - blocks for each robot pose, and R of 3×2 - blocks coupling landmarks and robot poses. So with n landmarks, m measurements, and p robot poses, P is a $2n \times 2n$, R a $3p \times 2n$, and S a $3p \times 3p$ matrix. Throughout the whole paper subscripts P_{ll}, R_{li}, and S_{ij} refer to the block corresponding to landmark l and robot pose i or j respectively. The matrix A has a very specific sparsity pattern being analyzed in the following (Fig. 2):

Therefore denote by subscript oi the i-th odometry measurement measuring robot pose $i+1$ relative to robot pose i with C_{oi} measurement covariance and J_{oi} measurement Jacobian. Further let $\mathcal{L}_i$ denote the set of landmark measurements taken from robot pose i and conversely $\mathcal{O}_l$ the set of robot poses from which landmark l has been observed. Clearly $l \in \mathcal{L}_i$ holds if and only if $i \in \mathcal{O}_l$. For a landmark $l \in \mathcal{L}_i$ let C_{li} be the covariance of the measurement of landmark l from robot pose i and J_{li} the measurement Jacobian. The information matrix A of all landmarks and all robot poses is the sum of $J_j^T C_j^{-1} J_j$ over all measurements with J_j Jacobian and C_j covariance. With the definitions made above, these measurements can be grouped by the robot pose and separated into odometry and landmark measurements as

$$A = \sum_{j=1}^{m} J_j^T C_j^{-1} J_j = \sum_{i=1}^{p} \left(J_{oi}^T C_{oi}^{-1} J_{oi} + \sum_{l \in \mathcal{L}_i} J_{li}^T C_{li}^{-1} J_{li} \right). \tag{1}$$

The Jacobian J_{oi} is sparse having a 3×3 nonzero block J_{oi3} at the columns corresponding to robot pose i and another block J_{oi4} at the columns corresponding to robot pose $i+1$. Similarly, J_{li} has a 2×3 nonzero block J_{li5} at the columns corresponding to robot pose i and a 2×2 block J_{li6} at the columns corresponding to landmark l (Fig. 2a, b). Expressions for $J_{oi3}, J_{oi4}, J_{li5}, J_{li6}$ can be derived from a concrete measurement model but do not matter here.

The structure of the Jacobians can be formally expressed using projection matrices. Let therefore I_i denote the block row of the identity matrix corresponding to robot pose i and I_l the block row corresponding to landmark l:

$$J_{oi} = J_{oi3} I_i + J_{oi4} I_{i+1}, \quad J_{li} = J_{li5} I_i + J_{li6} I_l \tag{2}$$

$$\begin{aligned}
A = & \sum_{i=1}^{p} (J_{oi3} I_i + J_{oi4} I_{i+1})^T C_{oi}^{-1} (J_{oi3} I_i + J_{oi4} I_{i+1}) \\
& + \sum_{i=1}^{p} \sum_{l \in \mathcal{L}_i} (J_{li5} I_i + J_{li6} I_l)^T C_{li}^{-1} (J_{li5} I_i + J_{li6} I_l) \\
= & \sum_{i=1}^{p} I_i^T (J_{oi3}^T C_{oi}^{-1} J_{oi3}) I_i + I_{i+1}^T (J_{oi4}^T C_{oi}^{-1} J_{oi3}) I_i \\
& + \sum_{i=1}^{p} I_i^T (J_{oi3}^T C_{oi}^{-1} J_{oi4}) I_{i+1} + I_{i+1}^T (J_{oi4}^T C_{oi}^{-1} J_{oi4}) I_{i+1} \\
& + \sum_{i=1}^{p} \sum_{l \in \mathcal{L}_i} I_i^T (J_{li5}^T C_{li}^{-1} J_{li5}) I_i + I_l^T (J_{li6}^T C_{li}^{-1} J_{li5}) I_i \\
& + \sum_{i=1}^{p} \sum_{l \in \mathcal{L}_i} I_i^T (J_{li5}^T C_{li}^{-1} J_{li6}) I_l + I_l^T (J_{li6}^T C_{li}^{-1} J_{li6}) I_l .
\end{aligned} \tag{3}$$

An expression like $I_i^T(\ldots)I_l$ places the 3×2 matrix in parentheses at the row and column corresponding to robot pose i and landmark l. The other combination $I_i^T(\ldots)I_j$, $I_l^T(\ldots)I_i$ and $I_l^T(\ldots)I_l$ act similar. From (3) it can be seen that each odometry measurement generates four small blocks at the intersections of the rows and columns corresponding to two successive robot poses (Fig. 2a). Correspondingly, each landmark measurement is generating four small blocks at the intersections of the rows and columns corresponding to the robot pose and the landmark (Fig. 2b).

The terms in expression (3) can be separated into those that belong to P, R and S. From the result it can be seen, that P is block diagonal, S is block tridiagonal, and R is sparse with a block being non-zero, if the landmark corresponding to that column has been observed from the robot pose corresponding to that row (Fig. 2c):

$$P = \sum_{i=1,\, l\in\mathcal{L}_i}^{p} I_l^T \underbrace{(J_{li6}^T C_{li}^{-1} J_{li6})}_{P_{ll}^i :=} I_l, \tag{4}$$

$$R = \sum_{i=1,\, l\in\mathcal{L}_i}^{p} I_i^T (J_{li5}^T C_{li}^{-1} J_{li6}) I_l, \tag{5}$$

$$S = \sum_{i=1,\, l\in\mathcal{L}_i}^{p} I_i^T \underbrace{(J_{li5}^T C_{li}^{-1} J_{li5})}_{S_{ii}^l} I_i + \sum_{i=1}^{p} \begin{pmatrix} I_i^T (J_{oi3}^T C_{oi}^{-1} J_{oi3}) I_i + I_{i+1}^T (J_{oi4}^T C_{oi}^{-1} J_{oi3}) I_i \\ + I_i^T (J_{oi3}^T C_{oi}^{-1} J_{oi4}) I_{i+1} + I_{i+1}^T (J_{oi4}^T C_{oi}^{-1} J_{oi4}) I_{i+1} \end{pmatrix} \tag{6}$$

In the following discussion the block diagonals of P and S will be of great importance, so an explicit formula for the diagonal block S_{ii} corresponding to robot pose i and P_{ll} corresponding to landmark l is derived. This is performed by grouping (4) and (6) by values of l and i respectively. The result for P_{ll} is a sum over $\mathcal{O}_l$, the set of robot poses from which landmark l has been observed. Similarly, the result for S_{ii} is a term from odometry plus a sum over $\mathcal{L}_i$ the set of landmarks observed from robot pose i:

$$\begin{aligned} P_{ll} &= I_l P I_l^T = \sum_{i\in\mathcal{O}_l} (J_{li6}^T C_{li}^{-1} J_{li6}) = \sum_{i\in\mathcal{O}_l} P_{ll}^i \quad (7)\\ S_{ii} &= I_i S I_i^T = J_{oi3}^T C_{oi}^{-1} J_{oi3} + J_{o(i-1)2}^T C_{o(i-1)}^{-1} J_{o(i-1)2} \\ &\quad + \sum_{l\in\mathcal{L}_i} J_{li5}^T C_{li}^{-1} J_{li5} \\ &= \underbrace{J_{oi3}^T C_{oi}^{-1} J_{oi3} + J_{o(i-1)2}^T C_{o(i-1)}^{-1} J_{o(i-1)2}}_{S_{ii}^{\mathcal{O}} :=} + \underbrace{\sum_{l\in\mathcal{L}_i} S_{ii}^l}_{S_{ii}^{\mathcal{L}} :=} . \quad (8) \end{aligned}$$

It can be observed that one part ($S_{ii}^{\mathcal{O}}$) of S_{ii} originates from odometry measurements and another part ($S_{ii}^{\mathcal{L}}$) originates from landmark observations. In the latter part matrices S_{ii}^l from all landmark observations made from that robot pose accumulate. Nevertheless, S_{ii} and $S_{ii}^{\mathcal{L}}$ are bounded, since the number of landmark observations from a certain robot pose depends on the sensor / landmark trait and will not grow when the map gets larger. As for the diagonal block P_{ll} corresponding to landmark l this is different. Here matrices from all observations of this landmark accumulate. Since the same landmark may be observed over and over again, P_{ll} will usually grow linear with time. Each block R_{il} of R is affected only by a single measurement of landmark l from robot pose i. So it is bounded and 0, if the landmark has not been observed from there. Table I gives an overview of the different parts defined above.

V. Schur Complement

After eliminating (old) robot poses by Schur complement the resulting information matrix for the landmarks alone is $P' = P - R^T S^{-1} R$. and the inverse of the corresponding covariance matrix maintained by EKF. It is not sparse, since S^{-1} is dense. This section will prove that that an entry $P'_{l_1 l_2}$ decays exponentially with the distance $d_{l_1 l_2}$ between observation of landmarks l_1 and l_2. It is defined as

$$d_{l_1 l_2} := \min\{|i-j| \mid i \in \mathcal{O}_{l_1}, j \in \mathcal{O}_{l_2}\}, \tag{9}$$

the number of robot movements[3] between observation of l_1 and l_2. Let $l_1 \neq l_2$ and consider the corresponding block

$$P'_{l_1 l_2} = -\sum_{i\in\mathcal{O}_{l_1}} \sum_{j\in\mathcal{O}_{l_2}} R_{il_1}^T (S^{-1})_{ij} R_{jl_2}. \tag{10}$$

This equation is of high importance, because since R_{il_1} and R_{jl_2} are bounded, asymptotically $P'_{l_1 l_2}$ behaves like block $(S^{-1})_{ij}$ of S^{-1} corresponding to the robot from where l_1 and l_2 have been observed.

VI. Exponential Decay of Off-Diagonal Entries

This subsection proves that an off-diagonal entry $(S^{-1})_{ij}$ decays exponentially with the distance $|i-j|$ to the diagonal. The result is then used to derive that $P'_{l_1 l_2}$ decays exponentially with the distance $d_{l_1 l_2}$ between observation of l_1 and l_2. Thereby the approximate sparsity of the SLAM information matrix P' is proven. The rate of decay depends on the ratio between $S_{ii}^{\mathcal{O}}$ and $S_{ii}^{\mathcal{L}}$, i.e. between the information gained from odometry and landmark observations:

Definition 1: For a sequence of odometry and landmark observations the characteristic parameters are:

$$\omega := \max\left\{\omega | S_{ii}^{\mathcal{L}} \geq \omega S_{ii}^{\mathcal{O}} \; \forall i\right\} \tag{11}$$

$$\eta := \max_{i,l} \|P_{ll}^i\| \tag{12}$$

$$\rho := \min_{i,\, l\in\mathcal{L}_i} \|P_{ll}^i - R_{il}^T (S_{ii}^{\mathcal{L}})^{-1} R_{il}\| \tag{13}$$

Intuitively, this definition means: a) In each robot pose the information gained from landmarks is at least ω times the information transported from the last pose by odometry. b) The information gained from a single landmark measurement assumed the robot pose was known is at most η. c) The information gained about a landmark from all observations from a certain unknown robot pose assumed that all other landmarks were known is a least ρ.

Nevertheless, b) and c) need some explanation: P_{ll}^i is a submatrix of the information matrix for a single landmark observation of l from pose i (Fig. 2b). Thus, as discussed before, it represents the information known about landmark l from that measurement if all other random variables, in this case the robot pose, were known. Similarly, $\begin{pmatrix} P_{ll}^i & R_{il} \\ R_{il}^T & S_{ii}^{\mathcal{L}} \end{pmatrix}$ is a submatrix of the information matrix of all landmark observations from pose i. Therefore it represents the information that were known about landmark l and robot pose i if all other random variables, in this case the other landmarks, were known. Thus, the Schur complement

$$P_{ll}^i - R_{il}^T (S_{ii}^{\mathcal{L}})^{-1} R_{il} \geq 0 \tag{14}$$

represents the same information without information about the robot pose. So in the end the term describes the

[3]We assume that the robot observes a landmark at regular intervals. Indeed, if the robot moves blindly between two landmarks of arbitrary long distance the coupling entry decays only reciprocal not exponentially.

TABLE I

SYMBOLS USED IN THE PROOF OF THEOREM 2.

Symbol	Equation	Format	Definition
P_{ll}	(7)	2×2	Diagonal block of A corresponding to landmark l
P_{ll}^i	(4)	2×2	Contribution of the observation of landmark l from pose i to P_{ll}
R_{il}	(4)	3×2	Block of R corresponding to landmark l and robot pose i defined from the observation of landmark l from pose i
S_{ii}	(8)	3×3	Diagonal block of S corresponding to robot pose i
S_{ii}^l	(6)	3×3	Contribution of the observation of landmark l from pose i to S_{ii}
$S_{ii}^{\mathcal{L}}$	(8)	3×3	Contribution of all landmark observations from pose i to S_{ii}
$S_{ii}^{\mathcal{O}}$	(8)	3×3	Contribution of both odometry observations from pose i and $i-1$ to S_{ii}
$P'_{l_1 l_2}$	(10)	2×2	Block corresponding to landmark l_1 and l_2 of the information matrix P' of all landmarks without robot poses.
$\mathcal{L}_i$			Landmarks observed from robot pose i
$\mathcal{O}_l$			Robot poses from which landmark l has been observed
$d_{l_1 l_2}$	(9)		Number of movements between observation of landmarks l_1 and l_2

information, if all other landmarks were known but the robot pose was unknown. The parameter ρ gives a lower bound on this information.

All three parameters ω, η, ρ depend on the sensor / landmark / environment characteristic and do not change when the map size grows. So they may be considered as being constant $\omega = O(1)$, $\eta = O(1)$, $\rho = O(1)$.

The central argument of the overall proof uses a theorem by Demko, Moss and Smith [13, theorem 2.4] that provides an exponentially decaying bound for the entries of the inverse of a symmetric positive definite (SPD) band matrix S. The bounds refer to a single entry of S denoted by $S_{\#ij} \in \mathbb{R}$ to avoid confusion with the 3×3 matrix block $S_{ij} \in \mathbb{R}^{3\times 3}$ corresponding to robot pose i and j. The bound depends on the norm $\|S\|$ and condition number $\operatorname{cond}(S)$ of S. The matrix norms refer to the usual spectral or 2-norm $\|S\| := \max_{|v|=1} |Sv|$ being equal to the largest eigenvalue of S (largest singular value for a non-symmetric matrix). The derived condition number is equal to the ratio between largest and smallest eigenvalue.

Theorem 1 (Demko, Moth, Smith [13]): Let S be an SPD w-banded matrix. Then for entry $(S^{-1})_{\#ij}$ of S^{-1}

$$|(S^{-1})_{\#ij}| \le \alpha \lambda^{|i-j|}, \text{ with} \tag{15}$$

$$\lambda := \left(\frac{\sqrt{\operatorname{cond}(S)} - 1}{\sqrt{\operatorname{cond}(S)} + 1} \right)^{\frac{2}{w}} \quad \text{and } \alpha \le 2\|S^{-1}\|. \tag{16}$$

Some technical lemmas are needed. Due to lack of space, their proof is omitted referring the reader to [8].

Lemma 1: Let S be a block diagonal SPD matrix with block bandwidth w. Then the norm $\|S\|$ is at most $2w-1$ times the norm of any diagonal block ($= \max_i \|S_{ii}\|$).

Lemma 2: For all $\omega \ge 0$ the following inequality holds:

$$\left(\sqrt{1+3/\omega}+1\right)\left(\sqrt{1+3/\omega}-1\right)^{-1} \ge 1 + \tfrac{4}{3}\omega \tag{17}$$

From theorem 1, lemma 1, and 2 follows:

Lemma 3: Let S be a block tridiagonal SPD matrix with 3×3 blocks, smallest eigenvalue $\lambda_{\min} \ge 1$ and largest eigenvalue $\lambda_{\max} \le 1 + \frac{3}{\omega}$. Then for $i \ne j$ the norm of block $(S^{-1})_{ij}$ of the inverse is at most $\|(S^{-1})_{ij}\| \le 6\left(1 + \frac{4}{3}\omega\right)^{1-|i-j|}$.

The next step is to derive an exponentially decaying bound for the off-diagonal entries of S^{-1}. For technical reasons in the proof of theorem 2 the matrix to be considered is $R^T S^{-1} R$ not S^{-1}. So instead of deriving a bound for $(S^{-1})_{ij}$, directly a bound for $R_{il_1}^T (S^{-1})_{ij} R_{jl_2}$ is given.

Lemma 4: For a sequence of odometry and landmark observations with parameter ω, η, ρ, and all robot poses i, j and all landmarks $l_1 \ne l_2$ the following bound holds:

$$\|R_{l_1 i}^T (S^{-1})_{ij} R_{jl_2}\| \le 6\eta \left(1 + \frac{4}{3}\omega\right)^{1-|i-j|}. \tag{18}$$

Proof: Let $S^{\mathcal{L}} := \operatorname{diag}_i(S_{ii}^{\mathcal{L}})$ be the part of S that originates from the landmark measurements and $S^{\mathcal{O}} := S - S^{\mathcal{L}}$ be the remaining part originating from odometry. The block diagonal of $S^{\mathcal{O}}$ is $\operatorname{diag}_i(S_{ii}^{\mathcal{O}})$, but $S^{\mathcal{O}}$ itself is block tridiagonal (compare Fig. 2).

The first step is to scale $S^{\mathcal{O}}$, so matrices of comparable norm appear on the block diagonal. Let therefore $L_i L_i^T = S_{ii}^{\mathcal{L}}$ be a Cholesky decomposition of $S_{ii}^{\mathcal{L}}$ and define the inverse of $L := \operatorname{diag}_i(L_i)$ as scale matrix. This way the scaled matrix $L^{-1} S^{\mathcal{L}} L^{-1T}$ is the identity matrix and the scaled matrix $L^{-1} S^{\mathcal{O}} L^{-1T}$ is normalized relative to $S^{\mathcal{L}}$. So it is possible to bound it by ω:

Matrix $S^{\mathcal{O}}$ is block tridiagonal and L is block diagonal. Thus, $L^{-1} S^{\mathcal{O}} L^{-1T}$ is block tridiagonal, too. Further $S_{ii}^{\mathcal{L}} \ge \omega S_{ii}^{\mathcal{O}}$ by definition 1. So for each diagonal block it follows

$$(L^{-1} S^{\mathcal{O}} L^{-1T})_{ii} = L_i^{-1} S_{ii}^{\mathcal{O}} L_i^{-1T} \overset{\text{definition 1}}{\le} \tag{19}$$
$$\frac{1}{\omega} L_i^{-1} S_{ii}^{\mathcal{L}} L_i^{-1T} = \frac{1}{\omega} L_i^{-1} L_i L_i^T L_i^{-1T} = \frac{1}{\omega} I.$$

The matrix $L^{-1}S^{\mathcal{O}}L^{-1T}$ is tridiagonal, so lemma 1 can be applied with $w = 2$ and the eigenvalue $\lambda_{\max}\left(L^{-1}S^{\mathcal{O}}L^{-1T}\right)$ is at most $\frac{3}{\omega}$. By construction $L^{-1}S^{\mathcal{L}}L^{-1T} = I$, so the eigenvalues of $L^{-1}SL^{-1T} = L^{-1}(S^{\mathcal{L}} + S^{\mathcal{O}})L^{-1T}$ lie in the interval $\left[1 \dots 1 + \frac{3}{\omega}\right]$. It follows from lemma 3 that

$$\|L_i^T(S^{-1})_{ij}L_j\| = \|((L^{-1}SL^{-1T})^{-1})_{ij}\| \overset{\text{lemma 3}}{\leq} 6\left(1 + \tfrac{4}{3}\omega\right)^{1-|i-j|} \quad (20)$$

$$\begin{aligned}\|R_{il_1}^T(S^{-1})_{ij}R_{jl_2}\| &= \|R_{il_1}^T L_i^{-1T} L_i^T(S^{-1})_{ij}L_j L_j^{-1}R_{jl_2}\| \\ &\leq \|R_{il_1}^T L_i^{-1T}\| \, \|L_i^T(S^{-1})_{ij}L_j\| \, \|L_j^{-1}R_{jl_2}\| \\ &\leq \|R_{il_1}^T L_i^{-1T}\| \, 6\left(1 + \tfrac{4}{3}\omega\right)^{1-|i-j|} \, \|L_j^{-1}R_{jl_2}\|. \quad (21)\end{aligned}$$

The next step is a bound for $\|L_j^{-1}R_{jl_2}\|$ and $\|R_{il_1}^T L_i^{-1T}\|$:

$$\begin{aligned}\|L_j^{-1}R_{jl_2}\|^2 &= \|(L_j^{-1}R_{jl_2})^T(L_j^{-1}R_{jl_2})\| \\ &= \|R_{jl_2}^T L_j^{-1T} L_j^{-1} R_{jl_2}\| \\ &= \|R_{jl_2}^T(S_{jj}^{\mathcal{L}})^{-1}R_{jl_2}\| \overset{(14)}{\leq} \|P_{l_2l_2}^j\| \overset{\text{definition 1}}{\leq} \eta. \quad (22)\end{aligned}$$

It follows $\|L_j^{-1}R_{jl_2}\| \leq \sqrt{\eta}$, which is substituted into (21):

$$\|R_{il_1}^T(S^{-1})_{ij}R_{jl_2}\| \leq 6\eta\left(1 + \tfrac{4}{3}\omega\right)^{1-|i-j|} \quad (23)$$

■

To prove approximate sparsity of P', the norm of its off-diagonal blocks $P'_{l_1l_2}$ must be bounded relative to the corresponding diagonal blocks $\|P'_{l_1l_1}\|$ and $\|P'_{l_2l_2}\|$. Therefore, a lower bound for a diagonal block $\|P'_{ll}\|$ is derived in the following lemma:

Lemma 5: For a sequence of observations with parameter ω, η, ρ, and all landmarks l it holds that $\|P'_{ll}\| \geq \rho\,|\mathcal{O}_l|$.

Proof: Since $S \geq S^{\mathcal{L}}$ it follows $S^{-1} \leq (S^{\mathcal{L}})^{-1}$ and

$$P' = P - R^T S^{-1} R \geq P - R^T (S^{\mathcal{L}})^{-1} R \quad (24)$$

$$\Rightarrow \quad P'_{ll} \geq P_{ll} - \sum_{i,j \in \mathcal{O}_l} R_{il}^T((S^{\mathcal{L}})^{-1})_{ij}R_{jl}. \quad (25)$$

$S^{\mathcal{L}}$ is block diagonal, so $(S^{\mathcal{L}})^{-1} = \operatorname{diag}_i((S_{ii}^{\mathcal{L}})^{-1})$ and $((S^{\mathcal{L}})^{-1})_{ij} = 0$ for $i \neq j$

$$\begin{aligned}P'_{ll} &= P_{ll} - \sum_{i \in \mathcal{O}_l} R_{il}^T(S_{ii}^{\mathcal{L}})^{-1}R_{il} \quad (26) \\ &= \sum_{i \in \mathcal{O}_l} P_{ll}^i - R_{il}^T(S_{ii}^{\mathcal{L}})^{-1}R_{il} \overset{\text{definition 1}}{\geq} \rho\,|\mathcal{O}_l|. \quad (27)\end{aligned}$$

■

The final step is to substitute the bound of lemma 4 into (10) to derive an overall bound. A sum of different powers of $(1 + \frac{4}{3}\omega)^{-1}$ appears. The exact exponents depend on the robot poses the landmark has been observed from, so it is difficult to find a closed expression. However, all exponents are at least $d_{l_1l_2}$. Using this property the sum can be bounded by the following lemma (proof in [8]):

Lemma 6: Let $0 \leq \gamma < 1$ and $\mathcal{A}, \mathcal{B} \subset \mathbb{N}$ with a minimal distance d between elements of $\mathcal{A}$ and $\mathcal{B}$: $d \leq |i - j| \; \forall i \in \mathcal{A}, j \in \mathcal{B}$. Then the following inequality holds:

$$\sum_{i \in \mathcal{A},\, j \in \mathcal{B}} \gamma^{|i-j|} \leq 2\frac{\gamma^d}{1-\gamma}\min\{|\mathcal{A}|, |\mathcal{B}|\} \quad (28)$$

Theorem 2 (Information Matrix Sparsity): Consider a sequence of odometry and landmark observations with parameter ω, η, ρ. Then the resulting SLAM information matrix of all landmarks P' is approximately sparse. The off-diagonal block $P'_{l_1l_2}$ corresponding to two landmarks $l_1 \neq l_2$ decays exponentially with the smallest number of steps $d_{l_1l_2}$ traveled between observation of l_1 and l_2.

$$\frac{\|P'_{l_1l_2}\|}{\min\left\{\|P'_{l_1l_1}\|, \|P'_{l_2l_2}\|\right\}} = O\left(\left(1 + \frac{4}{3}\omega\right)^{-d_{l_1l_2}}\right)$$

Proof: By equation (10) $P'_{l_1l_2}$ is a sum over different robot poses i, j. Each term can be bounded by lemma 4:

$$\begin{aligned}\|P'_{l_1l_2}\| &\overset{(10)}{=} \left\| - \sum_{i \in \mathcal{O}_{l_1}} \sum_{j \in \mathcal{O}_{l_2}} R_{il_1}^T(S^{-1})_{ij}R_{jl_2} \right\| \quad (29) \\ &\leq \sum_{i \in \mathcal{O}_{l_1}} \sum_{j \in \mathcal{O}_{l_2}} \left\|R_{il_1}^T(S^{-1})_{ij}R_{jl_2}\right\| \\ &\overset{\text{lemma 4}}{\leq} 6\eta \sum_{i \in \mathcal{O}_{l_1}} \sum_{j \in \mathcal{O}_{l_2}} \left(1 + \frac{4}{3}\omega\right)^{1-|i-j|} \\ &= 6\eta\left(1 + \frac{4}{3}\omega\right) \sum_{i \in \mathcal{O}_{l_1}} \sum_{j \in \mathcal{O}_{l_2}} \left(1 + \frac{4}{3}\omega\right)^{-|i-j|}.\end{aligned}$$

The sum can be bounded by lemma 6 with $\mathcal{A} := \mathcal{O}_{l_1}, \mathcal{B} := \mathcal{O}_{l_2}, \gamma := \left(1 + \frac{4}{3}\omega\right)^{-1}$ and $d := d_{l_1l_2}$:

$$\begin{aligned}\|P'_{l_1l_2}\| &\leq 6\eta\left(1 + \frac{4}{3}\omega\right)\left(\frac{2\min\{|\mathcal{O}_{l_1}|, |\mathcal{O}_{l_2}|\}}{1-\gamma}\gamma^{d_{l_1l_2}}\right) \\ &= 12\eta\left(1 + \frac{4}{3}\omega\right)\left(1 + \frac{3}{4\omega}\right)\min\{|\mathcal{O}_{l_1}|, |\mathcal{O}_{l_2}|\}\gamma^{d_{l_1l_2}} \\ &= \eta\min\{|\mathcal{O}_{l_1}|, |\mathcal{O}_{l_2}|\}\left(24 + 16\omega + \frac{9}{\omega}\right)\gamma^{d_{l_1l_2}} \quad (30)\end{aligned}$$

By lemma 5 it holds that

$$\begin{aligned}\min\left\{\|P'_{l_1l_1}\|, \|P'_{l_1l_1}\|\right\} &\geq \rho\min\{|\mathcal{O}_{l_1}|, |\mathcal{O}_{l_2}|\} \quad (31) \\ \Rightarrow \frac{\|P'_{l_1l_2}\|}{\min\left\{\|P'_{l_1l_1}\|, \|P'_{l_1l_1}\|\right\}} &\leq \frac{\eta}{\rho}\left(24 + 16\omega + \frac{9}{\omega}\right)\gamma^{d_{l_1l_2}}\end{aligned}$$

Even with an asymptotically increasing map size, ω, η and ρ remain constant, since they depend on the quality of the measurements such as sensor noise, typical distance to landmarks, typical number of landmarks. They do not depend on how many measurements were made. So the final asymptotic formula is

$$\frac{\|P'_{l_1l_2}\|}{\min\left\{\|P'_{l_1l_1}\|, \|P'_{l_1l_1}\|\right\}} \leq O\left(\left(1 + \frac{4}{3}\omega\right)^{-d_{l_1l_2}}\right). \quad (32)$$

■

VII. LOCAL VS. GLOBAL UNCERTAINTY

Apart from providing foundation for sparse information matrix based algorithms the theorem also allows characterization of SLAM uncertainty structure. It can be observed that there is a qualitative difference between local and global structures of SLAM, i.e. between relations of neighboring and of distant landmarks. Roughly speaking, the local uncertainty is small but complex and depends on actual observations, whereas the global uncertainty is large, rather simple and dominated by the map's geometry (*"certainty of relations despite uncertainty of positions" [5], [9]*). This is a consequence of theorem 2 and will be clarified in the following:

The measurements themselves define independent relations between landmarks and robot poses. For most sensors the uncertainty depends on the distance (laser scanner, stereo vision) or is even infinite in one dimension (mono vision). The information provided by the set of landmark observations from a single robot pose contains a highly coupled uncertainty originating from the uncertainty of the robot pose. From successive robot poses usually similar but different sets of landmarks are observed. So the parts of the information corresponding to different robot poses are highly coupled, but are always coupling different sets of landmarks. As a result the overall information on a local scale is also highly coupled and very complex. This corresponds to the coupling entries $P'_{l_1 l_2}$ in the information matrix being high for landmarks l_1, l_2 that are near to each other.

On a global scale the structure of the information is governed by theorem 2. The coupling entry $P'_{l_1 l_2}$ between distant landmarks is very low. So the uncertainty of the relation between them is approximately the composition of local uncertainties along the path from l_1 to l_2:

Consider the information matrix resulting from the integration of several local bits of information, for instance, the distance of each landmark to any other landmark nearby. This matrix is the sum of the information matrices for each bit of information. Each of them has non-zero coupling entries only for the landmarks involved. So the overall information matrix is sparse with all coupling entries being zero, except those of adjacent landmarks.

Thus, as local information corresponds to a sparse information matrix, an approximately sparse information matrix corresponds to information that can approximately be viewed as being the integration of local information.

If all measurements are uncertain, the global effect is approximately the sum of an uncertain rotation for each local region. The resulting uncertainty structure can best be described as an *uncertain bending* of the map [9], [5]. Compared to local uncertainty it is much larger, but simpler because the maps geometry is dominating it's structure.

The main target of SLAM is modeling global uncertainty. But often representation of local uncertainty is necessary to support landmark identification or allow task planning based on objects represented in the map. The approach of using sparse information matrices ideally suits the SLAM uncertainty structure: Local uncertainty is precisely captured in the small matrix representing a local region, whereas the global uncertainty structure does not need to be represented explicitly, since it is in very good approximation nothing more than the composition of local uncertainties.

VIII. CONCLUSION

In this paper it has been proven that the SLAM information matrix is approximately sparse, i.e. that an off-diagonal entry corresponding to two landmarks decays exponentially with the distance traveled between observation of those two landmarks. The theorem ensures that it is possible to approximate the SLAM information matrix by a sparse matrix by neglecting or conservatively eliminating off-diagonal entries of distant landmarks. It also indicates that if odometry becomes more imprecise compared to landmark observation the approximation error will become smaller and more aggressive sparsification is possible.

Apart from it's algorithmic implications the theorem also highlights the structure of SLAM uncertainty verifying from a formal perspective the intuitive characterization as "certainty of relations despite uncertainty of positions".

REFERENCES

[1] F. Lu and E. Milios, "Globally consistent range scan alignment for environment mapping," *Autonomous Robots*, vol. 4, pp. 333 – 349, 1997.

[2] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery, *Numerical Recipes, Second Edition*. Cambridge University Press, Cambridge, 1992.

[3] T. Duckett, S. Marsland, and J. Shapiro, "Fast, on-line learning of globally consistent maps," *Autonomous Robots*, vol. 12, no. 3, pp. 287 – 300, 2002.

[4] U. Frese, P. Larsson, and T. Duckett, "A multigrid algorithm for simultaneous localization and mapping," *IEEE Transactions on Robotics*, 2004, (to appear).

[5] U. Frese and G. Hirzinger, "Simultaneous localization and mapping - a discussion," in *Proceedings of the IJCAI Workshop on Reasoning with Uncertainty in Robotics, Seattle*, Aug. 2001, pp. 17 – 26.

[6] S. Thrun, D. Koller, Z. Ghahmarani, and H. Durrant-Whyte, "SLAM updates require constant time," School of Computer Science, Carnegie Mellon University, Pittsburgh, Tech. Rep., 2002.

[7] M. Paskin, "Thin junction tree filters for simultaneous localization and mapping," in *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, San Francisco, 2003, pp. 1157–1164.

[8] U. Frese, "An $O(\log n)$ algorithm for simulateneous localization and mapping of mobile robots in indoor environments," Ph.D. dissertation, University of Erlangen-Nürnberg, 2004.

[9] ——, "A discussion of simultaneous localization and mapping," *Autonomous Robots*, 2004, (submitted).

[10] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," in *Autonomous Robot Vehicles*, I. Cox and G. Wilfong, Eds. Springer Verlag, New York, 1988, pp. 167 – 193.

[11] W. Hager, "Updating the inverse of a matrix," *SIAM Review*, vol. 31, no. 2, pp. 221 – 239, June 1989.

[12] G. Tee, "Bounds for eigenvalues of positive-definite band matrices," *Australian Mathematical Society Gazette*, vol. 27, pp. 155 – 157, 2000.

[13] S. Demko, W. Moss, and P. Smith, "Deacy rates for inverses of band matrices," *Mathematics of Computation*, vol. 43, no. 168, pp. 491 – 499, Oct. 1984.

ERRATA

In figure 2 J_{li5} should be J_{li6} and vice versa. In (8) $J_{o(i-1)2}$ should be $J_{o(i-1)4}$. Thanks to S. Huang for pointing out, that the proof does not include information on the initial robot pose appearing on the left-upper 3×3 block of S. The corresponding term as well as any other global position information can be included in $S_{ii}^{\mathcal{L}}$ with the arguments of the proof carrying over. Note, that theorem 1, lemma 1 and lemma 3 are applied to $L^{-1}SL^{-1T}$ not S in (20) and the eigenvalues of that matrix lie in the interval $\left[1 \ldots 1 + \frac{3}{\omega}\right]$.