Simultaneous Localization and Mapping - A Discussion

U. Frese and **G. Hirzinger**
Deutsches Zentrum für Luft- und Raumfahrt (DLR)
Institut für Robotik und Mechatronik
Email: *Udo.Frese@dlr.de* *www.robotic.dlr.de*

Abstract

This papers provides two contributions to the problem of Simultaneous Localization and Mapping (SLAM): First we discuss properties of the problem itself and of the intended semantics of an uncertain map representation, with the main idea of "representing certainty of relations despite the uncertainty of positions". We propose some requirements an ideal solution of SLAM should have concerning uncertainty, memory space and computation time and discuss existing approaches in the light of these requirements. The second part proposes a representation based on sparse information matrices together with some properties that motivate this approach. This is shown to comply to the uncertainty and space requirements. To derive an estimated map from the representation a sparse linear equation system has to be solved. However, an update of the representation itself needs only constant time, making it highly attractive for building a SLAM algorithm.

1 Introduction

Navigation is the "*science of getting ships, aircraft, or spacecraft from place to place*" [Merriam-Webster's Collegiate Dictionary]. It is also the science of getting mobile robots from place to place, a problem that is central to mobile robotics and has been subject to extensive research. In a very general sense, every approach has to use some kind of map first to plan a path and then to localize the moving robot by comparing the perceived environment to the map.

Often a manual a-priori map is used, which is however difficult and expensive to obtain, making approaches favorable, that can build their map autonomously. For instance Thrun et al. [16] report, that obtaining a map for a robotic tour guide installation manually took one week, whereas building a map autonomously took about one hour.

In this paper we discuss the problem of Simultaneous Localization and Mapping (SLAM), which is an online version of the mapping problem. It involves building, extending and improving a map of the environment while the robot is moving and simultaneously localizing the robot with respect to the map.

The paper consists of two parts. The first and larger part attempts to "*understand the intended semantics*" of an uncertain map representation as suggested by the workshop's main focus. The key idea is the need to represent "*certainty of relations despite uncertainty of positions*" (section 2). We try to address the problem by taking an intentionally naive view, blinding out its known difficulty. This leads us to propose some requirements, concerning uncertainty, memory space and computation time, which an ideal solution should meet (section 3). We look at existing approaches in the light of these requirements (section 4).

The second part (section 5) proposes a representation for a map that uses information matrices to represent uncertainty. We identify some important properties of SLAM which have a counterpart in the structure of the information matrix, thereby motivating this approach.

The representation meets the uncertainty and memory requirements stated in section 3. Updating the representation takes constant time. This makes it highly attractive to build a SLAM algorithm based on this representation.

2 General Discussion

In this paper, we consider SLAM based on point-shaped landmarks, which are selected features of the environment. Most parts of the discussion however are valid even for non-landmark based approaches, e.g. using evidence-grids for representing the environment (see subsection 4.2).

Throughout the paper we assume that the observed landmarks can be identified, i.e. if a landmark is observed a second time, it can be recognized to be the very same as previously observed (see subsection 2.8).

The term "map instance" refers to an assignment of coordinates to the landmarks, whereas "map" refers to an assignment of coordinates together with information about their possibly coupled uncertainty. A "representation" is a concrete data structure representing a map. A "relation" between a set of landmarks is some property like distance, angle, etc., that depends on the landmark's coordinates but is invariant to rigid body movements of the whole set. We use the scenario shown in figure 1a to illustrate key ideas.

2.1 Graph of Measurements

SLAM can be viewed as an estimation theoretic problem. The parameters to be estimated are the p different robot poses at

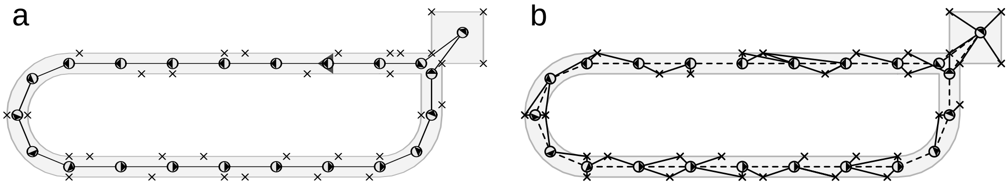

Figure 1: a) The robot starts at the position indicated by the triangle - without knowing the building. It moves through the long oval doorway observing many landmarks along the way. Then it enters the room observing four landmarks in the room's corners and finally moves back to the start position, re-identifying both of the nearby landmarks. b) Graph of measurements

various points of time and the positions of the n landmarks. The measurements are taken from odometry and landmark observations. A landmark observation yields the position of the landmark relative to the robot's pose at some point of time. The odometry defines the relative robot pose between two successive points of time. We assume to have an *a-priori* model of the uncertainty of each measurement.

An important property is, that every measurement involves only two objects, each having either 2 or 3 parameters. Thus it is natural to view the whole setup as a graph with the robot's poses and the landmark's positions being nodes and the measurements forming edges (figure 1b).

This representation gives a good intuition for the structure of the problem by a mechanical analogy: Imagine the graph as a truss build from bolts as nodes and elastic metal bars connecting the bolts as edges. Let the stiffness of a bar correspond to the certainty of a measurement. The bolts can be moved with respect to each other, where easy movement corresponds to uncertain relations and hard movement corresponds to relations precisely known. The odometric sequence can be viewed as a long thin metal spline, with the landmark observation bars connected to the spline [7].

2.2 Error Accumulation

If the robot moves through a known environment, i.e. by using an a-priori map, uncertainty of the robot's pose can be kept low, as each observation of a landmark reduces the uncertainty down to the landmark's uncertainty plus the uncertainty of the observation.

However if the robot moves through an unknown region, the uncertainty of its pose will get arbitrarily large, because the odometric error accumulates over time (figure 2a). The uncertainty can be reduced by fusing the odometry with several measurements of a new landmark as the landmark passes by (figure 2b). For most sensors this produces much better results than using odometry alone [16]. Nevertheless, estimating the robot's position after traveling a long distance is still subject to accumulated error: due to the limited sensor range the position is derived from a chain of several relative landmark relations.

For outdoor applications the problem can be relieved by using a compass [11], which is however known not to work properly in buildings which contain large amount of steel.

The fact that errors may accumulate to arbitrarily high values distinguishes SLAM from many other estimation problems and gives rise to the problems discussed in sections 2.3 and 2.6.

2.3 Representation of Relativity

We believe that the dominant aspect of SLAM is the need to model *"Certainty of Relations despite Uncertainty of Positions"*, which we call "representing relativity". In our scenario for instance, the pose of the room will be quite uncertain, while its shape will be very certain.

If the robot moves through a previously unknown region and observes a sequence of landmarks, the precision of the relative positions of the landmarks depends only on the measurement errors of the landmarks by the robot and on the odometric error between those measurements. So the most precisely known relations are those concerning the relative location of adjacent landmarks.

The uncertainty of the absolute robot pose before observing the first landmark however increases the uncertainty of the absolute position of all landmarks, acting as an unknown rigid body transformation on the whole set of observed landmarks. As the absolute robot pose is subject to error accumulation, the common situation is that relations are quite certain, whereas absolute positions can be arbitrarily uncertain. In very large maps this effect can appear at different scales: the relative positions of some landmarks in a room are much more precisely known than the position of the room in the building, which, seen as a relative position with respect to other rooms is in turn much more precisely known than the absolute position of the building.

Thus a SLAM system should be able to represent the certainty of relations between landmarks despite large uncertainty in the absolute position of the landmarks. In particular, a representation where only a single uncertainty value is assigned to each landmark, is insufficient.

Although it is theoretically possible to approach absolute precision by repeating all measurements often enough (with the robot's initial pose defined to be perfectly known), this is in general neither practical nor necessary, as most applications can be exclusively based on relative information: When navigating for instance, it is not necessary to compute the exact trajectory from start to finish in some global coordinate system. Path planning will rather result in a sequence of waypoints. The location of each waypoint will be known relative to the surrounding landmarks, so that the robot, knowing its own pose relative to those landmarks, will be able to navigate from one waypoint to the next.

2.4 Closing Loops

Let us assume that the robot moves along a closed loop and returns to the begin of that loop, but has not yet re-identified any landmark, so the latter fact is not known to the robot.

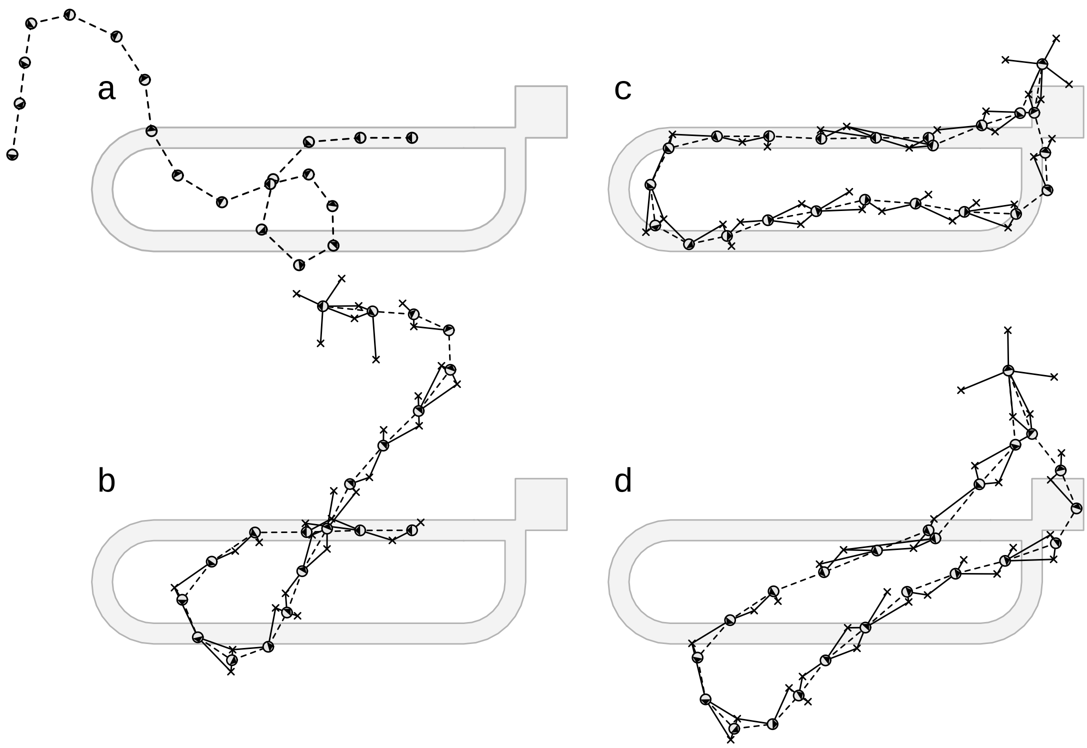

Figure 2: a) Odometric error b) Odometry fused with landmarks c) Loop closed by re-identifying two landmarks d) Closing the loop by EKF results in the room being too large.

Typically the loop is not closed in the map due to the error accumulated along the loop.

Now a landmark at the begin of the loop is re-identified and the corresponding measurement is integrated into the map causing the loop to get closed. To achieve this, the SLAM system has to ”deform” the whole loop to incorporate the information of a connection between both ends of the loop without introducing a break somewhere else (figure 2c).

This goal is sometimes referred to as the map being ”topologically consistent”, meaning that two parts of the map are represented to be adjacent if and only if this was observed by the robot. Within a landmark based approach adjacency is not explicitly modeled, so topological consistency has to be interpreted in the sense that two landmarks are represented being near to each other (the distance being low with low uncertainty), exactly if this was observed by some measurement.

It has to be stressed that correct treatment of the uncertainty contained in the measurements will implicitly yield the necessary deformation. More specifically the precisely known relative location of each landmark with respect to adjacent landmarks prevents any break in the loop. Because if there was a break, the relative positions of the landmarks on both sides of the break would be highly incorrect, thus being inconsistent with the measurements made in that vicinity. So the map estimate consistent with all measurements automatically deforms smoothly when closing large loops. To retain this property in some map which integrates all measurements instead of storing them individually, the representation must be able to “represent relativity”.

2.5 Maximum Likelihood Estimation

If we assume independent gaussian measurement errors with a-priori known covariance and do not care about computation time, SLAM can be solved in a thorough but straightforward fashion by least square nonlinear model fitting [14, chapter 15]. This is performed by finding the minimum $\hat{x}$ of the quadratic error function $Q(x)$:

$$q_i(x) = \frac{1}{2}(y_i - f_i(x))^T C_i^{-1}(y_i - f_i(x)), \tag{1}$$

$$Q(x) = \sum_i q_i(x), \quad \hat{x} := \arg\min_x Q(x), \tag{2}$$

$$\mathcal{C}_\alpha := \{x \| Q(x) - Q(\hat{x}) \leq \alpha\}, \tag{3}$$

with the landmark positions and different robot poses forming the parameter vector x, y_i being the i-th measurement, C_i its covariance and $f_i(x)$ being the corresponding measurement equation, i.e. the value the measurement should have had if the landmark and robot poses were x. Even though f_i is nonlinear, the minimum $\hat{x}$ of Q is a maximum likelihood estimator for the map instance and can be found by the Levenberg-Marquardt algorithm (figure 2a-c have been computed this way). The set $\mathcal{C}_\alpha$ surrounding $\hat{x}$ is a confidence region defining the map's uncertainty with α depending on the desired level of confidence.

However, this approach is not a practical solution for SLAM as it requires all measurements to be saved and an iteration performed with several linear equation systems to be solved each time a new measurement is added. With n landmarks and p robot poses, this takes $O((n+p)^3)$ computation time. Its invaluable benefit however lies in the fact that it can provide a reference for discussion and for comparison with efficient approaches.

When linearizing the measurement equations f_i, Q becomes quadratic in the parameter vector: $Q(x) = x^T Ax + x^T b + \gamma$. The matrix A is called information matrix. High entries in A correspond to precisely known relations (see subsection 2.7).

It is interesting to note that in the mechanical analogy explained in section 2.1 the function Q corresponds to the elastic energy stored in the system. Finding a maximum likelihood solution thus corresponds to minimizing the elastic energy, which is just the mechanical system's natural behavior. "Closing a loop" can therefore be interpreted literally, with the begin and end of the loop both connected to the bolt representing the re-observed landmark and the system smoothly deforming to achieve the state of minimal energy [7].

2.6 Extended Kalman Filter

The EKF is the tool most often applied to SLAM [15] using the same measurement equations as for maximum likelihood estimation. The EKF integrates all measurements into a covariance matrix of the landmark positions and the actual robot pose, without having to store any measurements afterwards. For linear measurement models it nevertheless yields the maximum likelihood estimate. For nonlinear measurement equations, which appear in SLAM due to the robots rotation, the maximality holds to the extent, that the equation can be adequately linearized. If after each new measurement *all* measurements could be linearized at the current estimate, the result would be equivalent to the nonlinear maximum likelihood estimate. With the EKF however, changing the point of linearization after integrating a measurement is impossible, so the measurements are usually linearized at the estimation in the moment of that measurement.

As a consequence the point of linearization can be significantly wrong when moving through an unmapped area, since the odometric error accumulates thereby. Especially the robot's orientation error can easily exceed $45°$ in practical settings rendering all linearizations of $\sin$ and $\cos$ useless.

The effect of processing the example scenario with an EKF instead of using maximum likelihood estimation is bad (figure 2d). Start and finish of the loop do not match and even worse, the room, although precisely known gets significantly larger than before. The reason for this is the following: The EKF would have to implicitly move and rotate the room to make the map consistent. Instead it performs this by a rotation linearizing the angle at 0:

$$R(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \quad (4)$$

$$R(\phi) \overset{\phi=0}{\longrightarrow} \begin{pmatrix} 1 & -\phi \\ \phi & 1 \end{pmatrix} = \sqrt{1+\phi^2} \cdot R(\arctan\phi) \quad (5)$$

The consequence is that the room is larger than before and rotated by a too small angle.

2.7 Covariance vs. Information Matrix

EKF and least square become equivalent, yielding the same estimate, when using measurement equations that are linearized at the estimate in the moment of measurement. Figure 3 shows the relation between the EKF's covariance matrix C and the information matrix A used in linearized least square, both representing the uncertainty of the estimate.

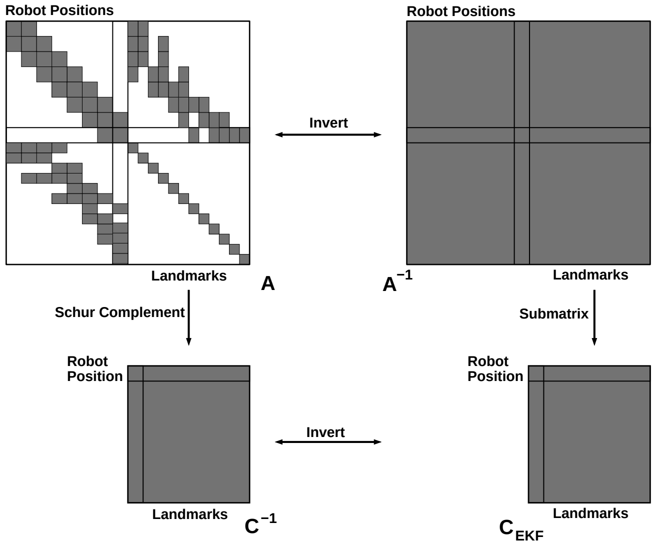

Figure 3: Relation between the least square information matrix A and the covariance matrix C used by the EKF.

Two stochastically independent information matrices can be combined by simply adding both. This corresponds to the fact, that the information matrix for a set of measurements is the sum of the information matrices for each measurement (2). Integrating two independent covariance matrices C_1, C_2 is done by computing $\left(C_1^{-1} + C_2^{-1}\right)^{-1}$. Indeed, the EKF can be seen as a clever way to add the information matrix corresponding to a single measurement to the inverse of the covariance matrix, where actually inverting the covariance matrix is avoided by application of the Woodbury formula for updating the inverse of a matrix [14, chapter 2.7].

The main difference between information and covariance matrix lies in the representation of indirect relations. Assume that 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 A is sparse, having non-zero off-diagonal entries only for those pairs of parameters which are involved in a common measurement. The inverse A^{-1} is the covariance matrix for the positions of the landmarks and *all* robot poses. A^{-1} represents all indirect relations explicitly and is thus not sparse. Removing the rows and columns corresponding to old robot poses yields the covariance matrix C of the EKF. Its inverse C^{-1} is the information matrix of all landmark positions and the actual 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 parameters by the Schur complement which destroys sparsity.

2.8 Landmark Identification

Throughout the paper we assume that the observed landmarks can be identified. This is a very difficult but highly important problem, since for instance, the event of closing a loop is only evident from the re-identification of some landmarks. There are some approaches that take advantage of tightly integrating mapping and identification [1]. Indeed the uncertainty information from the map can be used to support identification [13]. However it is often a good idea to separate both, as the SLAM problem can be formulated relatively independently of the sensors used, whereas landmark identification usually depends heavily on them.

2.9 Local vs. Global 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 the actual observation, whereas the global uncertainty is large, rather simple and dominated by the map's geometry. We will clarify this in the following:

The measurements themselves define independent relations between landmarks and robot poses. For some sensors (laser range finder) the same uncertainty can be assumed for each measurement. However for many sensors the uncertainty depends on the distance (stereo vision) or is even infinite in one dimension (mono vision).

The covariance matrix of a set of landmarks generated by measurements from a single robot pose consists of an independent uncertainty for each landmark described by a (block) diagonal matrix D_i plus a correlated rank 3 uncertainty $J_i C_i J_i^T$ with J_i being the measurement Jacobian with respect to the robot pose and C_i the covariance of the robot pose. The latter is generated by the uncertain rigid body transform on the set of landmarks that originates from the uncertain robot pose. If a set of landmarks is observed from several robot poses the resulting matrix C_{total} can be expressed using the Woodbury formula:

$$C_{\mathrm{total}}^{-1} = \sum_i (D_i + J_i C_i J_i^T)^{-1} \quad (6)$$

$$= \sum_i D_i^{-1} - \sum_i (D_i^{-1} J_i)(C_i + J_i^T D_i^{-1} J_i)^{-1} (D_i^{-1} J_i)^T \quad (7)$$

The term in the second sum is a symmetric positive definite rank 3 matrix with row- / column-space equal to the column space of $D_i^{-1} J_i$. As in general the D_i and J_i are different, the overall sum can have up to full rank. This corresponds to quite complex correlations between the different landmarks that heavily depend on the uncertainties of the different measurements and on whether a landmark is observed from a certain robot pose. The situation is similar if the different robot poses are related by uncertain odometric measurements.

To see the uncertainty structure of global relations, we assign suitably spaced reference frames to the whole map and make the plausible assumption, that the effect of the local uncertainty structure can be approximated by an uncertain relation between adjacent reference frames (figure 4). Let us

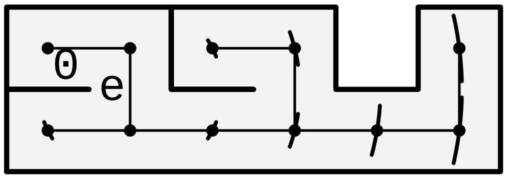

Figure 4: Global uncertainty generated by the uncertainty in edge e with node 0 defined to be the coordinate origin.

first consider a map containing no loop with an arbitrary reference frame defined to be the origin. The resulting graph of relations will be a tree, i.e. containing no circles. Each edge e divides the graph into two parts, where the uncertainty of the edge affects only that part that does not contain the origin. If all relations except e were exact, the dominant error would be an uncertain rotation around e combined with an uncertain but much smaller translation. This effect introduces high correlations both between different frames and between the frames position and orientation.

The total uncertainty is roughly the sum of these effects for all edges e in the map. The general case of a map containing loops can be understood by removing one edge of each resulting circles in the graph. The uncertainty of the original graph equals the uncertainty of the resulting graph constrained by the additional information of the removed edges. We can conclude, that the global uncertainty is dominated by the map's geometry and is much larger than the complex local uncertainty. This conclusion justifies our original assumption about the approximation of local uncertainty.

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.

3 Requirements for an Ideal Solution

In this section we postulate some properties a SLAM solution should have. They are based on an intentionally naive view of the problem, blinding out its apparent difficulty, but asking how mapping should work based on a common sense understanding of maps. The intention is both to clarify the discussion of existing approaches in section 4 and to make the motivation for our research more explicit.

> **R1: Bounded Uncertainty**: *The uncertainty with which any relation is represented in the map should not be much larger than the minimal uncertainty that could be theoretically derived from the measurements.*

This postulate is quite general, saying that if something can be known from the measurements it should at least roughly be represented in the map. Consistently approximating some relations for the sake of efficiency is acceptable to the extend, to which relations get slightly less precise, but not if loosing all or almost all information about certain relations. This includes the ability to represent relativity, as many relations can be known precisely from the measurements, so not representing one would violate the principle stated. As explained above, representing relativity implies to be able to close large loops achieving topologically consistent maps.

> **R2: Linear Memory Size**: *The memory size of a map that covers a large area should be linear in the number of landmarks ($O(n)$).*

The soundness of this postulate can be seen from the following example: Imagine a building consisting of two parts, $\mathcal{X}$ and $\mathcal{Y}$, being connected only by a few doorways. Then the map of the whole building consists of the map of both parts plus some information concerning the connections and should thus have a size only slightly larger than the size of a map of $\mathcal{X}$ plus the size of a map of $\mathcal{Y}$.

It is worth noting that simply storing all measurements will not meet (R2), since the memory size is proportional to the number of measurements not to the number of landmarks. Thus the map's size would grow even if repeatedly moving through the same area.

> **R3: Linear Update Cost**: *Incorporating a measurement into a map covering a large area should have a computational cost at most linear in the number of landmarks ($O(n)$).*

This postulate is more difficult to justify than the preceeding ones: Let us assume that the same setting as above holds, with a measurement made in $\mathcal{X}$. At first the measurement has to be incorporated into the map of $\mathcal{X}$, taking the known effect of $\mathcal{Y}$ on the connection between $\mathcal{X}$ and $\mathcal{Y}$ into account. Then the effect of these connections onto $\mathcal{Y}$ must be computed. This is equivalent to incorporating *several* measurements concerning the connections into the map of $\mathcal{Y}$. However, the computation can be deferred until the robot actually enters $\mathcal{Y}$, sharing the computational cost with all other measurements that generate effects on the connections and have to be integrated until then. As the number of landmarks in the connections is small, this should take not more time per measurement than incorporating the original measurement into the map of $\mathcal{X}$, at least for large maps.

So the total cost for integrating a measurement into a map containing $\mathcal{X}$ and $\mathcal{Y}$ should not be larger than the cost of integrating it into $\mathcal{X}$ plus the cost for integrating it into $\mathcal{Y}$, thus being linear in the number of landmarks.

(R1) states that the map shall represent nearly all information contained in the measurements, thus binding the map to reality. The other postulates regard efficiency, requiring linear space and time consumption: All asymptotic statements have to be interpreted with respect to a map not only consisting of an increasing number of landmarks but covering an increasing large area. We think this is essential, because above a certain scale a map consists of weakly coupled parts, whereas below that scale coupling is much stronger and more complicated and we think not suitable to a $O(n)$ solution (section 2.9). The most important postulate from a practical point of view is (R3), limiting the amount of time spent on each measurement.

4 Related Work

The problem of Simultaneous Localization and Mapping has found considerable interest in the mobile robotics community for more than a decade with first works reaching back to the mid 80s [3]. Two main ideas evolved with the fundamental work of Smith, Self and Cheeseman [15] and Elfes [5]. The latter is to represent the environment by an aligned grid of small elements similar to pixels in an image. The first approach is to extract features called landmarks, that are described by geometrical parameters (usually position).

4.1 Landmark based Approaches

Most landmark based approaches employ a statistical view. They treat the robot's pose and the positions of all landmarks as a state vector and maintain an estimate with corresponding covariance matrix using the EKF equations for updating [15; 2; 9; 10]. Due to the Kalman filter being sound, doing so is consistent and represents all information available to the extent that linearization of the measurement function is adequate, thus meeting (R1). The biggest drawback is the $O(n^2)$ size of the covariance matrix and the $O(n^2)$ per measurement cost for maintaining it. This clearly violates (R2-R3).

Several researchers have divided the environment into submaps with a fixed number of landmarks k, executing an EKF on each [11]. This reduces the space requirements to $O(nk) = O(n)$ and the computational cost to $O(k^2) = O(1)$ per landmark, thus meeting (R2) and (R3). However, not representing correlations between submaps at all, is like not representing correlations between landmarks only on a larger scale and results in not being able to represent relativity or to close loops on that scale, violating (R1).

Recently, impressive progress has been made by Guivant and Nebot, with the "Compressed SLAM Algorithm" [8]. This allows to use a small EKF of the current active submap for accumulating all observations of the submap's k landmarks, with $O(k^2)$ cost per observation. When leaving the submap, the whole information is transfered to a global EKF with computational cost $O(kn^2)$. The result is identical to integrating every measurement into the global EKF. Performing the update in a background process yields $O(k^2 + \frac{k}{m}n^2)$ cost if m landmark observations are made in each submap. With $m \gg k$, this is much better than plain EKF, but far away from being linear. Further reduction of the computational cost is achieved by employing a "relative map" with each landmark being represented by its relative location to a pair of reference landmarks. This makes all non-reference landmarks basically decoupled and allows to apply an update scheme that consistently ignores some of the correlations, reducing the cost to $O(k^2 n)$ per transfer or equivalently to $O(k^2 + \frac{k^2}{m}n)$ per measurement, thus meeting (R3).

This approach is much less conservative than [11]. Whether (R1) holds in a strict sense has to be subject of a thorough analysis. The most critical case for submap based representations is an adjacent pair of landmarks with very precisely known relative location belonging to different submaps.

Lu and Milios [12] avoid the $O(n^2)$ storage space overhead of the covariance matrix by storing a graph of relations between robot poses. They estimate "globally consistent" poses by applying non-linear least square, having to solve a $O(p)$ sized linear equations system each time an estimate of the map is desired (p = number of robot poses), which takes

$O(p^3)$. This can be reduced to $O(p^2)$ by using a sparse matrix solver if the covariance matrix is not required. The approach does not satisfy (R2), since it permanently stores every robot pose having a memory consumption of $O(p)$. This is not linear in the size of the map, but grows if the robot repeatedly moves through the same area.

Golfarelli et al. [7] and Duckett et. al [4] build a graph of relations between certain "places" (in [7] called "landmarks") instead of robot poses and landmarks. Using places as basic entities on the one hand does not represent local uncertainty structure. On the other hand it allows to immediately integrate measurements between the same places thus achieving linear storage space (R2). (R1) is met as far as global uncertainty is concerned. Duckett et al. employ a relaxation scheme equivalent to one Gauss-Seidel iteration after each measurement to avoid the $O(n^2)$ (by exploiting sparsity) cost of equation solving. This is a clever way of meeting (R3), since normally each measurement produces only a small change in the map. However when closing a loop a single measurement has a large effect on the whole map which results in many iterations necessary and is presumably significantly slower than utilizing a direct equation solver, which are known to be much faster than Gauss-Seidel iteration. Golfarelli et al. apply an update procedure based on the mechanical analogy with parameterizable trade off between computation time and update accuracy.

To avoid storing correlations without loosing consistency, Uhlman et al. use covariance intersection [17]. However due to assigning a single individual uncertainty measure to each landmark, their approach cannot represent relativity and thus is not compliant with (R1).

In previous work [6], we have build a graph like the one shown in figure 1b with the edges representing relations with uncertainty similar to covariance intersection. We employed a mechanism to keep the number of represented robot poses low and thus achieved a very compact representation meeting (R1) and (R2) and treating even all nonlinearities conservatively.

4.2 Dense Approaches

A consequent extension of representing maps by evidence grids is to represent uncertainty by grid based probability distributions. Each grid element represents an event and stores the probability for that event. The greatest advantage of this approach is being able to store arbitrary, even multimodal distributions, thus for instance representing ambiguous identifications. The computational cost however is exponential in the number of dimensions, limiting them to 3 or 4. Representing a joint distribution of all robot poses or even of all map instances as in the covariance based approaches is not feasible.

Thrun et al. [1] solve this problem by building the map from possibly overlapping "patches" each described by an evidence grid and a distribution of the pose of that patch in the map. The distribution of the different patches are assumed to be independent. While the robot moves, the raw sensor data (ultrasonic) is integrated into a new patch every 5m.

To achieve global consistency, the grids are aligned by an Expectation-Maximization algorithm modified by an annealing technique to avoid local minima. In the E-step distributions for the robot poses based on the actual map instance are estimated (like Markov Localization). In the M-step distributions for the poses of the patches based on the estimated robot poses are computed. The result is modified by an annealing parameter $\theta \in [0..1]$ which performs a smooth transition from estimating distributions to maximum likelihood estimation (like normally done by the EM algorithm). The process is iterated with decreasing θ until convergence is achieved.

The algorithm exhibits impressive behavior being able to construct a map even from very noisy and low resolution measurements like that produced by ultrasonic sonar. It does not require any landmark or place to be identified, since identification is implicitly performed by the E-step with ambiguous identification yielding multimodal distributions. This is a clear breakthrough, since sonar data is usually so bad, that it is very hard to derive any reliable landmark information or even identify a landmark from it.

The drawback of the algorithm is its computational expense and the assumption of independence between the poses of different patches, that does not allow to represent relativity above patch scale, which is necessary for large maps. Altogether the approach is very different from landmark based SLAM and thus difficult to compare.

A possible way of treating dense environment representations with a methodology similar to landmark based SLAM is the following according to Lu and Milios [12]: Each robot pose is associated with a single laserscan that has been taken from that pose. If two scans cover an overlapping region of the environment, the relation between the two corresponding robot poses can be estimated by comparing them. The resulting graph of relative pose constraints can be treated just like that appearing in landmark based SLAM.

5 Proposing the Use of an Information Matrix

As can be seen from the discussion above, there is a large discrepancy between what one would think a SLAM solution should provide and the actual performance of current approaches. On the other hand, at least the core representations of most approaches are much more general than actually needed. For instance EKF can handle arbitrary measurement equations, covariance matrices can represent any linear relation between variables and maximum likelihood can be used for even more general problems. Thus we think that SLAM has some specific properties that are not utilized yet and that could be a key in meeting the proposed requirements.

In what follows we present a set of properties we have identified and a representation that makes use of them. The representation meets requirements (R1), (R2) and with some restrictions (R3).

Our approach is to achieve memory and time efficiency by representing the landmark uncertainties with a sparse matrix. SLAM covariance matrices are seldom sparse, whereas the information matrix A, used in linearized least square is sparse. The problem is however, that when storing all robot poses, the dimension of the matrix will grow unbounded, violating (R1), while eliminating old robot poses by the Schur complement destroys sparsity.

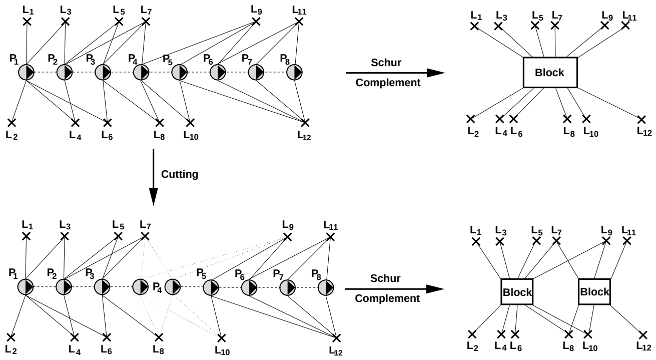

Figure 5: "Cutting" the odometric sequence by duplicating a robot pose with all corresponding landmark observations. Eliminating the robot poses of the resulting two parts yields two dense but small matrix blocks instead of one large dense block when not cutting before elimination.

To avoid this dilemma, we make a consistent approximation. The sequence of odometric measurements is "cut" into pieces connected via shared landmarks (figure 5). Eliminating now the robot poses for a single piece destroys the sparsity only of the submatrix of involved landmarks. This introduces only a small dense block into the whole matrix, which overall remains sparse.

Cutting the odometry sequence is done by duplicating a robot pose together with all landmark observations taken at that pose. Each copy belongs to one piece. To preserve consistency, the covariance of the landmark observation is doubled. The resulting system would be equivalent to the original one when adding the information, that both copies of the robot pose are exactly equal. The error introduced by not adding that information depends on the precision with which that pose is defined by the landmark observations alone (figure 5). Using larger pieces results in less error, but the resulting dense block will be larger and thus the whole matrix less sparse.

After cutting the odometry sequence the information matrix for a single piece is computed. Eliminating all robot poses by the Schur complement introduces off-diagonal entries but only between the involved landmarks. Thus the resulting matrix consists of a small dense block with 0 everywhere else (figure 6). Therefore the integration of a piece - which means to add its information matrix to A - does no destroy the sparsity of A and can be done in computation time independent from the number of landmarks. This procedure does not sacrifice information except that lost due to cutting.

As can be seen from the above discussion, maintaining the representation can be done very efficiently. Extracting information like estimating a map instance requires to solve a linear equation system with A. This is the remaining challenge considering computational efficiency.

An effective method for maintaining an estimated map instance is to use an iterative equation solver like conjugated gradients [14, chapter 2.7]. As the map estimate normally does not change rapidly one iteration per measurement will suffice. More steps are only needed if a large loop is closed.

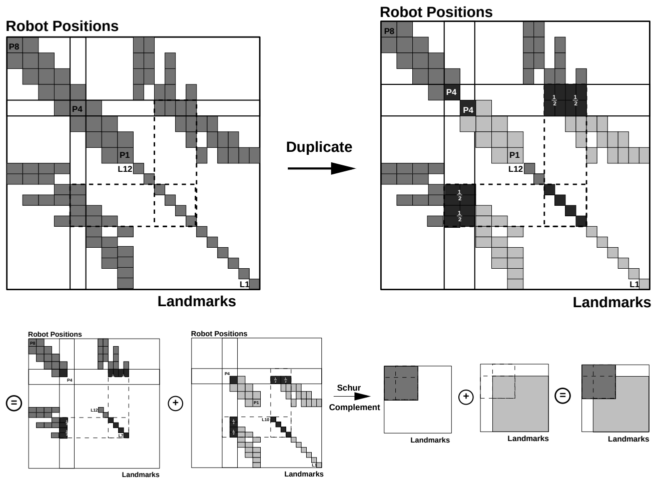

Figure 6: Matrix operation performed, when cutting the odometric sequence into two pieces. The original matrix is the information matrix of L1-L12 and P1-P8. P4 is duplicated together with its landmark observations, cutting the information by 2 to maintain consistency. The result is the sum of two independent information matrices representing both parts. Applying the Schur complement to each yields matrices containing a small dense block. The sum of those blocks is a sparse information matrix of the landmarks positions.

In the following we discuss the properties that motivate this approach (P1-P3) or that we think can be exploited to further speed up equation solving (P3, P4):

> **P1: Odometric Scale**: *There exists a distance ("odometric scale") above which the uncertainty of odometry is significantly higher than of relocalizing with respect to known landmarks.*

In general this is clear from the fact that odometry accumulates error, whereas determining a pose from landmark observations does not. For typical mobile robots the relocalization error is 0.1m and $3°$, which is comparable to the odometric error after moving 2m. So the odometric scale is $\approx 3..6$m.

This property yields that our approximation is compliant with (R1). The odometric scale defines the minimum length for the odometric pieces that does not introduce too much additional error compared to the accumulated odometrical error.

> **P2: Sparse Relations**: *The number of landmarks that can be observed in context with a certain landmark is constant and small.*

For every possible robot pose, all landmarks but a few are either occluded or out of sensor range. This and the fact that the robot moves continously are the reasons for this property.

Properties (P1) and (P2) together, imply that when cutting the whole odometry sequence into pieces, only a constant number of landmarks will appear in the same block as a certain landmark. Thus the whole information matrix remains sparse containing a number of non-zero entries linear in the number of landmarks and meeting (R2).

This further implies that one iteration of conjugate gradients needs linear time, since the dominating operation of one step is computing a matrix vector product. If the application only needs an estimated map instance without having easily accessible uncertainty information, (R3) is met.

> **P3: Area of Interest**: *All of the time the only relevant landmarks are those in the vicinity of the robot. This subset changes slowly and predictably, except when closing large circles.*

Due to limited sensor range and occlusion, only landmarks in the vicinity of the robot can be observed and used for localization. Distant landmarks could be involved in path planning, but with a map that might change at any moment, it is desirable to use only those relations that are known precisely. This is entirely possible, as discussed in section 2.3, justifying this property. All submap based approaches and especially compressed SLAM [8] utilize it, although in general the area of interest constitutes rather something like a sliding window than a set of submaps.

> **P4: Decomposability**: *Any large map can be split into parts with a small shared boundary.*

This property holds for any planar surface, with the shared boundaries length being at most proportional to the square root of the area of the surface. For large indoor maps the length can be made even smaller, as it suffices to cut a small number of doorways.

To speed up equation solving, one could divide the set of landmarks into parts decomposing the matrix A accordingly. For any non-zero entry a_{ij} the landmarks corresponding to i and j must be represented in a common submatrix of A. So some landmarks will be represented in several submatrices, making the decomposition less effective.

By construction $a_{ij} \neq 0$ holds only if the corresponding landmarks can both be observed with the robot moving less than the odometric scale. Together with (P4), this means that only a small fraction of the landmarks will be so near to the boundary that they can be involved in a dense matrix block together with some landmark from the other side.

6 Conclusion

In this paper we have discussed the intended semantics of an uncertain map representation together with the structure of the SLAM problem itself. We have identified some intuitive requirements which a SLAM solution should meet. We conjectured that the discrepancy between these requirements and current approaches is due to not exploiting some key properties of SLAM and proposed a set of properties as candidates.

We describe an uncertainty representation that utilizes these properties to meet the proposed uncertainty and memory requirements. To derive an estimate from it, a sparse linear equation system must be solved. However updating the representation itself takes only constant time.

Our current research is addressing the problem of building a complete SLAM algorithm using the proposed representation. Here the most important point will be to utilize SLAM properties for efficiently solving the appearing equation systems. As our representation has constant update time, achieving linear time for equation solving would completely fulfill all proposed requirements.

References

[1] W. Burgard, D. Fox, H. Jans, C. Matenar, and S. Thrun. Sonar-based mapping with mobile robots using EM. In *Proc. 16th International Conf. on Machine Learning*, pages 67–76. Morgan Kaufmann, San Francisco, CA, 1999.

[2] J. A. Castellanos, J. Montiel, J. Neira, and J. Tardos. The spmap: A probablistic framework for simultaneous localization and map building. *IEEE Trans. on Robotics and Automation*, 15(5):948 – 952, 10 1999.

[3] R. Chatila and J. Laumond. Position referencing and consistent world modeling for mobile robots. In *IEEE Int. Conf. on Robotics and Automation*, pages 138 – 145, 1985.

[4] T. Duckett, S. Marsland, and J. Shapiro. Learning globally consistent maps by relaxation. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'2000)*, San Francisco, CA, 2000.

[5] A. Elfes. *Occupancy Grids: A probabilistic framework for robot perception and navigation.* PhD thesis, Dep. of Eletrical Engineering, Carnegie Mellon Univertsity, 1989.

[6] U. Frese, M. Hörmann, B. Bäuml, and G. Hirzinger. Globally consistent visual localization without a-priori map (german). *at, automatisierungstechnik*, 3:273 – 280, 6 2000.

[7] M. Golfarelli, D. Maio, and S. Rizzi. Elastic correction of dead-reckoning errors in map building. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'98)*, pages 905–911, Victoria, Canada, 1998.

[8] J. Guivant and E. Nebot. Optimization of the simultaneous localization and map building algorithm for real time implementation. *IEEE Journal of Robotic and Automation*, (subm.).

[9] P. Hebert, S. Betge-Brezetz, and R. Chatila. Probabilistic map learning, necessity and difficulty. In *Proceeding of the Int. Workshop Reasoning with Uncertainty in Robotics*, 1995.

[10] J. J. Leonard and H. F. Durrant-Whyte. Dynamic map building for an autonomous mobile robot. *The Int. Journal on Robotics Research*, 1992.

[11] J. J. Leonard and H. J. S. Feder. Decoupled stochastic mapping. Technical report, MIT, 12 1999.

[12] F. Lu and E. Milios. Globally consistent range scan alignment for environment mapping. In *Autonomous Robots 1997*, volume 4, pages 333 – 349, 1997.

[13] J. Neira and J. Tardos. Robust and feasible data asssociation for simultaneous localization and map building. Workshop ICRA2000, 2000.

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

[15] P. C. R. Smith, M. Self. Estimating uncertain spatial relationships in robotics. *Autonomous Robot Vehicles*, pages 167 – 193, 1988.

[16] S. Thrun, W. Burgard, and D. Fox. A probabilistic approach to concurrent mapping and localization for mobile robot. *Machine Learning and Autonomous Robots*, 31(5):1 – 25, 1998.

[17] J. K. Uhlmann, S. J. Julier, and M. Csorba. Nondivergent simultaneous map building and localization using covariance intersection. In *Proc. - SPIE the Int. Society of Optical Engineering, 2, 3087*, 1997.