Probabilistic Qualitative Localization and Mapping
Supplementary Material

Roee Mor and Vadim Indelman

This document provides supplementary material to the paper [2]. Therefore, it should not be considered a self-contained document, but instead regarded as an appendix of [2].

Throughout this report, all notations and definitions are with compliance to the ones presented in [2].

1 Extended geometrical analysis

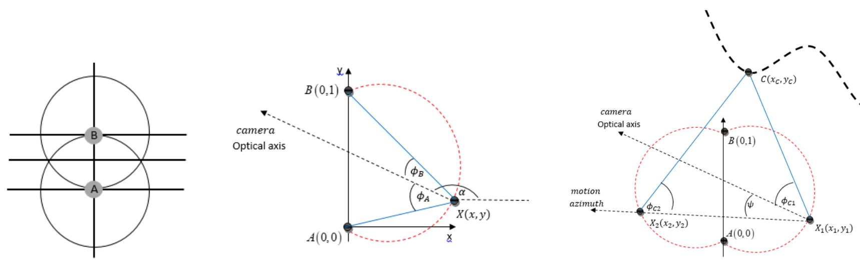
The basic metric geometrical problem underling our approach is determining camera poses $X_{1:n}^{AB}$ and landmark C location $L^{AB:C}$ given a set of measurements $Z_{1:n}^{ABC}$. In this section we analyze the geometric properties of our problem, and demonstrate the effect of incorporating a motion model on the quality of the solution. In [**?**, Section 3a] a simple degrees of freedom analysis is done to show the effect of incooperating a motion model to the inference. Here we will look at this effect in more detail.

In this section we consider the basic geometry of our problem, and do a metric deterministic analysis. Thus, for the sake of explanation, we consider the measurement and motion models ideal, noise-free observations that we shall denote by $\bar{Z}$.

First, we shall look at the single view case. Landmarks A,B are known (we work in the AB frame), while the camera pose and landmark C are unknown. Determining camera pose by 2D azimuth measurements to two known landmarks is a well known 2D 2-point camera resection problem. A full and efficient analytic solution to this problem can be found in [3]. Given these azimuth measurements, the camera's location must be on a *circle* that goes through the two landmarks A,B. The "locus circle" center and radius calculation is given in [3]. Azimuth ordering of ϕ_n^A, ϕ_n^B further confines the camera location to left

---

R. Mor is with the Department of Computer Science, Technion - Israel Institute of Technology, Haifa 32000, Israel. V. Indelman is with the Department of Aerospace Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel. roeeki.mor@gmail.com vadim.indelman@technion.ac.il. This work was partially supported by the Israel Ministry of Science & Technology (MOST).



(a) (b) (c)

Figure 1: (a) Extended double cross spatial partition sugested in [1]. (b) 2D 2-point camera resection. (c) camera resection and landmark triangulation with two measurements

or right to the AB vector, see Figure 1b. It is not difficult to show that camera orientation α_n is directly determined by its location:

$$\alpha_n = \text{atan}_2(y_n^X, x_n^X) + \phi_A = \text{atan}_2(y_n^X - 1, x_n^X) + \phi_B. \tag{1}$$

Per every possible camera pose, landmark C can be located somewhere on the line of sight corresponding to the azimuth measurement ϕ_n^C. A diagram illustrating these aspects is given in Figure 1b, while a simulative example can be seen in Figures 2a and 2b. A crucial insight is that even in this deterministic setting the problem is not fully observable, i.e. solutions for metric camera poses and landmark locations are continuously distributed. We denote this joint distribution by $\mathbb{P}(X_n^{AB}, L^{AB:C}|\bar{Z}_n^{ABC})$.

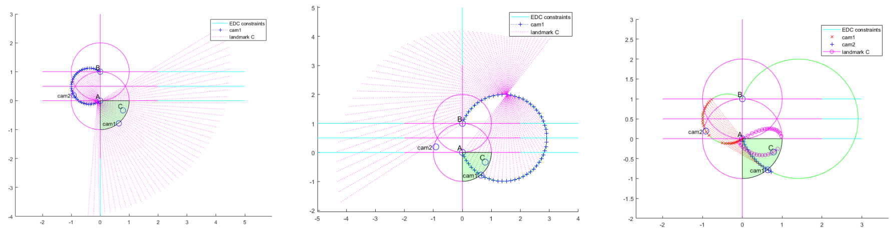
If measurements from several time instances are available, the joint pdf is:

$$\mathbb{P}(X^{AB}, L^{AB:C}|\bar{Z}_{1:n}^{ABC}) = \prod_{i=1}^{n} \mathbb{P}(X^{AB}, L^{AB:C}|\bar{Z}_i^{ABC}) \tag{2}$$

As described in [**?**], the posterior distribution over qualitative states can be extracted by integration.

Now we examine the impact of incorporating a motion model. Given a specific pose for the first camera $(x_1^*, y_1^*, \alpha_1^*)$, the action ψ_1, and the second measurement $Z_2 = \{\phi_2^A, \phi_2^B, \phi_2^C\}$ obtained after executing the action, the second camera location (x_2^*, y_2^*) can be calculated as the intersection of the line of motion with the valid part of the second measurement locus circle (see Figure 1c). As this intersection can occur once, twice or not at all. As a result, some camera poses are disqualified. Another consequence is geometric ambiguity: some measurements can support two solutions for landmark location and the corresponding second camera pose.

Given the two camera poses $(x_1^*, y_1^*, \alpha_1^*), (x_2^*, y_2^*, \alpha_2^*)$ and azimuth measurements ϕ_1^C, ϕ_2^C to landmark C, its location (x_c^*, y_c^*) can be triangulated. Considering all possible poses for the first camera, the mapping of landmark C is reduced to a curve (might be split into two curves) - see Figure 1c. A simulation



(a) **(b)** **(c)**

Figure 2: Simulation example for first camera resection and landmark estimation. True camera positions, and landmark positions marked in circles.(a) First view. Estimated camera position marked by blue +. Estimated landmark C location marked in magenta (.) - see legend. (b) Second view. Same markings. (c) Estimation by both views using motion model. Estimated camera position marked by +,x. Estimated landmark C location marked in magenta circles (see legend).

example that displays how our motion model allows to disqualify certain camera poses, and triangulate the landmark C can be seen in Figure 2c.

When considering three or more measurements, triangulation consistency further disqualifies most camera poses and landmark locations, leaving only one or few discrete possible locations for landmark C, and the corresponding camera trajectories. The surviving landmark C locations are in fact an intersection of the curve estimates for pairs of consecutive views.

These geometric properties of the problem are the main motivation of our work. The deterministic analysis reveals that incorporating a motion model into the formulation improves results in the metric problem. Since probability for qualitative states is an integration on the metric pdf, it should improve as well. We also notice that the geometric ambiguity enables several distinct solutions to the problem. Therefore probabilistic solutions based on linearizion might converge to local minima and achieve bad results. A global solution, or a hypothesis based approach is therefore needed.

2 detailed algorithm

As described in [**?**], our approach addressees each triplet of landmarks viewed together at time steps $1, \dots, n$. The landmarks are noted as A, B, C. Given the motion model for each transition, our goal is to infer the posterior probabilities of landmark C qualitative state, $\mathbb{P}(S^{AB:C}|H_n)$, and camera qualitative trajectory $\mathbb{P}(S^{AB:X}_{1:n}|H_n)$, both in a landmark centric AB frame. In this section, we will describe the details of our algorithm for estimation the qualitative states.

Our approach and considerations are specified in [**?**]. In short, we first solve the metric SLAM problem $\mathbb{P}(X_{1:n}, L_C|H_n)$ using a sample based method. The result is a set of valid samples of landmark C location L^C, and camera poses $X^{AB}_{1 \dots n}$ in the AB frame, and a probability for each sample. Then we calculate the qualitative state probabilities by summing over our weighted samples.

For solving this SLAM problem, usually camera frame is used by fixing

4 degrees of freedom: first camera pose, and scale taken to be the distance between first to second cameras. We use a different landmark centric frame by fixing landmarks $A = (0, 0), B = (0, 1)$. It is more intuitive to infer qualitative landmark related states this way. A deeper analysis and an efficient analytic solution to this problem geometry are given in [3], and discussed in 1.

We start by sampling the first camera pose X_1^{AB} from the probability $\mathbb{P}(X_1^{AB}|\phi_1^A, \phi_1^B)$ using azimuth measurements to landmarks A,B. This step handles both the fact that we have no prior knowledge, and the measurement noise. These samples are denoted as $\hat{X}_{1,k}^{AB}$ where i is time index, $k \in 1...m_1$ is sample index and m_1 is the number of camera pose samples. If we have measurements from only one time step for a specific landmark triplet, we sample landmark C location L^C for each camera pose from $\mathbb{P}(L^C|X_1^{AB}, \phi_1^C)$.

If we have measurements from multiple time steps we proceed by iteratively implement a three step algorithm for each time step. We define "trajectory hypothesis" $th_j = \{\hat{X}_{1,k1}^{AB}, ..., \hat{X}_{i,ki}^{AB}\}$ as a set of camera poses - one from each time step. $TH = \{th_j\}$ is the set of all trajectory hypothesis. We initialize TH to be the set of the sampled camera poses from time step 1 - $th_j = \hat{X}_{1,j}^{AB}$ (so each trajectory hypothesis contains only one camera pose). Now we iteratively implement three steps for each time step:

Sampling step: sample camera pose $\hat{X}_{i,j}^{AB}; j = 1...m_i$ in the same way, using azimuth measurements to landmarks A,B in time i. Now we build a new trajectory hypothesis set TH. For each existing trajectory hypothesis from time $i-1$, we generate m_i new ones by adding each new $\hat{X}_{i,j}^{AB}$ sample to it. TH is now m_i times bigger.

Motion step: Use action a_{i-1} (in our case, the motion azimuth ψ_{i-1}) and motion model to test the probability $\mathbb{P}(\psi_{i-1}|\hat{X}_{i,j}^{AB}, \hat{X}_{i-1,j}^{AB})$ for each trajectory hypothesis in time steps $i, i-1$. Low probability hypothesis are dismissed, and removed from TH. We also keep motion model weight $wm_{j,i} = \mathbb{P}(\psi_i - 1|\hat{X}_{i,j}^{AB}, \hat{X}_{i-1,j}^{AB})$ for each hypothesis.

Resection step: For each valid trajectory hypothesis th_j, we test consistency of bearing measurements from all cameras to landmark C. First, we use camera poses $\hat{X}_{1...i,j}^{AB}$ and bearing measurements $\psi_{1...i}^C$ to triangulate landmark C location. Trajectory hypothesis that don't intersect, are disqualified. For each camera set that do intersect, landmark C location $\hat{L}_j^C$ is estimated to be the centroid of all line of sight pairs intersection points. Then we estimate the probability for this configuration as: $wr_j = \prod_{k=1}^{i} \mathbb{P}(\phi_k^C|\hat{L}_j^C, \hat{X}_{k,j}^{AB})$ (assuming independent measurement noise). This is actually a consistency test that indicates if the camera pose set is consistent with landmark C bearing measurements. Again, we disqualify low probability sets, and keep a resection weight wr_j for each valid trajectory hypothesis.

Note: using a single sample for landmark C location as the line of sights intersection centroid for each camera pose set is heuristic. We can uniformly sample over a larger area to cover all probable locations. We choose the simpler way for run-time considerations.

Finally, we sum over all valid camera pose sets in each qualitative state to get an estimate for qualitative state probability distribution:

$$\mathbb{P}(s_k^C|H_n) \approx \sum_{j;\hat{L}_j^C \in s_n^C} wr_j \prod_{k=1}^{n} wm_{j,k}. \tag{3}$$

The camera qualitative state can be similarly inferred:

$$\mathbb{P}(s_k^{X_i}|H_n) \approx \sum_{j;\hat{X}^A B_i, j \in s_n^{X_i}} wr_j \prod_{k=1}^{n} wm_{j,k}. \tag{4}$$

This algorithm seems to handle a large number of trajectory hypothesis that grows exponentially in time. Practically, the geometric constraints enforced in the "motion" step and the "resection" step dramatically decreases the number of hypothesis for two views, and leave a very small number of consistent hypothesis for three views or more. This makes our algorithm actually very fast, and is possible only due to the fact that we introduce the use of a motion model. It also means that using this algorithm incrementally, requires to save only a small number of hypothesis, and therefore is not memory intensive. A pseudo-code for a simplified version of this approach is given in Algorithm 1.

Algorithm 1 single triplet probabilistic qualitative estimation

sample camera pose$\hat{X}_{1,k}^{AB}; k = 1 : m_1$ from $\mathbb{P}(X_1^{AB}|\phi_1^A, \phi_1^B)$
initialise trajectory hypothesis set: $TH = \{\hat{X}_{1,k}^{AB}\}$
for $i = 2, ..., n$ **do**
  sample camera pose$\hat{X}_{i,k}^{AB}; k = 1 : m_i$ from $\mathbb{P}(X_i^{AB}|\phi_i^A, \phi_i^B)$
  extend TH with new samples
  estimate motion model consistency: $wm_{j,i} = \mathbb{P}(\psi_{i-1}|\hat{X}_{i,j}^{AB}, \hat{X}_{i-1,j}^{AB})$
  dismiss low probability hypothesis
  estimate resection consistency: $wr_j = \prod_{k=1}^{i} \mathbb{P}(\phi_k^C|\hat{L}_j^C, \hat{X}_{k,j}^{AB})$
  dismiss low probability hypothesis
end for
estimate landmark qualitative state probability by 3
estimate camera qualitative state probability by 4

3 extended result analysis

In [?] we present part of the results to best explain the our main insights. Actually we used more metrics, and various noise level instances to get a good understanding of the meaning of the results. In this section we give some additional details to support out conclusions.

3.1 Result Metrics

In our work [?] we use various metrics to analyses inference performance. This is actually an important addition for QSR SLAM, since it gives us deeper understanding than what was presented in previous works. We use two types of metrics:

ground truth (GT) related metrics:

- Probability DMSE: This is the mean square error for the difference between estimated qualitative state vector to GT state vector $DMSE = \sqrt{\sum_{i=1}^{m} (s_i^{AB:C} - s_{GT}^{AB:C})^2}$. It tests how accurate is the estimation.
- Ground truth rating: This is the position of the GT qualitative state when ordering qualitative states by their estimation likelihood (1 is most likely)
- Geometric distance: The mean distance from qualitative state centroids to GT state centroind weighted by the state estimated probability. $gmd = \sum_{i=1}^{m} ||c_i^{AB:C} - c_{GT}^{AB:C}||_2$ This metric tells us how close are the estimated states to GT (1 is the distance between A,B).
- Ground truth likelihood: estimated posterior probability of GT qualitative state $\mathbb{P}(s_{GT}^{AB:C})$. This measures the accuracy of estimation, but ignores false qualitative states.
- Ground truth likelihood ratio: the ration between estimated probability of GT qualitative state to estimated probability of the most likely state $\frac{\mathbb{P}(s_{GT})}{max(S^{AB:C})}$. This metric measures how close the GT state is to being most likely.

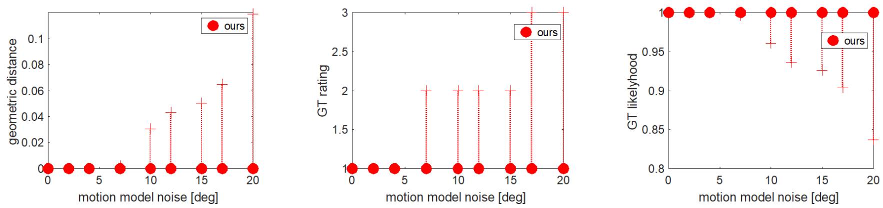
information (or entropy) related metrics:

- Entropy: estimation probability entropy $E = -\sum_{i=1}^{m} \mathbb{P}(s_i^{AB:C} \log s_i^{AB:C})$ measures how distributed is the qualitative state distribution (or how much information is in the distribution).
- Likelihood ration: the ration between the second most likely qualitative state to the most likely qualitative state. This measure how "decisive" the estimation result is. very important to maximal likelihood approaches.

Using all these metrics in different stages of result analysis teaches us much about the quality of our algorithms.

3.2 Motion Model Inference Extended Results

After noticing that the qualitative localization and mapping fast approximation algorithm presented in our main paper [?] gives results close to the full algorithm, we made a more extensive analysis of the approximated algorithm performance. In this section we'll test the different effects of motion model and



(a) **(b)** **(c)**

Figure 3: Sensitivity to motion model error. The plot shows median and percentiles 25 and 75. (a) mean geometric distance Vs motion model noise. (b) ground truth rating Vs motion model noise (c) ground truth likelihood Vs motion model noise.

azimuth noise separately. We use the same framework, and scenarios as specified in [**?**], but compare different sets of noise levels.

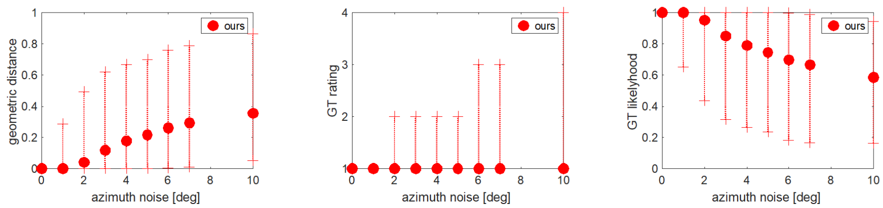
To test motion model noise effect on our estimation, we use scenarios with measurement noise 0, and various motion model noise levels. results are presented in figure 3. It can be seen that motion model noise affect weekly ground truth likelihood and geometric distance. It does however have more effect on the ground truth rating (but only on the 75 percentile).

This is an intuitive result. Scenarios with harder geometry (landmark close to qualitative state edges, or ill conditioned camera poses) are the first to suffer from the motion model errors. It gives more weight to false qualitative states, but does not significantly change probability distribution. Generally the estimation handles motion model noise well. Even with very high noise of 20°, the estimated probabilities are just minorly affected. GT rating is affected, but GT is still on the top 3 (and most likely in most cases - median).

To test measurement model noise effect on out estimation, we use scenarios with motion model noise 0, and various measurement model noise levels. results can be seen in figure 4. The effect of measurement noise is stronger. Both geometric distance, and ground truth likelihood respond harder as noise increases. GT rating responds a little harder on 75 percentile but stable on median.

Generaly it seem that estimation handles the noise pretty well in most scenarios. We note for camera based platforms a measurement noise of up to 2° is very reasonable. under these conditions estimation handles noise well.

In summary, it seems that our approximated qualitative inference algorithm is more sensitive to measurement model noise than to motion model noise. This result is encouraging, since measurement noise is usually smaller in real world scenarios. In addition, we note response to high levels of noise is gradual and generally stable. In addition, we also see that our algorithm can handle reasonable amounts of noise well. Remembering that this is a low compute algorithm, this is another demonstration to support our conclusion that our method has potential to be practical for qualitative autonomous navigation and



(a) **(b)** **(c)**

Figure 4: Sensitivity to measurement model error. The plot shows median and percentiles 25 and 75. (a) mean geometric distance Vs measurement model noise. (b) ground truth rating Vs measurement model noise (c) ground truth likelihood Vs measurement model noise.

mapping, for online qualitative active planning. A more comprehensive analysis should be done to evaluate more realistic scenarios.

References

[1] Mark McClelland, Mark Campbell, and Tara Estlin. Qualitative relational mapping for mobile robots with minimal sensing. *Journal of Aerospace Information Systems*, 11(8):497–511, 2014.

[2] R. Mor and V. Indelman. Probabilistic qualitative localization and mapping. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2020. Submitted.

[3] Vincent Pierlot and Marc Van Droogenbroeck. A new three object triangulation algorithm for mobile robot positioning. *IEEE Transactions on Robotics*, 30(3):566–577, 2014.