Robotic Vision Summer School
Fusing map data

February 3, 2020

To finish finally identify the location of the animals, there are two key analysis problems that must be resolved. The first is to register the SLAM maps obtained by the different robots. That is the process of aligning the data generated by each of the robots so that they share a common reference and the information about the animals can be combined. The second is to estimate the position of an animal from multiple independent measurements of bearings taken at different points in the terrain. Both these problems involve combining information from different sources and can be solved using stochastic data fusion techniques. Such techniques are probably the most fundamental set of techniques in modern robotics.

1 Notation

1.1 Kroneker products

For two matrices $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{p\times q}$ then

$$A \otimes B = \begin{pmatrix} A_{11}B & \cdots & A_{1m}B \\ \vdots & & \vdots \\ A_{n1}B & \cdots & A_{nm}B \end{pmatrix} \in \mathbb{R}^{np\times mq}$$

1.2 SLAM configuration state

The state of a differential drive robot is often modelled as position (x, y) and orientation θ. That is the state $\xi = (x, y, \theta) \in \mathbb{R}^3$ where we unwrap the angle variable θ from the circle $[0, 2\pi)$ to the real line. This state representation is nice for formulating the extended Kalman filter but poor for computing geometric transformations.

The geometric pose of a robot is represented[1] as an element of $P \in \mathbf{SE}(2)$ Fig. 1. This pose representation the position and orientation of the robot as a

[1] Using $\mathbf{SE}(3)$ to represent the pose is not formally correct, however, it is a convenient and accepted notation where the rotation columns encode the directions of the axes of the robot body-fixed frame of reference while translation column encodes the robot position.

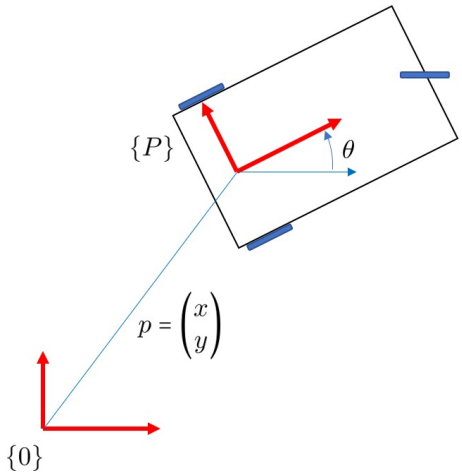

Figure 1: Geometric pose of a robot moving on a plane.

rigid transformation from an fixed reference frame $\{0\}$ to the robot body-fixed frame of reference $\{P\}$.

An element of $\mathbf{SE}(2)$ is written

$$P = \begin{pmatrix} R_P & x_P \\ 0 & 1 \end{pmatrix} \in \mathbf{SE}(2)$$

with $R_P \in \mathbf{SO}(2)$ a rotation and $x_P \in \mathbb{R}^2$ its location. The rotation matrix is written

$$R_P = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

where θ is the relative orientation angle of the robot.[2] Thus, the geometric pose P is constructed from the robot state (x, y, θ) by

$$P = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & x \\ \sin(\theta) & \cos(\theta) & y \\ 0 & 0 & 1 \end{pmatrix} \tag{1}$$

The inverse of a matrix $A \in \mathbf{SE}(2)$ is

$$A^{-1} = \begin{pmatrix} R_A^\top & -R_A^\top x_A \\ 0 & 1 \end{pmatrix}.$$

That is $AA^{-1} = I_3 = A^{-1}A$.

[2]This expression for rotation uses the robotics convention - representing the robots state from the perspective of a reference frame. The Computer Vision community often uses the inverse convention, representing an object viewed in the world from the perspective of the camera.

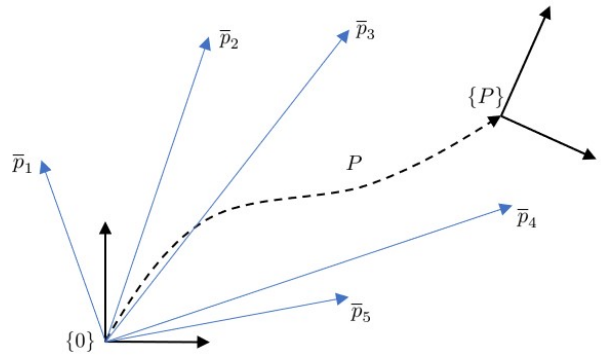

Figure 2: The state of a SLAM problem with 5 landmarks.

Landmarks are 2D points on the plane $p_i \in \mathbb{R}^2$ for $i = 1, \ldots, n$ represented with respect to the reference frame $\{0\}$. A typical state configuration including geometric robot pose and landmark points are shown in Figure 2

1.3 Rigid body transformations

A rigid-body transformation is a transformation of space that preserves distances and angles between all points. Rigid-body transformations consist of rotations and translations and a general rigid-body transformation of $\mathbb{R}^2$ is an element of $\mathbf{SE}(2)$. Landmarks transform under rigid body transformation by $A \in \mathbf{SE}(2)$ by

$$p \mapsto R_A p + x_A$$

The algebra of homogeneous coordinates are often used to simplify this notation. Let the homogeneous vector for $p \in \mathbb{R}^2$ be defined as

$$\overline{p} = \begin{pmatrix} p \\ 1 \end{pmatrix} \hookrightarrow \mathbb{R}^3$$

where we use the hook right arrow to indicate the embedding of homogeneous coordinates into $\mathbb{R}^3$, the vector has three entries. With this algebra one has

$$\overline{R_A p + x_A} = \begin{pmatrix} R_A p + x_A \\ 1 \end{pmatrix} = \begin{pmatrix} R_A & x_A \\ 0 & 1 \end{pmatrix} \begin{pmatrix} p \\ 1 \end{pmatrix} = A\overline{p}$$

or $\overline{p} \mapsto A\overline{p}$.

The matrix $P \in \mathbf{SE}(3)$ representing the position and orientation of a robot will also transform by a rigid-body transformation $A \in \mathbf{SE}(2)$ by

$$P \mapsto AP.$$

That is the old P is translated and rotated by A to a new pose AP.

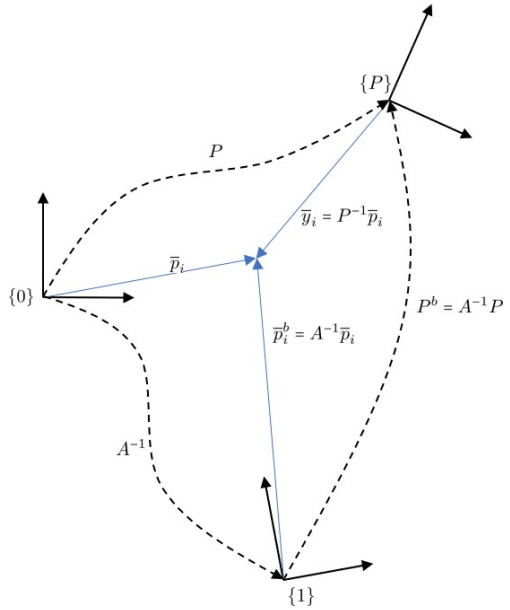

Figure 3: A gauge transform is visualised as a transformation of the reference frame that changes coordinates but doesn't move the robot or landmark points. The rigid body transformation $A : \{1\} \mapsto \{0\}$ and maps $\overline{p}_i^b \mapsto A\overline{p}_i^b = \overline{p}_i$ and $P^b \mapsto AP^b = P$.

2 Fusing two maps

The geometric robot pose $P(t) \in \mathbf{SE}(2)$ and the landmark points $\overline{p}_i \in \mathbb{R}^2$ are represented with respect to a reference frame $\{0\}$ that is implicit in the coordinate formulation. This reference frame, however, has no physical ground truth in the SLAM problem. Indeed, if $(P, \overline{p}_1, \ldots, \overline{p}_n)$ are consistent with your measurements, then for any rigid-body transformation $A \in \mathbf{SE}(2)$ one has that $(AP(t), A\overline{p}_1, \ldots, A\overline{p}_n)$ are also consistent with the measurements. That is translating and rotating the robot and the landmark points by the same transformation is consistent with all the observations. This property is termed an *invariance* of the SLAM problem, and the transformations $A \in \mathbf{SE}(2)$ are known as *gauge transformation*.

Although gauge transforms do not change the relative information contained in a SLAM configuration estimate, they do change the coordinates of a configuration. Thus, for two configurations $(P, \overline{p}_1, \ldots, \overline{p}_n)$ and $(AP(t), A\overline{p}_1, \ldots, A\overline{p}_n)$ that represent the same physical configuration of robot and landmark points, $P \neq AP$ and $\overline{p}_i \neq A\overline{p}_i$. A useful visualisation of this difference is two consider two different coordinate representations of the same physical configuration of robot and landmarks. The different coordinates can be related to each other

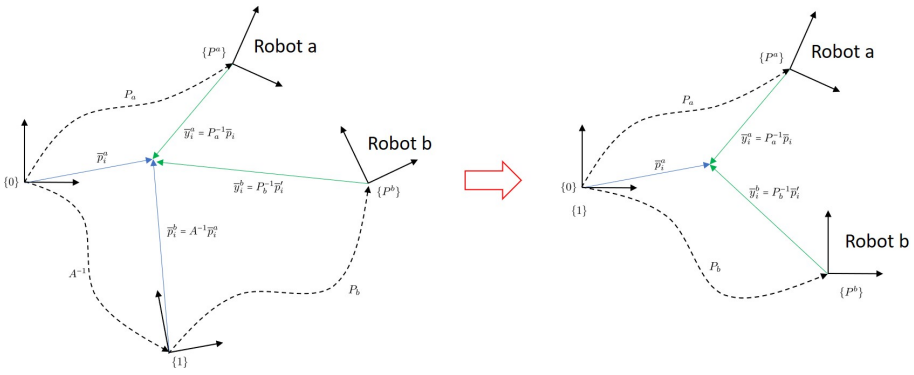

Figure 4: A figure demonstrating map registration for two robot SLAM. The maps on the left are of the same physical points but derived from SLAM algorithms running on different robots with different measurements. The coordinates of the SLAM solution differ by a rigid-transformation A that is visualised as a gauge transformation A^{-1} of the reference frame $\{0\} \to \{1\}$. The right hand side shows the situation after the rigid body transformation $A \in \mathbf{SE}(2)$ is applied to the reference $\{1\}$, as the SLAM configuration $(P_b, \overline{p}_i^b)$. In the registered coordinates the map points correspond. The robot poses are different since the two robots were not collocated.

by applying a gauge transformation to the reference frame by A^{-1} as shown in Figure 3.

This relationship is key to understand when fusing maps from two different robots. When two maps are produced by different robots, the two reference frames implicit in the coordinates used by the robots will be different Fig. 4. The coordinates of the map points estimated by the SLAM algorithm will be different. To combine the data from both robots one needs to *register* the two maps so that they share the same reference frame. Registering the maps refers to computing the relative gauge transform between the reference frames and transforming the coordinates of both maps into a single shared reference frame. In Figure 4, the reference frame from robot a was chosen without loss of generality as common frame, and the gauge transformation A computed that transforms $\{1\}$ to $\{0\}$. This gauge transform is applied to the SLAM solution $(P^b, \overline{p}_i^b) \mapsto (AP^b, A\overline{p}_i^b)$. In the new registered coordinates then $\overline{p}_i^a = A\overline{p}_i^b$ at least up to noise in the estimates. The robot poses $AP^b \neq P^a$ of course since the two robots were not co-located.

2.1 Registering two maps without noise

Let us begin with the case where there is no noise in either map. That is consider two maps $\{\overline{p}_1^a, \ldots, \overline{p}_n^a\}$ and $\{\overline{p}_1^b, \ldots, \overline{p}_n^b\}$ and ask the question: Find $A \in \mathbf{SE}(2)$

such that

$$A\overline{p}_i^b = \overline{p}_i^a, \text{ for } i = 1, \ldots, n. \tag{2}$$

This problem is linear in the unknown $A \in \mathbf{SE}(2)$, although the nonlinearity of the Lie-group $\mathbf{SE}(2)$ does complicate matters.

Expressing (2) as

$$R_A p_i^b + x_A - p_i^a = 0 \tag{3}$$

then it follows that $x_A = p_i^a - R_A p_i^b$ for all $i = 1, \ldots, n$. Since we know that there will be noise even though at the moment we are not modelling it, we will also average over all landmarks

$$x_A = \frac{1}{n}\sum_{j=1}^{n}(p_j^a - R_A p_j^b) = \frac{1}{n}\sum_{j=1}^{n} p_j^a - R_A \frac{1}{n}\sum_{j=1}^{n} p_j^b = p_{\text{av}}^a - R_A p_{\text{av}}^b \tag{4}$$

where

$$p_{\text{av}} := \frac{1}{n}\sum_{j=1}^{n} p_j$$

since we have used the overbar notation for homogeneous coordinates.

Substituting (4) into (3) one obtains

$$R_A(p_i^b - p_{\text{av}}^b) = (p_i^a - p_{\text{av}}^a), \quad \text{for } i = 1, \ldots, n. \tag{5}$$

Define data matrices

$$M_a = \begin{pmatrix}(p_1^a - p_{\text{av}}^a) & \cdots & (p_n^a - p_{\text{av}}^a)\end{pmatrix} \in \mathbb{R}^{2\times n} \tag{6}$$

$$M_b = \begin{pmatrix}(p_1^b - p_{\text{av}}^b) & \cdots & (p_n^b - p_{\text{av}}^b)\end{pmatrix} \in \mathbb{R}^{2\times n} \tag{7}$$

then

$$R_A M_b = M_a \tag{8}$$

As long as $\text{rank}(M_b) = 2$ then this equation has a unique solution

$$R_A = M_a M_b^\top (M_b^\top M_b)^{-1} \tag{9}$$

If there is no noise in the data, then this matrix will be orthogonal $R_A \in \mathbf{SO}(2)$.

In practice, it is always a good idea to make the resulting matrix orthogonal. Take the SVD of the data

$$M_a M_b^\top (M_b^\top M_b)^{-1} = U^\top \Lambda V$$

for $U, V \in \mathbf{SO}(2)$ and Λ diagonal. If the data is ideal then $\Lambda = I_2$ is the identity matrix. If not then make it so, that is, set $R_A = U^\top V$. Of course you should check that the singular values (diagonal entries of Λ) are close to zero. If they are not close to zero, it indicates that the two maps do not represent the same physical landmark points and this flags an error.

Algorithm 1: Registration of ideal maps.

Result: Compute (R_A, x_A) to register ideal maps.
Choose thresh for error criteria ;
Compute x_A (4) ;
Compute M_a and M_b (6) (7) ;
Compute the SVD $U^\top \Lambda V = M_a M_b^\top (M_b^\top M_b)^{-1}$. ;
if $\|\Lambda - I_2\|^2 >$ thresh **then**
  Error: "maps are not consistent" ;
  return
end
Set $R_A = U^\top V$, $x_A = x_A$.

2.2 Registration of a noisy map to known data

In §2.1 we discussed registering two maps assuming ideal data. A real robot never generates a perfect map. Indeed, the standard SLAM algorithms generate both an estimate and a covariance, or estimate of uncertainty for the map. Indeed, these algorithms are fundamentally stochastic in nature, and the correct interpretation is that they generate an information state for a landmark variable. That is the estimate of the landmark variable is the Gaussian probability distribution with mean (the estimate) and variance (the estimate uncertainty). In order to register noisy data to a ground truth the ideal equations in §2.1 do not hold and a stochastic approach is appropriate.

Formulating the maximum likelihood cost: Consider a known landmark configuration $\{p_1^a, \ldots, p_n^a\}$ and a set of noisy data generated by a SLAM algorithm $\{\hat{p}_1^b, \ldots, \hat{p}_n^b\}$

$$\hat{p}_i^b \sim N(p_i^b, \Sigma_i^b).$$

where $N(p_i^b, \Sigma_i^b)$ is the Gaussian with mean[3] p_i^b and (positive definite) variance $\Sigma_i^b \in \mathbb{R}^{2\times 2}$. This is typically written as an explicit *generative noise model*

$$\hat{p}_i^b = p_i^b + \eta_i^b, \quad \eta_i^b \sim N(0, \Sigma_i^b). \tag{10}$$

The underlying problem is the same: Find $A \in \mathbf{SE}(2)$ such that

$$A\overline{p}_i^b \approx \overline{p}_i^a.$$

however now we do not have p_i^b directly and we need to consider some sort of approximate solution. To begin with we must define what sort of approximation is appropriate. The most common approach is to use maximum likelihood.

Applying A^{-1} to p_i^a is equivalent to applying A to p_i^b. Thus, from (10) one can write

$$R_A^\top p_i^a - R_A^\top x_A - \hat{p}_i^b = -\eta_i^b \sim N(0, \Sigma_i^b)$$

[3]We will assume that the SLAM algorithm is unbiased, that is that the mean of the Gaussian from which the estimate is drawn is the true map.

That is $R_A^\top p_i^a - R_A^\top x_A - \hat{p}_i^b \sim N(0, \Sigma_i^b)$. Note that the left hand side depends on unknown parameters R_A and x_A. The *likelihood* of a data sample, in this case the evaluation of $R_A^\top p_i^a - R_A^\top x_A - \hat{p}_i^b$ is the evaluation of the probability distribution $N(0, \Sigma_i^b)$. The *maximum likelihood* solution is obtained by choosing the parameters R_A and x_A to maximize the likelihood of the data!

Now, for a 2 dimensional Gaussian distribution

$$N_{(0,\Sigma_i^b)}(\zeta) = \frac{1}{2\pi \det(\Sigma_i^b)^{\frac{1}{2}}} \exp\left(-\zeta^\top (\Sigma_i^b)^{-1} \zeta\right) \tag{11}$$

where $\zeta \in \mathbb{R}^2$ is just a placeholder variable here. Maximising (11) is difficult due to the exponential non-linearity in the Gaussian distribution function. However, the log function is monotonic increasing to maximising $\log N_{(0,\Sigma_i^b)}(\eta_i^b)$ is the same as maximising $N_{(0,\Sigma_i^b)}(\eta_i^b)$. Moreover, multiplying by -1 makes the problem a minimization, which will match classical least squares methods. Thus, we consider minimizing $-\log N_{(0,\Sigma_i^b)}(\eta_i^b)$ over parameters R_A and x_A. That is

$$\begin{aligned}(\hat{R}_A, \hat{x}_A) &= \underset{(\hat{R},x)}{\operatorname{argmin}} \left((\eta_i^b)^\top (\Sigma_i^b)^{-1} (\eta_i^b) + \log(2\pi \det(\Sigma_i^b)^{\frac{1}{2}})\right) \\ &= \underset{(R_A,x_A)}{\operatorname{argmin}} (\eta_i^b)^\top (\Sigma_i^b)^{-1} (\eta_i^b)\end{aligned}$$

where $\eta_i^b = R_A^\top p_i^a - R_A^\top x_A - \hat{p}_i^b$. Note that the term $\log(2\pi \det(\Sigma_i^b)^{\frac{1}{2}})$ can be discarded from the optimisation since it does not depend on the parameters $(\hat{R}, x)$. Define the scaled least squares norm to be

$$\|\zeta\|_{\Sigma_i^b}^2 := \zeta^\top (\Sigma_i^b)^{-1} \zeta$$

and note that this problem can be written

$$\hat{R}_A, \hat{x}_A = \underset{(R_A,x_A)}{\operatorname{argmin}} \|R_A^\top p_i^a - R_A^\top x_A - \hat{p}_i^b\|_{\Sigma_i^b}^2.$$

This is a least squares problem with respect to a norm that is scaled by the covariance Σ_i^b.

The above discussion leads to a nice formulation for a single data point, in this case a single map estimate of a landmark. In fact, generalising to multiple landmarks is straightforward. The output of a SLAM system is a vector state that combines all the variables of the system, typically

$$\xi = (x, y, \theta, p_1^x, p_1^y, \ldots, p_n^x, p_n^y) \in \mathbb{R}^{3+2n}$$

The covariance estimate is one large matrix $\Sigma \in \mathbb{R}^{(3+2n)\times(3+2n)}$. That is the state $\xi \sim N(\hat{\xi}, \Sigma)$ is a multi-variate Gaussian in $3 + 2n$ dimensions. It is important to maintain the full state structure since the estimates of the landmarks and robot state are highly correlated. The map registration problem depends only on the landmark variables of the state and the robot pose variables can be *marginalised*

out. For a Gaussian distribution, the marginal obtained by integrating out for the robot state is just the map part of the state $m = (p_1, \ldots, p_n) \in \mathbb{R}^{2n}$ corresponding to the n landmarks along with the $2n \times 2n$ partition of Σ^m corresponding to the map. The only tricky part is acting on the measurements with R_A and x_A as this must be done landmark by landmark. This can be vectorised by forming the following Kroneker products

$$I_n \otimes R_A = \begin{pmatrix} R_A & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & R_A \end{pmatrix}, \quad \mathbf{1}_n \otimes x_A = \begin{pmatrix} x_A \\ \vdots \\ x_A \end{pmatrix}$$

where I_n is the $n \times n$ identity and $\mathbf{1}_n$ is the n-vector with ones in each entry. Then maximum likelihood solution can be written as a function of the full map coordinates as a nonlinear least squares problem

$$(\hat{R}_A, \hat{x}_A) = \underset{(R_A, x_A)}{\operatorname{argmin}} \|(I_n \otimes R_A)^\top \mathbf{m}_a - (I_n \otimes R_A)^\top (\mathbf{1}_n \otimes x_A) - \hat{\mathbf{m}}_b\|^2_{\Sigma_b}. \tag{12}$$

where $\mathbf{m}_a = (p_1^a, \ldots, p_n^a)$ and $\hat{\mathbf{m}}_b = (\hat{p}_1^b, \ldots, \hat{p}_n^b)$.

Solving the least squares problem: The next question is how to solve (12). This is no longer a simple algebraic solution of a linear system of equations. The least squares problem, when weighted by a general Σ_b matrix is quadratic in the variables R_A and x_A and in addition, $R_A \in \mathbf{SO}(2)$ is a rotation matrix. This problem must be solved iteratively by an optimisation algorithm. The standard tool used is a *Levenberg–Marquardt algorithm*, this is a modification of the Gauss–Newton algorithm[4] modified specifically for cost functions of the nonlinear least squares form (12). The Levenberg-Marquardt algorithm is based on linearising the argument of the nonlinear least squares cost around a point, and then solving the resulting quadratic least squares problem in the linearised variable.

The first step is to locally parameterise the rotation matrix R_A. Let[5]

$$\mathfrak{e} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Then the matrix exponential of $\mathfrak{e}$ (skew symmetric) is a rotation matrix $\exp(\theta\mathfrak{e}) \in \mathbf{SO}(2)$. Indeed,

$$\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} = \exp(\theta\mathfrak{e})$$

We only need to parameterize R_A locally around a previous iterate of the optimization algorithm $R_A(k)$ so we write

$$R_A = R_A(k) \exp(\delta\mathfrak{e})$$

[4]It also has a trust region modification to improve its basin of attraction that will not be discussed in these notes.

[5]The matrix $\mathfrak{e}$ is the generator for the matrix representation of the Lie-algebra $\mathfrak{so}(2)$.

for $\delta \in \mathbb{R}$ a small angle. There is still the nonlinear dependence on δ through the exponential function. Thus, finally we linearize around $\delta = 0$ to get

$$R_A \approx R_A(k)(I_2 + \delta\mathfrak{e}). \tag{13}$$

Note that right-hand side is not actually an orthogonal matrix, however, it is this linearisation that is required to compute the Levenberg-Marquardt update. Similarly, we write a local linear approximation for x_A around an estimate $x_A(k)$ by

$$x_A = x_A(k) + \epsilon. \tag{14}$$

This local model is already linear and does not require linearisation.

The solution to the nonlinear least squares problem (12) can now be written locally around an estimate $(R_A(k), x_A(k))$ as

$$(\hat{R}_A, \hat{x}_A) = (R_A(k)\exp(\mathfrak{e}\delta_k), x_A(k) + \epsilon_k) \tag{15}$$

where

$$\begin{aligned}(\delta_k,\epsilon_k) = \underset{(\delta,\epsilon)}{\operatorname{argmin}} \|&(I_n \otimes R_A(k))^\top \mathbf{m}_a + (I_n \otimes R_A(k))^\top (\mathbf{1}_n \otimes x_A(k)) - \hat{\mathbf{m}}_b \\ &- \delta(I_n \otimes \mathfrak{e}R_A(k))^\top \mathbf{m}_a - \delta(I_n \otimes \mathfrak{e}R_A(k))^\top (\mathbf{1}_n \otimes x_A(k)) + \mathbf{1}_n \otimes R_A^{\otimes}(k)^\top \epsilon\|^2_{\Sigma_b}.\end{aligned} \tag{16}$$

This equation is obtained by substituting (13) and (14) into (12) and cancelling the quadratic $\epsilon\delta$ term. Although equation (16) looks horrible it is in fact very simple. Define data structures

$$a(k) = (I_n \otimes R_A(k))^\top \mathbf{m}_a + (I_n \otimes R_A(k))^\top (\mathbf{1}_n \otimes x_A(k)) - \hat{\mathbf{m}}_b \in \mathbb{R}^{2n} \tag{17}$$

$$b(k) = (I_n \otimes \mathfrak{e}R_A(k))^\top \mathbf{m}_a + (I_n \otimes \mathfrak{e}R_A(k))^\top (\mathbf{1}_n \otimes x_A(k)) \in \mathbb{R}^{2n} \tag{18}$$

$$C(k) = \mathbf{1}_n \otimes R_A(k)^\top \in \mathbb{R}^{2n\times 2} \tag{19}$$

Then the nonlinear least squares problem at step k of the iteration can be written

$$(\delta_k, \epsilon_k) = \underset{(\delta,\epsilon)}{\operatorname{argmin}} \|a(k) - \delta b(k) + C(k)\epsilon\|^2_{\Sigma_b}. \tag{20}$$

The solution of this is just least squares in the two linear variables δ and ϵ with respect to the scaled norm $\|\cdot\|^2_{\Sigma_b}$.

Algorithm 2: Levenberg Marquardt for map registration

Result: Compute $(\hat{R}_A, \hat{x}_A)$ maximum likelihood estimates for map registration of noisy data to known landmarks

Fix a required accuracy a_{thresh} ;
Choose $(R_A(1), x_A(1))$;
while $\|a(k)\|^2_{\Sigma_b} > a^2_{thresh}$ **do**
  Compute $a(k)$, $b(k)$ and $c(k)$ from (17), (18), (19) ;
  Solve (20) for (ϵ_k, δ_k). ;
  Compute $(R_A(k+1), x_A(k+1)) = (R_A(k)\exp(\mathfrak{e}\delta_k), x_A(k) + \epsilon_k)$;
end
Set $(\hat{R}_A, \hat{x}_A) = (R_A(k), x_A(k))$;

The Levenberg-Marquardt algorithm solves the (20) recursively. Each iteration finds the exact optimum of the linearised problem and then subsequent linearisation is undertaken closer to the true optimum and contains information to improve the approximation. The iteration is terminated when the cost of the estimated solution

$$\|a(k)\|^2_{\Sigma_b} = \|(I_n \otimes R_A(k))^\top \mathbf{m}_a + (I_n \otimes R_A(k))^\top (\mathbf{1}_n \otimes x_A(k)) - \hat{\mathbf{m}}_b\|^2_{\Sigma_b}$$

is small, less than a fixed threshold a^2_{thresh}.

Of course, like any locally convergent algorithm, the initial guess $(R_A(1), x_A(1))$ needs to be pretty good before you start the algorithm, or it will diverge. A good choice for $(R_A(1), x_A(1))$ is to solve the ideal case, using the ideal case algorithm 1 then use this as a starting point. The ideal case will always compute something that is not too bad, although you may need to set the error threshold fairly high to start with and look at deleting outliers if the maps are not consistent.

2.3 Registering and fusing two noisy maps

> There is data you can trust and data you should distrust but 'true data' is an oxymoron.[6]

The reality in robotics is that the concept of ground truth is nonsensical. All estimates are noisy in some sense, even if they are what you may have thought was ground truth. In the case of fusing two maps generated by separate robots, both maps come with explicit estimates of their uncertainty and the problem is one of both registering and fusing the stochastic information.

Formulating the maximum likelihood cost: Consider two noisy maps generated by robots a) and b). That is $\hat{m}_a = (\hat{p}^a_1, \dots, \hat{p}^a_n)$ and $\hat{m}_b = (\hat{p}^b_1, \dots, \hat{p}^b_n)$

$$\hat{m}_a \sim N(\mathbf{m}_a, \Sigma_a), \quad \hat{m}_b \sim N(\mathbf{m}_b, \Sigma_b).$$

[6] Mahony 2020 .

Here we jump straight to representing the map as $\mathbb{R}^{2n}$ objects with correlated variance. The underlying problem is the same: Find $A \in \mathbf{SE}(2)$ such that

$$(I_n \otimes R_A)\mathbf{m}_b + \mathbf{1}_n \otimes x_A \approx \mathbf{m}_a$$

using the Kroneker notation for the vector structure of the full map. The ideal map coordinates $\mathbf{m}_a$ and $\mathbf{m}_b$ are not available, and an exact solution is impossible. We must formulate a criteria to define which approximation for the map registration parameters (R_A, x_A) are appropriate, but also find an approximate solution for the true map $\hat{\mathbf{m}}_\star$ under some appropriate criteria.

Since there is not ground truth, we must start by choosing a reference. Choose robot a) as the reference system. That is, choose the reference in which we will fuse the data to be the arbitrary reference associated with robot a) SLAM solution. Which is to say the "true" map is $\mathbf{m}_\star = \mathbf{m}_a$.

It follows that $\hat{\mathbf{m}}_a \sim N(\mathbf{m}_\star, \Sigma_a)$ and we can consider a coupled set of equations[7]

$$\begin{aligned}(I_n \otimes R_A)\mathbf{m}_b + \mathbf{1}_n \otimes x_A &\approx \mathbf{m}_\star \\ \mathbf{m}_a &= \mathbf{m}_\star\end{aligned}$$

This may appear to be sleight of hand, however, the point is that we now treat $\mathbf{m}_\star$ as a parameter. In particular, applying the inverse and rearranging on has

$$(I_n \otimes R_A)^\top \mathbf{m}_\star - (I_n \otimes R_A)^\top(\mathbf{1}_n \otimes x_A) - \hat{\mathbf{m}}_b \sim N(0, \Sigma_b) \tag{21}$$

$$\mathbf{m}_\star - \hat{\mathbf{m}}_a \sim N(0, \Sigma_a). \tag{22}$$

Note particularly the trick to get the estimate $\hat{\mathbf{m}}_a$ and $\hat{\mathbf{m}}_b$ as independent terms here so that the Gaussian distributions generated by the SLAM maps can be used directly. The consequence of this trick is that the parameters get combined in a nonlinear manner $(I_n \otimes R_A)^\top \mathbf{m}_\star$.

The resulting nonlinear least squares problem is obtained by taking the log-likelihood of (21) and (22) to get

$$\begin{aligned}(\hat{R}_A, \hat{x}_A, \hat{\mathbf{m}}_\star) = \underset{(R_A, x_A, \mathbf{m}_\star)}{\operatorname{argmin}} \big(&\|(I_n \otimes R_A)^\top \mathbf{m}_\star - (I_n \otimes R_A)^\top(\mathbf{1}_n \otimes x_A) - \hat{\mathbf{m}}_b\|^2_{\Sigma_b} \\ &+ \|\mathbf{m}_\star - \hat{\mathbf{m}}_a\|^2_{\Sigma_a} \big)\end{aligned} \tag{23}$$

Solving for the data fusion problem: Solving for the optimum using the Levenberg-Marquardt involves the same construction as we undertook for the early noisy registration process.

Recall the linearisations (13) and (14) around the point $(R_A(k), x_A(k))$. The additional parameter is $\mathbf{m}_\star$. We can write

$$\mathbf{m}_\star = \mathbf{m}_\star(k) + \mu$$

for $\mu \in \mathbb{R}^{2n}$ small.

[7] This trick is equivalent to applying total least squares.

Define data structures

$$a(k) = (I_n \otimes R_A(k))^\top \mathbf{m}_\star(k) + (I_n \otimes R_A(k))^\top (\mathbf{1}_n \otimes x_A(k)) - \hat{\mathbf{m}}_b \in \mathbb{R}^{2n} \quad (24)$$

$$b(k) = (I_n \otimes \mathfrak{e} R_A(k))^\top \mathbf{m}_\star(k) + (I_n \otimes \mathfrak{e} R_A(k))^\top (\mathbf{1}_n \otimes x_A(k)) \in \mathbb{R}^{2n} \quad (25)$$

$$C(k) = \mathbf{1}_n \otimes R_A(k)^\top \in \mathbb{R}^{2n \times 2} \quad (26)$$

$$D(k) = (I_n \otimes R_A(k))^\top \in \mathbb{R}^{2n \times 2n} \quad (27)$$

and note that the linearised least squares can now be written locally around an estimate $(R_A(k), x_A(k), \mathbf{m}_\star(k))$ as

$$(\hat{R}_A, \hat{x}_A, \hat{\mathbf{m}}_\star) = (R_A(k) \exp(\mathfrak{e}\delta_k), x_A(k) + \epsilon_k, \mathbf{m}_\star(k) + \mu_k) \quad (28)$$

where

$$(\delta_k, \epsilon_k, \mu_k) = \underset{(\delta,\epsilon,\mu)}{\operatorname{argmin}} \|a(k) - \delta b(k) + C(k)\epsilon + D(k)\mu\|^2_{\Sigma_b} + \|\mathbf{m}_\star(k) + \mu - \hat{\mathbf{m}}_a\|^2_{\Sigma_a}. \quad (29)$$

This is a straightforward linear least squares problem and can be solved with standard matrix techniques.

Algorithm 3: Levenberg Marquardt for map fusion

Result: Compute $(\hat{R}_A, \hat{x}_A), \hat{m}_\star)$ maximum likelihood estimates for fusing two noisy maps

Fix a required accuracy a_{thresh} ;

Choose $(R_A(1), x_A(1), \mathbf{m}_\star(1))$;

while $\|a(k)\|^2_{\Sigma_b} > a^2_{thresh}$ **do**

  Compute $a(k)$, $b(k)$, $C(k)$ and $D(k)$ from (24), (25), (26), (27) ;

  Solve (29) for $(\epsilon_{k+1}, \delta_{k+1}, \mu_{k+1})$;

  Compute $(R_A(k+1), x_A(k+1), \mathbf{m}_\star(k+1)) =$
  $(R_A(k) \exp(\mathfrak{e}\delta_k), x_A(k) + \epsilon_k, \mathbf{m}_\star(k) + \mu_k)$;

end

Set $(\hat{R}_A, \hat{x}_A, \hat{\mathbf{m}}_\star) = (R_A(k), x_A(k), \mathbf{m}_\star(k))$;

The output from this process will both estimate the maximum likelihood registration parameters (R_A, x_A) as well as the maximum likelihood map estimates $\mathbf{m}_\star$. These parameters are estimated with respect to the reference frame for robot a).

3 Estimating position of an animal from bearing measurements

In this section, we assume that all the data has been registered into a single reference frame. Thus, the question of locating an animal in the terrain is one of fusing a collection of separate noisy measurements of the animals position into a single "best" estimate of the animal position. The problem is slightly more

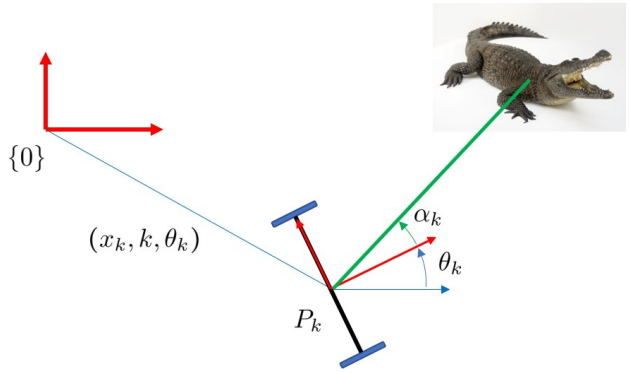

Figure 5: The various states representing the kth measurement bearing measurement of a target (crocodile) as seen from a robot. The robot state is (x_k, y_k, θ_k) and the crocodile bearing is α_k.

complex due to the fact that we only have bearing measurements of the animals position and we will have to triangulate their position from noisy measurements. Figure 5 shows the measurement process.

3.1 Ideal triangulation: computing position from a collection of bearings

Consider a collection of $k = 1, \ldots, N$ measurements of the bearing to a target (in this case an animal) taken from a moving robot. Each measurement consists of two parts, the robot pose (x_k, y_k, θ_k) and the relative bearing from the robot to the animal α_k as shown in Fig. 5. Recall that we write the geometric pose P_k (1) corresponding to the robot state and that one has $x_{P_k} = (x_k, y_k)$ and

$$R_{P_k} = \begin{pmatrix} \cos(\theta_k) & -\sin(\theta_k) \\ \sin(\theta_k) & \cos(\theta_k) \end{pmatrix}$$

For the moment we will assume that these measurements are ideal, or noise free.

Consider the bearing of the target written as a unit vector $y_k \in \mathrm{S}^2$ from the robot in the direction of the target. Then

$$y_k = \begin{pmatrix} \sin(\alpha_k) \\ \cos(\alpha_k) \end{pmatrix}$$

Let $\xi \in \mathbb{R}^2$ denote the location of the target in the reference frame. A model for the direction of the target in terms of the robot state is

$$y_k = \frac{R_{P_k}^{\mathsf{T}}(\xi - x_{P_k})}{\|\xi - x_{P_k}\|} \tag{30}$$

By using this model then is possible to estimate for ξ from multiple measurements y_k. Define an orthogonal bearing

$$z_k = \begin{pmatrix} \cos(\alpha_k) \\ -\sin(\alpha_k) \end{pmatrix}$$

and note that $z_k^\top y_k = 0$. Now multiply (30) by $z_k^\top$ and one has

$$z_k^\top y_k = 0 = \frac{z_k^\top R_{P_k}^\top (\xi - x_{P_k})}{\|\xi - x_{P_k}\|}$$

or equivalently

$$z_k^\top R_{P_k}^\top (\xi - x_{P_k}) = 0$$

Define data

$$a_k = z_k^\top R_{P_k}^\top \in \mathbb{R}^{1\times 2} \tag{31}$$

$$b_k = z_k^\top R_{P_k}^\top x_{P_k} \in \mathbb{R} \tag{32}$$

and note that

$$a_k \xi = b_k$$

In the ideal case one needs two measurements in order to solve this set of linear equations. In general, we will use all the measurements by forming a data matrices

$$A = \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} \in \mathbb{R}^{N\times 2}, \qquad b = \begin{pmatrix} b_1 \\ \vdots \\ b_N \end{pmatrix} \in \mathbb{R}^{N\times 1} \tag{33}$$

and solving the linear equations

$$A\xi = b. \tag{34}$$

In the ideal case this stack of equations is consistent. In the real world of course, that will not be the case and a 'least squares' solution is a reasonable choice to generate a feasible solution

$$\xi = (A^\top A)^{-1} A^\top b. \tag{35}$$

That is ξ is the location of the target expressed in the reference frame as computed from N bearing measurements and poses.