An alternative to the Mahalanobis distance for determining optimal correspondences in data association

Jose-Luis Blanco, Javier González-Jiménez and Juan-Antonio Fernández-Madrigal

***Abstract*— The most common criteria for determining data association rely on minimizing the squared Mahalanobis distance (SMD) between observations and predictions. We hold that the SMD is just a heuristic, while the alternative *matching likelihood* (ML) is the optimal statistic to be maximized. Thorough experiments undoubtedly confirm this idea, with false positive reductions of up to 16%.**

***Index Terms*— SLAM, data association.**

I. Introduction

The problem of Simultaneous Localization and Mapping (SLAM) has been one of the most studied topics in mobile robotics during the last decade [1]–[4]. Maps of the robot environment learned by SLAM can be roughly classified as either continuous representations (such as occupancy grid maps [5], elevation maps [6] or gas concentration maps [7]) or discrete (object-based) representations, comprised of a variable number of natural or artificial *landmarks* present in the environment. In this work we focus on this latter approach. Using a map of discrete elements has some advantages such as the relative efficiency of graph SLAM [8] and EKF-like [9] algorithms in comparison to alternatives for continuous maps – e.g. grid mapping with particle filters [10].

However, discrete maps introduce two hurdles: first, in most cases sensors do not directly detect landmarks, thus an additional *detection* step must be introduced whose failure would severely degrade the overall mapping performance. Secondly, once a set of observed landmarks is available from the sensor, they must be paired with those already in the map. This is the *data-association* (DA) problem, which is the central concern of this work.

The DA problem can be stated as follows: at some time step t, and given the vector with N landmark observations $\mathbf{z}_t$, compute the N-length association vector $\mathbf{n}_t$ which states to which map landmark does each observation corresponds (or whether it is a new landmark not observed before). Each landmark observation is a point in the *observation space* whose dimensionality depends on the specific problem, e.g. two-dimensional in planar range-bearing SLAM with point features [4], [11] and in Monocular SLAM [12], or one-dimensional in range-only SLAM [13], [14]. Each of these observation points must be paired with either one or none of a set of *predictions*, or expected observation for each known landmark in the map. Given that the sensor model is stochastic and both the vehicle pose and the map are represented as probability densities, these predictions are probability distributions as well – typically, Gaussians.

As we will discuss in Section II, the most popular methods for solving DA are the Nearest Neighbor (NN) [4] and the Joint-Compatibility Branch and Bound (JCBB) [11] algorithms. As described in the literature, these methods aim at establishing the most likely pairings by minimizing the squared Mahalanobis distance (SMD) between the observations and their associated predictions.

The central claim of this work is that minimizing the SMD does not always lead to the most likely pairings, as can be easily demonstrated. Consider the probability mass function over all the possible associations $\mathbf{n}_t$ for a time step t, given knowledge about the joint vehicle-map state vector $\mathbf{s}_t$ and the latest observation $\mathbf{z}_t$, which follows the conditional distribution $P(\mathbf{n}_t|\mathbf{s}_t, \mathbf{z}_t)$. By definition, the most likely set of associations is the value of $\mathbf{n}_t$ that maximizes this distribution. Applying the Bayes rule over the observation $\mathbf{z}_t$:

$$\begin{aligned} P(\mathbf{n}_t|\mathbf{s}_t, \mathbf{z}_t) &\propto \cancelto{\eta}{P(\mathbf{n}_t|\mathbf{s}_t)} p(\mathbf{z}_t|\mathbf{s}_t, \mathbf{n}_t) \\ &\propto p(\mathbf{z}_t|\mathbf{s}_t, \mathbf{n}_t) \end{aligned} \quad (1)$$

using the fact that the a priori distribution of the associations not conditioned to any observation must be uniform, leading to an irrelevant constant term η. A natural and expected consequence of the equation above is that optimal correspondences are those that "best explain" the observations.

If we let $\mathcal{N}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denote the evaluation at $\boldsymbol{x}$ of the probability density function (pdf) of a multivariate Gaussian with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, the observation likelihood in EKF-based or graph SLAM can be denoted as:

$$p(\mathbf{z}_t|\mathbf{s}_t, \mathbf{n}_t) = \mathcal{N}(\mathbf{z}_t; h(\mathbf{s}_t, \mathbf{n}_t), \mathbf{S}(\mathbf{n}_t)) \quad (2)$$

where $h(\cdot)$ is the observation model that gives us the prediction vector $\mathbf{h}_{\mathbf{n}_t}$, and $\mathbf{S}(\mathbf{n}_t)$ is its covariance[1]. We will refer to Eq. (2) as the *matching likelihood* (ML), since all the terms are known with the exception of $\mathbf{n}_t$, the vector of the matching hypothesis.

On the other hand, the SMD between observations and predictions for a given hypothesis $\mathbf{n}_t$ is given by:

$$D^2_M(\mathbf{n}_t) = (\mathbf{z}_t - \mathbf{h}_{\mathbf{n}_t})^\top \mathbf{S}(\mathbf{n}_t)^{-1} (\mathbf{z}_t - \mathbf{h}_{\mathbf{n}_t}) \quad (3)$$

It is straightforward to see that the SMD is one of the terms in the expression of the previous ML, as can be verified expanding Eq. (2) as a negative log-likelihood:

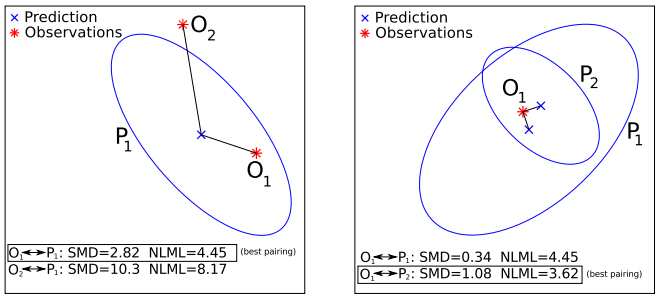
$$\begin{aligned} -\log p(\mathbf{z}_t|\mathbf{s}_t, \mathbf{n}_t) &= -\log \left\{ \frac{1}{\sqrt{(2\pi)^d |\mathbf{S}(\mathbf{n}_t)|}} \exp\left(-\frac{1}{2} D^2_M(\mathbf{n}_t)\right) \right\} \\ &= \frac{1}{2}\left(d \log 2\pi + D^2_M(\mathbf{n}_t) + \log|\mathbf{S}(\mathbf{n}_t)|\right) \end{aligned} \quad (4)$$

where d represents the dimensionality of the $\mathbf{z}_t$ vector. Since the constant factor $\frac{1}{2}$ becomes irrelevant while comparing ML values for different correspondence hypotheses, we conveniently define the NLML (negative logarithm of the ML) as twice Eq. (4), thus:

$$\mathrm{NLML}(\mathbf{n}_t) = d \log 2\pi + D^2_M(\mathbf{n}_t) + \log|\mathbf{S}(\mathbf{n}_t)| \quad (5)$$

Authors are with the Department of System Engineering and Automation, University of Málaga. Email: joseluisblancoc@gmail.com, {jgonzalez,jafma}@ctima.uma.es

[1]For an additive Gaussian noise sensor model with noise covariance matrix $\mathbf{R}$, this covariance becomes $\mathbf{H}\mathbf{P}_\mathbf{k}\mathbf{H}^\top + \mathbf{R}$ with $\mathbf{H}$ the observation Jacobian and $\mathbf{P}_\mathbf{k}$ the vehicle-map joint covariance. In this work we will assume that $\mathbf{R}$ is either constant or depends on the prediction $h(\cdot)$, thus it will be always constant when evaluating a prediction against multiple potentially-paired observations. However, our claims about ML and SMD still hold if the model for $\mathbf{R}$ was to depend on the observation readings $\mathbf{z}_t$.



(a) (b)

Fig. 1. A motivating example for this work. The values of the squared Mahalanobis distance (SMD) are compared to those of the negative logarithm of the matching likelihood (NLML) for two hypothetical situations. In both SMD and NLML, lower values mean "better" pairings. (a) Two observations are tested against a probabilistic prediction. In this case the minimum SMD can be safely used to detect the most likely pairing. (b) One observation is tested against two predictions. Here, minimizing SMD does not give us the optimal pairing.

Clearly, minimizing the SMD ($D^2_M(\mathbf{n}_t)$) is not equivalent to minimizing the NLML (or maximizing the ML). The other two terms (the dimensionality term and the determinant of the covariance $\mathbf{S}_t$) can also change among different potential pairings and their contribution may make the difference between establishing correct or wrong associations.

We can illustrate this discussion with the example in Fig. 1(a), where two observations are compared to one prediction. In this example, the same covariance $\mathbf{S}$ appears in both expressions of the ML, thus it is justified to employ the SMD to decide the correspondence, since the minima of SMD and NLML coincide. In contrast, a situation more commonly found in SLAM is similar to that depicted in Fig. 1(b), where one or more observations are to be tested against one or more predictions, thus Gaussians with different covariance matrices are involved. As shown in the figure, in this case the SMD would recommend to associate the observation with P_1 although the pairing with P_2 is more likely.

With this paper we make two contributions to the field of probabilistic mobile robotics: (i) an in-depth discussion of the principles that underlie the decisions made in DA, which is rarely found in the literature, and (ii) the proposal of the NLML as the *optimal statistic* (by definition) to determine correspondences. As validated below with experiments, using the NLML instead of the SMD reduces the average ratio of wrong correspondences in SLAM. The resulting improvement becomes more relevant as the level of noise in the sensor increases, which explains why the SMD has led to such good results in the past for sensors with relatively low levels of noise (e.g. [11]).

The rest of this paper is organized as follows. First, we review the existing literature on DA in EKF-SLAM. Section III is devoted to comparing the theoretical performance of the expected behavior of SMD and the ML, and then Section IV summarizes the algorithms for establishing optimal associations. Next we analyze the computational complexity of ML, while experimental results are provided in Section VI.

II. Data association in the SLAM literature

The origin of the basic EKF-based approach to SLAM can be found in the seminal paper by Smith *et al.* [15] which settled the bases for stochastic mapping, but which did not mention the problem of data association. A decade after this work, in the influential papers by Dissanayake *et al.* [9], [16] which theoretically established that EKF-SLAM actually converges to a solution, we can witness a first criterion to decide the observation-to-landmark associations: only those observations that, given a SMD threshold, clearly correspond to *one and only one* map landmark are paired, discarding those with multiple potential pairings. This method works perfectly well for environments where the sensor noise plus the vehicle uncertainty are always below the typical distance between any pair of landmarks. However, this limit can be overcome with the more advanced techniques discussed below. Another peculiarity about [9] is that DA is performed in the landmark space, that is, SMD is computed between landmarks in the map and projections of the observations into that map. This approach has some disadvantages, such as the fact that the projection of observations as Gaussians into the map is not always well-defined. For instance, in Monocular SLAM the inverse sensor model is not well approximated by a Gaussian – although smart parameterizations alleviate the problem [17]. In addition, projecting the observations into the map necessarily introduces a correlation between all of them, and also between them and the landmarks, via the vehicle uncertainty. In this situation, DA methods (specially those that consider joint densities [11]) have to deal with extra cross-covariance terms which can be avoided if DA is carried out in observation space. Therefore, it is not surprising that most recent works perform DA in observation space.

The simplest DA method that deals with ambiguity (multiple potential pairings for each observation) is the Nearest Neighbor (NN) algorithm [4], which can be traced back to the target tracking literature [18]. This method works on each observation independently, associating it to the prediction that, in theory, maximizes the likelihood of that individual observation. In practice, most works apply NN by using the SMD heuristic instead of the ML [4], [9]. The NN method is able to cope with ambiguity, although its decisions quickly become wrong as the level of noise in the system increases.

Superior robustness can be achieved with batch DA methods, such as the Joint Compatibility Branch and Bound (JCBB) algorithm [11], which, in contrast to NN, also takes into account the cross-correlations of the predictions. Again in this case, the proposed method tries to minimize the joint SMD between the observations and the predictions, in contrast to the maximization of the ML [2]. Another batch method is the Combined Constraint Data Association (CCDA) [20], which is also able to perform global localization of a mobile robot by searching in a graph containing all the potential pairwise correspondences.

Data association is also present in those approaches to SLAM based on Rao-Blackwellized Particle Filters (RBPF) [21], such as the implementation for landmark maps in [22] or the more advanced solution in [23]. Due to the properties of RBPFs, landmarks in each map hypothesis are uncorrelated, but it is still advisable to apply a method such as JCBB instead of the simpler NN in order to cope with ambiguity. Interestingly, optimal DA in the context of RBPF SLAM was correctly stated to be the maximization of the ML [24], whereas the heuristic technique of minimizing the SMD has dominated the literature on EKF-based SLAM.

III. On the usage of the Mahalanobis distance for data association

Looking for the set of optimal correspondences $\hat{\mathbf{n}}_t$ implies maximizing the ML in Eq. (1) or, taking logarithms, minimizing the NLML in Eq. (2). In this section we provide an insight into why minimizing the SMD has led to so many good results in the literature

[2]To the best of our knowledge, the only mention to using the theoretically "more correct" ML in the JCBB algorithm is a footnote on page 38 in [19]. The present work is the first one to provide an experimental measurement of the performance of each alternative.

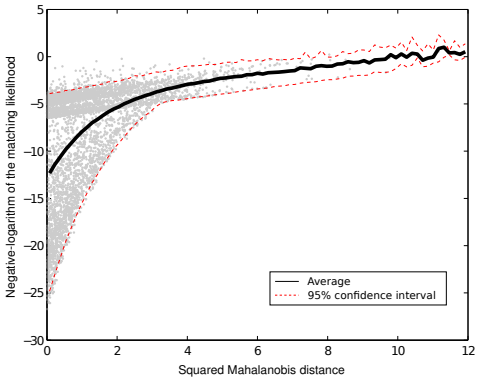


Fig. 2. Points representing the SMD and ML of a set of $2 \cdot 10^4$ valid correspondences evaluated during a simulation of 2D range-bearing EKF-SLAM for random maps (refer to Section VI). It is clear that a small Mahalanobis distance is not a guarantee for a high likelihood value, since correspondences within a wide range of ML may have the same SMD value.

(e.g. [9], [11]), despite it not being the optimal statistic. Our intention is to avoid potential confusions in future research, since the SMD in the context of EKF-based SLAM has been sometimes referred to as a maximum likelihood statistic to determine associations [4].

When considering potential pairings of one uncertain prediction against several observations, as depicted in Fig. 1(a), we find that the NLML for each possible association n_i is $\mathrm{ML}(n_i) = d \log 2\pi + D^2_M(n_i) + \log |\mathbf{S}|$ where both d and the matrix inside the determinant are constant over the candidate indexes i. In this situation, minimizing the SMD or the NLML are *exactly* equivalent.

In contrast, if we consider another scenario such as that in Fig. 1(b), where one observation has multiple potential feasible predictions, these two criteria are not equivalent any more. This follows from the contribution of the determinant of the prediction covariance, which models the different levels of uncertainty in each prediction. It is clear that the greater the range of covariance determinants found among the different predictions, the more different results will be obtained from the two statistics.

In our opinion, the fact that most of the prediction covariances have comparable determinant values in 2D range-bearing EKF-SLAM has led to small errors attributable to using SMD instead of NLML. In experiments previously reported in the literature, these errors might perfectly have been masked among those naturally occurring due to the stochastic nature of the decisions. To provide a quantitative verification of this idea, we present in Fig. 2 a plot of NLML vs. SMD where each point represents a *valid* correspondence of a prediction and the noisy observation of the corresponding landmark from a series of simulations with known ground truth. If both measures were exactly equivalent, the plot would describe a well-defined, thin trace. But that is clearly not the case: in general, NLML and SMD are correlated, but many correspondences with identical SMD values usually have very different likelihood values, i.e. any vertical cut to the graph will find a wide range of likelihood values.

Interpreting the criterion of minimizing the SMD under the light of these data we find that, indeed, a candidate pairing c_1 with a SMD smaller than an alternative c_2 will be also, *on average*, more likely:

$$D^2_M(c_1) < D^2_M(c_2) \rightarrow E[\mathrm{NLML}(c_1)] < E[\mathrm{NLML}(c_2)] \qquad (6)$$

Therefore, if we only had the SMD, its minimizing would be a good heuristic. However, provided that it is always possible to directly evaluate the NLML at the same cost (see Section V), we claim that employing the SMD to decide DA is *suboptimal* and leads to unnecessary errors.

We have evaluated the decisions between alternative candidates from all the possible pairs among the $2 \cdot 10^4$ measurements shown in Fig. 2. Our results show that $\sim 11\%$ of the decisions based on SMD would be wrong, i.e. in those cases we have $D^2_M(c_1) < D^2_M(c_2)$ but $\mathrm{NLML}(c_1) \not< \mathrm{NLML}(c_2)$. The result strongly validates our claim, but this particular figure should be taken just as an illustrative example; SMD-induced errors largely depend on the specific SLAM problem being solved, the sensor model, the sparsity of the landmarks, the noise levels and the path traversed by the robot.

IV. Determining Optimal Correspondences

Given the set of predictions and the set of actual observations, any solution to the DA problem must decide, for each individual observation, whether it corresponds to a map landmark or it is a new one, not previously observed. Each of these individual decisions can fall into one and only one of these categories:

- *True positive:* Observation correctly associated to its landmark in the map.
- *True negative:* Observation correctly classified as new.
- *False positive:* Observation incorrectly associated to a non-corresponding known landmark.
- *False negative:* Observation classified as a new landmark while it actually corresponds to one already in the map.

One of the most challenging aspects of DA is that not all the landmarks are observed simultaneously in all time steps, hence the need to detect observations of *new* landmarks not observed before. If all the map landmarks were simultaneously observed at all time steps, only true positives and false positives would be possible, and we could then directly associate each observation to its most likely landmark by optimizing a given measure (either SMD or ML). Actually, it is the possibility of some certain observations not to correspond to any known landmark, either for being a new landmark or a spurious, what forces us to introduce thresholding, or *gating*, as a first stage of any DA method.

In the following we describe the two stages of DA: (i) gating, to determine potential pairings, and (ii) the resolution of ambiguity in the case of multiple compatible correspondences. The distinction between SMD and NLML is relevant mainly in the latter.

A. The problem of gating

In all DA methods the set of potential pairings between observations and predictions is firstly pruned by only considering those whose SMD is below a given threshold T_{SMD}. This method has been only touched superficially in the literature despite being the main responsible of all false negatives.

For the case of SMD, the threshold typically used is the chi-squared inverse cumulative distribution function (cdf) $\chi^2_{d,c}$ for any desired confidence c level, e.g. 95%. Two crucial remarks are relevant at this point: (i) this confidence level c must not be confused with the actual probability of accepting true positives (in fact, $1 - c$ describes the probability of accepting false negatives, *conditioned* to the assumption of a correct pairing), and (ii) although the underlying assumption that SMD values for valid correspondences follow a chi-squared distribution is very accurate, it is not theoretically exact: due to errors in the reconstruction of landmark positions in the estimated map, the actual distribution for each landmark is a *non-central* chi-square with unknown parameters. However, the central chi-square is a good model in most practical situations and, being pragmatic, is the best (and possibly the unique) approximation available to us.

It is clear that there is no "perfect" (or "magic") threshold when performing DA gating, neither for the SMD nor for the ML: *any*

threshold will have an associated rate of errors, both false negatives and false positives. We have numerically evaluated the optimal thresholds, for both SMD and NLML, from a data set with millions of correct and incorrect potential pairings obtained from simulations. Our results confirm that, obviously, the optimal thresholds depend on the robot path, the noise levels, etc. Moreover, we found out that the overall rate of misclassification has only a weak sensitivity to these thresholds. In particular, any threshold in the range $[9, 16]$ for the SMD or within $[-1, 6]$ for the NLML leads to almost identical results[3]. Although optimal gating deserves further research, it seems clear that the choice between NLML or SMD at this point is not more relevant for gating than the election of the threshold. We now focus on the second stage of DA, which clearly encourages the usage of NLML.

B. Dealing with ambiguity

After gating, each observation has an associated set of potential, *individually compatible* [11] pairings. Empty sets can be assumed to correspond to new landmarks, thus we ignore them in the following discussion. In most practical situations, all the remaining sets will comprise multiple candidates which must be resolved. Even in the case where all observations have just one compatible pairing each, one should check whether all those pairings are mutually compatible.

We define the optimal correspondences $\hat{\mathbf{n}}_t$ as those that maximize the likelihood of the latest observation $\mathbf{z}_t$:

$$\begin{aligned}\hat{\mathbf{n}}_t &= \arg\max_{\mathbf{n}_t} P(\mathbf{n}_t|\mathbf{s}_t, \mathbf{z}_t) \quad \text{(See Eq. 1)} \\ &= \arg\max_{\mathbf{n}_t} p(\mathbf{z}_t|\mathbf{s}_t, \mathbf{n}_t) \\ &= \arg\max_{\mathbf{n}_t} \mathcal{N}(\mathbf{z}_t; h(\mathbf{s}_t, \mathbf{n}_t), \mathbf{S}_t) \end{aligned} \tag{7}$$

Note that for each time step t, observations $\mathbf{z}_t$ actually contain a variable number of observations of individual landmarks, i.e. $\mathbf{z}_t = \left\{\mathbf{z}_t^i\right\}_i$, each demanding a DA decision n_t^i. At this point, one can follow one of the following two approaches.

1) The Nearest Neighbor method: The Nearest Neighbor (NN) method is an approximative approach that solves DA on an individual observation basis [4]. Under a probabilistic viewpoint, the foundation of this method can be seen as the consideration of marginal distributions, ignoring all the cross-correlation terms. Then, the ML distribution in NN is factored into the product of the terms for each individual observation $\mathbf{z}_t^i$:

$$\begin{aligned}\hat{\mathbf{n}}_t = \left\{\hat{n}_t^i\right\}_i &= \arg\max_{\mathbf{n}_t} p(\mathbf{z}_t|\mathbf{s}_t, \mathbf{n}_t) \\ &\approx \arg\max_{\mathbf{n}_t} \prod_i p(\mathbf{z}_t^i|\mathbf{s}_t, n_t^i) \end{aligned} \tag{8}$$

thus each correspondence decision n_t^i can be obtained independently:

$$\begin{aligned}\hat{n}_t^i &\approx \arg\max_{n_t^i} p(\mathbf{z}_t^i|\mathbf{s}_t, n_t^i) \\ &= \arg\max_{n_t^i} \mathcal{N}(\mathbf{z}_t^i; h^i(\mathbf{s}_t, n_t^i), \mathbf{S}_t^i) \end{aligned} \tag{9}$$

As mentioned earlier, the implementation of NN found in the SLAM literature does not actually optimize the likelihood in Eq. (9) but the SMD which is only one of the terms involved – remember Eq. (5). This practice will reveal as suboptimal in the experimental results presented later on.

[3]The interested reader can find the experimental data and a figure summarizing them in the multimedia attachment

2) Methods considering the joint distribution: Another family of methods reported in the literature takes into account the full joint distribution in Eq. (7). As can be expected, these approaches are more computationally expensive but in turn can manage the ambiguity in DA more robustly than NN. Some of these techniques can even handle the problem of global localization, that is, to relocalize a robot without any prior belief for its pose. No description will be given here of these methods, since they have been thoroughly discussed elsewhere [11], [19], [20], [25]. Those algorithms can be employed for minimizing the NLML instead of the SMD by just replacing the evaluation of the latter by Eq. (5).

In the next sections we will use JCBB [11] as the most popular representative of this family of DA methods. Experiments confirm that minimizing the SMD instead of the ML within this algorithm undoubtedly also leads to suboptimal results.

V. Computational cost of the ML

We compare next the computational complexities of evaluating both the SMD and the NLML. As can be seen in Eq. (5), the SMD appears in the expression of the NLML, thus we firstly discuss the calculation of the former. Starting with Eq. (3), we proceed as follows:

$$\begin{aligned} D_M^2 &= (\mathbf{z} - \mathbf{h_n})^\top \mathbf{S}^{-1} \underbrace{(\mathbf{z} - \mathbf{h_n})}_{\tilde{\mathbf{y}}} = \tilde{\mathbf{y}}^\top \underbrace{\mathbf{S}}_{\mathbf{L}\mathbf{L}^\top}{}^{-1} \tilde{\mathbf{y}} \\ &= (\tilde{\mathbf{y}}^\top \mathbf{L}^{\top -1})(\mathbf{L}^{-1}\tilde{\mathbf{y}}) = (\mathbf{L}^{-1}\tilde{\mathbf{y}})^\top \underbrace{(\mathbf{L}^{-1}\tilde{\mathbf{y}})}_{\mathbf{q}} = \mathbf{q}^\top \mathbf{q} \end{aligned} \tag{10}$$

where $\mathbf{L}$ is the lower triangular matrix obtained from the Cholesky decomposition of $\mathbf{S}$, a transformation with a time complexity of $O(N^3)$ for an $N \times N$ matrix, with N proportional to the number of simultaneously-observed landmarks at a time step. The rest of the operations, that is, solving $\mathbf{L}\mathbf{q} = \tilde{\mathbf{y}}$ by back substitution and the dot product $\mathbf{q}^\top \mathbf{q}$, have complexities of $O(N^2)$ and $O(N)$, respectively, thus the overall complexity of computing the SMD is cubic with the size of the observation vector.

The relevant extra operation required for NLML is computing the determinant of $\mathbf{S}$, as can be seen in Eq. (5). Since we already have the Cholesky decomposition of this matrix, the determinant can be obtained as $|\mathbf{S}| = \left|\mathbf{L}\mathbf{L}^\top\right| = |\mathbf{L}|^2$, where $|\mathbf{L}|$ turns out to be the product of the diagonal elements, that is, an $O(N)$ operation.

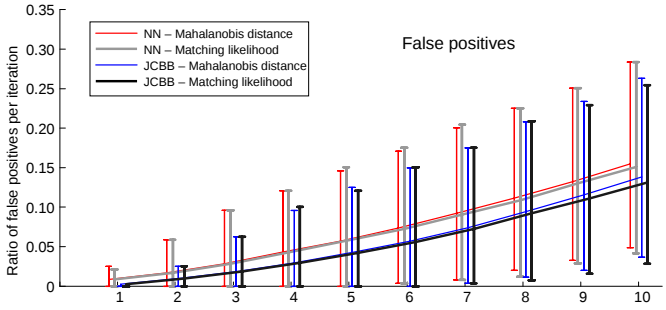
Therefore, we conclude that computing the NLML instead of the SMD does not modify the computational complexity and only introduces a very small overhead.

VI. Experimental evaluation

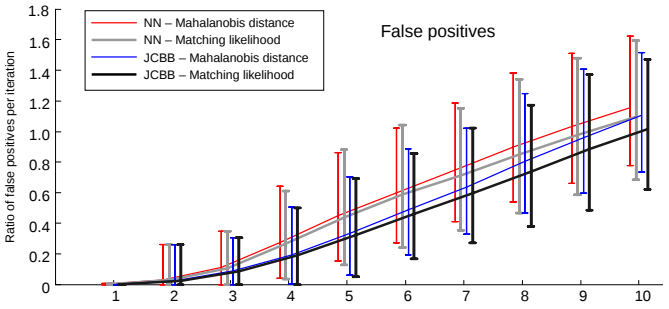
This section describes a series of experiments aimed at confirming our previous discussion defending the usage of the NLML instead of the SMD in order to reduce the number of incorrect associations in feature-based SLAM. We will focus on the particular case of EKF-based SLAM.

A. SLAM in two synthetic environments

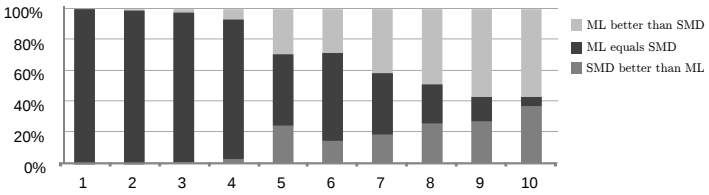
In the first place, we have carried out a comprehensive comparative benchmark. The performance of DA can be only measured in terms of correct or incorrect associations, thus we have relied on simulations where the real correspondences (ground truth) are available and we can study how performance evolves for different levels of sensor



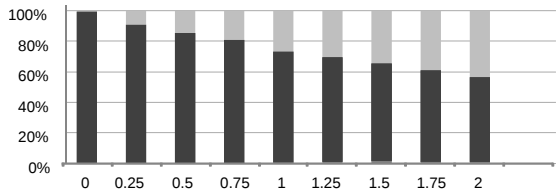
(a) Randomly-generated maps: DA vs. noise level



(b) Square corridor map (8x8m): DA vs. noise level



(c) DA with JCBB: Random maps vs. noise level



(d) DA with JCBB: Random maps vs. ratio of outliers

Fig. 3. Results obtained for (a) 1000 EKF-SLAM runs with maps of randomly-distributed landmarks and (b) 500 runs in a corridor-like fixed map. Confidence intervals (5%–95%) and mean value of the average ratio of false positives per time step for both scenarios. Please, notice the different vertical scales. (c)–(d) Bar graphs with the ratio of time steps at which DA criteria outperform or match to each other. The horizontal axes in (a)–(c) correspond to different levels of noise in the simulated sensors. The standard deviations of the range and bearing noises for each of the 10 cases are: $\sigma_{range} = \{1, 4, 7, 10, 13, 16, 19, 22, 25, 28\}(cm)$ and $\sigma_{bearing} = \{0.02, 0.05, 0.10, 0.30, 0.50, 0.75, 1.00, 1.25, 1.35, 1.45\}(deg)$, respectively. The horizontal axis in (d) represents the average number of spurious readings per time step.

noise. The benchmark has been implemented[4] for the problem of 2D range-bearing SLAM [9], which dominates the literature on DA in SLAM. We evaluated the performance of four different DA methods, namely, NN and JCBB, each in two versions: one based on SMD and another on ML. The four combinations have been tested in two different scenarios that present different challenges to DA. The first one consists of a $12m \times 12m$ map with 100 landmarks randomly distributed following a uniform distribution. The robot follows a $8m \times 8m$ square trajectory in this relatively-sparse scenario, including closing a loop once. In the second scenario the robot follows exactly the same trajectory but in this case in a "corridor-like" map, with landmarks intentionally positioned close to each other to present a challenge to DA. In all cases the sensor has a field of view of 140 degrees and detects landmarks within a minimum and maximum range of 0.15 and 5 meters, respectively. In all the experiments, the threshold for gating pairings in the first stage of DA was set to a 99% confidence interval for SMD and to 0 for NLML.

The results of 1000 and 500 runs in the first and second scenario, respectively, are summarized in Fig. 3(a)–(b). These series have been repeated for 10 different levels of sensor noise, represented by indices running from 1 to 10 in the horizontal axes. In each run, a high-quality pseudo-random number generator [26] has been employed to simulate the system noises, initializing its seed to an identical state for each of the four methods in order to obtain a fair comparison.

In Fig. 3(a)–(b) we represent the false positives attained by each method. As expected, JCBB performs, on average, better than the simpler NN method. An observation validating our claim is that, for either NN or JCBB, using the NLML improves the average results of the SMD version of the algorithm. The improvement is small for low levels of noise but becomes more evident for more noisy systems. Another remark is that the improvement is more patent for the "corridor-like" map than for random maps, which implies that the advantage of ML increases with the hardness of the DA being solved. Interestingly, for very noisy environments the usage of NLML with NN matches the performance of the much more complex JCBB algorithm with the SMD criteria (see the right part of Fig. 3(b)).

We also studied a number of other statistics for each scenario (such as false negatives and errors in the vehicle pose) and the effects of noise level in the vehicle odometry. All the results consistently show a small but unquestionable better performance of NLML over SMD[5].

Focusing on the most relevant case, the usage of the JCBB algorithm, we have plotted as vertical bars in Fig. 3(c) the ratio of time steps where: (i) the NLML criterion leads to a reduction in the number of false positives, (ii) both NLML and SMD recommend exactly the same associations and (iii) SMD outperforms NLML. It can be seen how for low levels of noise the dominant situation is an equivalence of both criteria but for noisy sensors the NLML becomes better than SMD up to in a $\sim 60\%$ of the observations, i.e. the light-gray portion of the graph bars. In terms of false positive reduction, NLML+JCBB reduces an average of $\sim 1\%$ to $\sim 8\%$ of the overall wrong associations with respect to SMD+JCBB.

B. Behavior against outliers

We also repeated the previous benchmark but, instead of sweeping over different levels of noise in the sensor, we investigated the effects of spurious readings, while keeping constant the sensor and odometry errors.

Again, we evaluated the same four DA techniques in two types of environments, the corridor-like scenario and the random maps. The improvement implied by the replacement of SMD with NLML is even more relevant here than in the previous benchmarks. As above, we represent the behavior of MLML and SMD for the JCBB method and the random maps as vertical bars in Fig. 3(d). The horizontal axis

[4]The simulator designed for this purpose is part of the open source MRPT project and is available for download, along with instructions to reproduce the benchmark (refer to the multimedia attachment of this paper). For the program itself, refer to `http://www.mrpt.org/Application:2d-slam-demo`.

[5]Graphs for these and other statistics, as well as the raw data, can be found in the multimedia attachment.

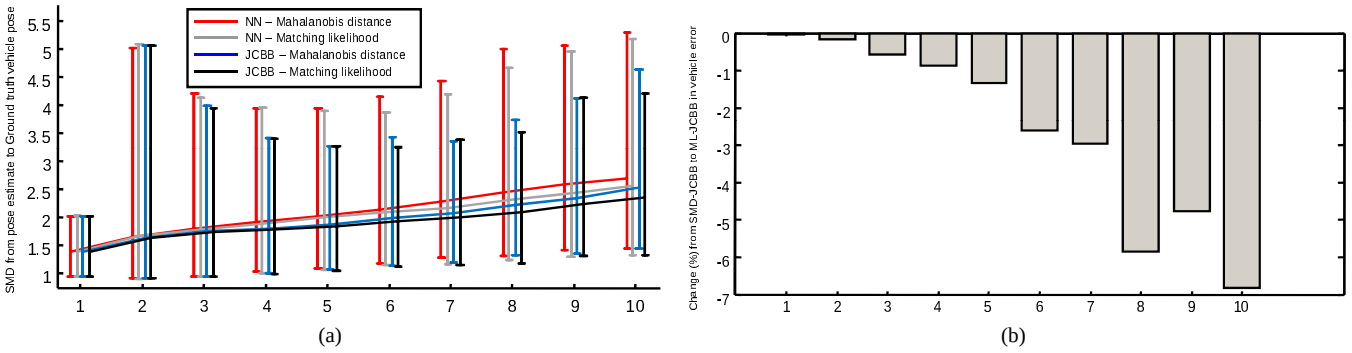


Fig. 4. Statistics about the error in the vehicle positioning for the different DA methods in the corridor-like scenario. Horizontal axes stand for increasing levels of sensor noise, with the same meaning than that of Fig. 3. (a) The mean and 5%-95% confidence intervals of the squared Mahalanobis distance between the EKF estimate of the vehicle 2D pose, i.e. (x, y, ϕ), and the ground truth. (b) Visualization of the same data only for the JCBB method and as the percentage reduction entailed by using NLML instead of SMD for DA.

now stands for different frequencies of outliers, and the statistics are the result of more than 3000 runs for each DA method.

The ratio of cases where SMD is preferable is here so reduced that the corresponding bar segment is barely visible in the graphs. About in $\sim 99\%$ of the observations the NLML leads to identical or better correspondences than SMD, with a $\sim 40\%$ of them being actually better. Furthermore, for the corridor-like scenario this ratio raises to over $\sim 60\%$, confirming once more that the advantage of NLML increases with the hardness of the DA problem. These results strongly support the main claim of this paper. Regarding the reduction of false positives, in this case NLML+JCBB reduces an average of $\sim 0\%$ to $\sim 16\%$ of the overall wrong associations with respect to SMD+JCBB.

C. SLAM with real data

The benchmarks above are based on a simulated sensor in synthetic environments due to the need for a data association ground truth. In order to completely rule out the possibility of our simulator being the source of any bias that favors NLML we also performed an analysis of 2D range-bearing EKF-based SLAM for a real dataset. This experiment employs the DLR dataset, released by J. Kurlbaum and U. Frese [27], which includes the ground truth for the data association of each landmark observation. We selected a sequence of 634 odometry-observation pairs extracted from the "DLR dataset number e", taking care of including one loop closure[6].

The average number of false positives obtained for each of the four DA methods are 10.89%, 10.45%, 10.82% and 8.02% for the SMD+NN, NLML+NN, SMD+JCBB and NLML+JCBB methods, respectively. Obviously, in this case a statistical comparison is not in order since there exists only one sequence of odometry and sensory data. These results are in perfect concordance with all those obtained in simulations: JCBB improves data association in comparison to NN, as it does when using NLML instead of SMD.

D. Effects in vehicle positioning error

The main criteria to evaluate the reliability of a DA method should be the rate of false positives. All other statistics, as false negatives or vehicle positioning errors, are only a consequence of them. Nevertheless, an interesting question is that of at what extent do in practice false positives affect the error in the vehicle positioning whose estimation is, after all, one of the ultimate purposes of SLAM.

[6]The obtained map, the estimated robot path, source code and instructions to reproduce this experiment can be found in the multimedia attachment.

We believe that it is not possible to categorically determine how false positives affect the error in the pose estimates, since this complex issue is determined by the sparsity of landmarks, their specific geometrical arrangement in the environment and, most importantly, whether wrong associations occur or not during the critical instants of a loop closure, in which case the negative effects would be magnified. Therefore, the best we can do here is to provide an insight on the effects of DA on the vehicle error from our benchmarks, without pretending that the results could be extrapolated.

In order to numerically quantify the vehicle error we have evaluated the Mahalanobis distance from the EKF estimate (including its uncertainty) to the known ground truth for the entire paths of the robot for all the simulation-based benchmarks above. Then, we have computed the reduction in this error for the JCBB method between using the SMD and the NLML. The most relevant conclusions we can obtain from our benchmarking are: (i) the error reduction is somewhat small (only $\sim 0.2\%$) for the benchmarks with spurious readings, (ii) considerably more relevant (about $\sim 1\%$ to $\sim 5\%$) for the benchmarks of noisy sensors and (iii) in any case, the gain, although small, typically grows as the hardness of the associations does. As an example of the expected improvement we provide the statistics in Fig. 4 for the particular case of the corridor-like scenario.

VII. Discussion

With this work we have provided a review of the literature on DA in SLAM from the point of view of the criteria employed for deciding against multiple feasible, but mutually incompatible, pairings. Although there exists a sound statistic (the NLML) to ground this search for correspondences we found out that most previous works employed the SMD instead. We have also provided a founded discussion exposing the differences between the SMD and the NLML, including why the SMD is *almost* as good classifier as the NLML in many situations. Comprehensive SLAM experiments with both simulated and real data have verified that, indeed, the usage of NLML reduces the ratio of false positives, potentially catastrophic errors for SLAM, for both NN and JCBB. The improvements are small, but unquestionable.

Therefore, our conclusion is that SMD should not be employed anymore when deciding data association: NLML always leads to (statistically) better pairings, apart from being, conceptually, a more sound statistic. As it was also shown, its added computational cost with respect to SMD is negligible.

It would be interesting to see other authors verifying our results in the future, not only for 2D range-bearing SLAM, but for other problems which must cope with data association. From the discussion

in this paper, we predict that the improvement due to the NLML will be more evident for those problems where the determinants of the innovation covariance span over wider ranges.

Apart from these facts, the modification of existing DA algorithms in order to employ the ML criteria opens at least two interesting new research fronts, which will be briefly mentioned here but would deserve thorough investigations in the future. Firstly, regarding RBPF-based SLAM [21] with landmarks, dubbed FastSLAM [22], [23], the existing methods solve DA at each particle independently. Provided that a method such as JCBB, when using ML, builds a tree where the N most likely correspondence hypotheses can be efficiently retrieved, a natural extension to FastSLAM would be the replication of each particle into several ones, each carrying a different DA hypothesis. The key point of using ML instead of SMD here is that the ML value would seamlessly integrate into the computation of the weights for those new particles. Interestingly, this idea would be a dual version of the EKF-based (non RBPF) approach proposed in [28].

A second interesting issue is the following. Any of the DA methods discussed in section II, when faced to two alternative matching hypotheses having a different number of pairings, always prefers the one with more paired features. In general, this leads to good decisions but one could devise an extra term, a *uniform* distribution, that would allow us to compare hypotheses with different numbers of pairings in a more probabilistically-grounded form. This term can be easily added to NLML while cannot be integrated into SMD, limited to Gaussian distributions.

REFERENCES

[1] H. Durrant-Whyte and T. Bailey, "Simultaneous localisation and mapping (slam): Part i the essential algorithms," *Robotics and Automation Magazine*, vol. 13, no. 2, pp. 99–110, 2006.

[2] T. Bailey and H. Durrant-Whyte, "Simultaneous localisation and mapping (SLAM): Part II-State of the art," *Robotics and Automation Magazine*, vol. 13, pp. 108–117, 2006.

[3] S. Thrun, *Robotic Mapping: A Survey*. School of Computer Science, Carnegie Mellon University, 2002.

[4] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics*. The MIT Press, September 2005.

[5] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *Computer*, vol. 22, no. 6, pp. 46–57, 1989.

[6] P. Pfaff, R. Triebel, C. Stachniss, P. Lamon, W. Burgard, and R. Siegwart, "Towards Mapping of Cities," in *International Conference on Robotics and Automation*, pp. 4807–4813, 2007.

[7] A. Lilienthal, M. Reggente, M. Trincavelli, J.-L. Blanco, and J. González, "A statistical approach to gas distribution modelling with mobile robots - the kernel dm+v algorithm," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2009.

[8] G. Grisetti, R. Kummerle, C. Stachniss, and W. Burgard, "A tutorial on graph-based slam," *IEEE Intelligent Transportation Systems Magazine*, vol. 2, no. 4, pp. 31–43, 2010.

[9] M. Dissanayake, P. Newman, S. Clark, H. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localization and map building (SLAM) problem," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 3, pp. 229–241, 2001.

[10] G. Grisetti, C. Stachniss, and W. Burgard, "Improved Techniques for Grid Mapping With Rao-Blackwellized Particle Filters," *IEEE Transactions on Robotics*, vol. 23, pp. 34–46, Feb 2007.

[11] J. Neira and J. Tardós, "Data association in stochastic mapping using the Joint Compatibility test," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 6, pp. 890–897, 2001.

[12] A. Davison, I. Reid, N. Molton, and O. Stasse, "MonoSLAM: Real-Time Single Camera SLAM," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1052–1067, 2007.

[13] J. Blanco, J. Fernández-Madrigal, and J. Gonzalez, "Efficient Probabilistic Range-Only SLAM," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1017–1022, 2008.

[14] G. Kantor and S. Singh, "Preliminary results in range-only localization and mapping," *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on*, vol. 2, 2002.

[15] R. Smith, M. Self, and P. Cheeseman, "Estimating uncertain spatial relationships in robotics," *Autonomous Robot Vehicles*, vol. 1, pp. 167–193, 1990.

[16] M. W. M. G. Dissanayake, P. Newman, H. F. Durrant-Whyte, S. Clark, and M. Csorba, "An experimental and theoretical investigation into simultaneous localisation and map building," in *The Sixth International Symposium on Experimental Robotics VI*, (Sydney, Australia), pp. 265–274, Springer-Verlag, March 2000.

[17] J. Montiel, J. Civera, and A. Davison, "Unified Inverse Depth Parametrization for Monocular SLAM," in *Robotics: Science and Systems*, July 2006.

[18] Y. Bar-Shalom, *Tracking and data association*. Academic Press Professional, Inc. San Diego, CA, USA, 1987.

[19] T. Bailey, *Mobile robot localisation and mapping in extensive outdoor environments*. PhD thesis, Australian Center for Field Robotics, University of Sydney, 2002.

[20] T. Bailey, E. Nebot, J. Rosenblatt, and H. Durrant-Whyte, "Data association for mobile robot navigation: A graph theoretic approach," in *IEEE International Conference on Robotics and Automation*, vol. 3, pp. 2512–2517, 2000.

[21] A. Doucet, N. de Freitas, K. Murphy, and S. Russell, "Rao-Blackwellised particle filtering for dynamic Bayesian networks," in *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pp. 176–183, 2000.

[22] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM: A factored solution to the simultaneous localization and mapping problem," *Proceedings of the AAAI National Conference on Artificial Intelligence*, pp. 593–598, 2002.

[23] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges," in *Proceedings of the International Joint Conference on Artificial Intelligence*, 2003.

[24] M. Montemerlo, *FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem With Unknown Data Association*. PhD thesis, University of Washington, 2003.

[25] J. Neira, J. Tardós, and J. Castellanos, "Linear time vehicle relocation in SLAM," in *International Conference on Robotics and Automation (ICRA)*, vol. 1, pp. 427–433, 2003.

[26] M. Matsumoto and T. Nishimura, "Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator," *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, vol. 8, no. 1, pp. 3–30, 1998.

[27] J. Kurlbaum and U. Frese, "A benchmark data set for data association," 2008.

[28] D. Maksarov and H. Durrant-Whyte, "Mobile vehicle navigation in unknown environments: a multiple hypothesis approach," in *IEEE Proceedings on Control Theory and Applications*, vol. 142, pp. 385–400, IET, 2002.