Uncertainty 2:

Combining Uncertainty

1 Combining Uncertainties

• The basic theory behind everything to do with the Kalman Filter can be obtained from two sources:

- 1) Uncertainty transformation with the Jacobian matrix
- 2) The Central Limit Theorem

1.1 Variance of a Sum of Random Variables

• We can use our uncertainty transformation rules to compute the variance of a sum of random variables. Let there be n random variables - each of which are normally distributed with the same distribution:

$$x_i \sim N(\mu, \sigma) \quad , \quad i = 1, n$$

• Let us define the new random variable y as the sum of all of these.

$$y = \sum_{i=1}^{n} x_i$$

• By our transformation rules:

$$\sigma_y^2 = J \sigma_x^2 J^T$$

• where the Jacobian is simply n ones:

$$J = \begin{bmatrix} 1 & \dots & 1 \end{bmatrix}$$

• Hence:

$$\sigma_y^2 = n\sigma_x^2$$

• Thus the variance of a sum of identical random variables grows linearly with the number of elements in the sum.

1.2 Variance of A Continuous Sum

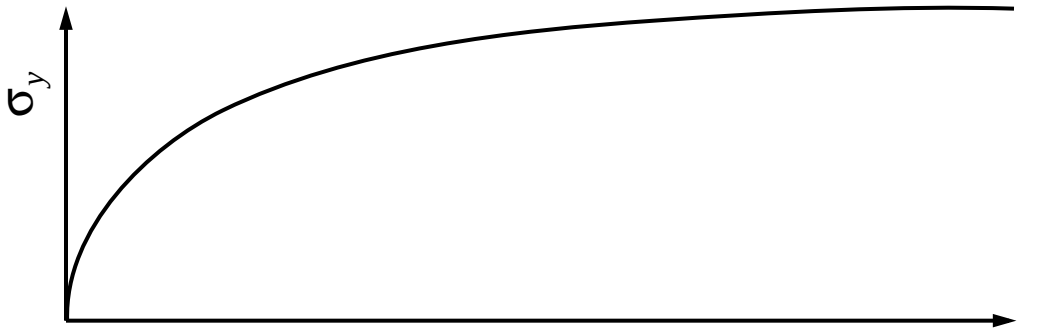
• Consider a problem where an integral is being computed from noisy measurements. This will be important later in dead reckoning models.

• In such a case where a new random variable is added at a regular frequency, the expression gives the development of the standard deviation versus time because:

$$n = \frac{t}{\Delta t}$$

in a discrete time system.

• In this simple model, uncertainty expressed as standard deviation ***grows with the square root of time***.



• Uncertainty grows rapidly and then levels off as time evolves. This arises from the fact that truly random errors tend to cancel each other if enough of them are added.

1.3 Central Limit Theorem

• This important theorem has several important interpretations. One was mentioned already. The sum of a large number of independent random variables is Gaussian - regardless of the distribution of each.

• The CLT also describes the so-called “sampling distribution of the mean” which is discussed below.

• Let $(x_1, x_2, x_3, \ldots)$ be a random sample of size n from a normal distribution having mean μ and variance σ^2.

• Then if $\bar{x}$ is the sample mean, as $n \to \infty$ the distribution of $\bar{x}$ is the standard normal distribution with mean μ and variance σ^2/n.

$$\bar{x} \sim N(\mu, \sigma/\sqrt{n})$$

• Intuitively, this means that the most likely sample mean is the population mean and that the approximation is likely (not guaranteed) to get better as the sample size increases[1].

• We have a different use for this.

1.4 Combined Observations of a Constant

• Consider the results of the CLT when the sample size n is taken as a variable. The CLT directly tells us that the uncertainty of the sample mean decreases with more measurements. That is, to get a better answer, measure the same thing over and over again and take the average.

• Suppose several redundant measurements of a constant are obtained and that they all have identical statistics. Then the true

1. Hence the opinion poll which predicts the behavior of 300 million people based on interviewing a few thousand.

value can be approximated as the mean of the observations. Under these circumstances, the uncertainty in this mean is:

$$y = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \frac{1}{\sigma_y}^2 = \sum_{i=1}^{n} \frac{1}{\sigma_{x_i}}^2 = \frac{n}{\sigma_{x_i}^2} \Rightarrow \sigma_y^2 = \frac{1}{n}\sigma_x^2$$

• Thus, the variance in the mean decreases with increasing numbers of measurements.

1.5 Variance of A Continuous Measurement of a Constant

• Consider a problem where a constant value must be estimated from a sequence of noisy measurements.

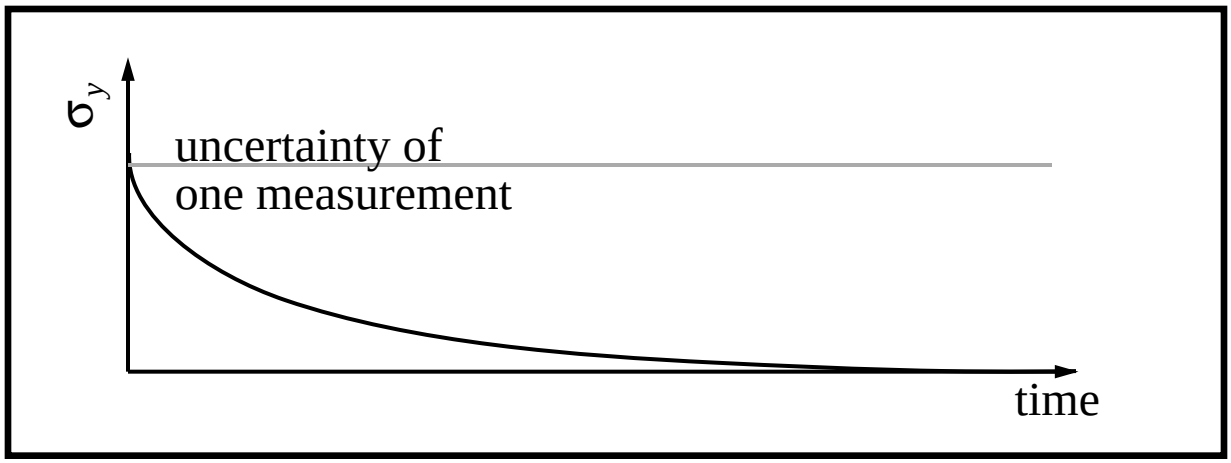
• In such a case where a new measurement is taken at a regular frequency, the expression gives the development of the standard deviation versus time because:

$$n = \frac{t}{\Delta t}$$

in a discrete time system.

• In this simple model, uncertainty expressed as standard deviation ***decreases with the square root of time***. This idea that taking and merging multiple observations ***reduces the***

uncertainty of the combined result is the basic idea of the Kalman filter.



• Uncertainty decreases rapidly and then levels off as time evolves.

2 Compounding Uncertainty in Measurement Sequences

• Our transformation rules give all the tools necessary to compute the uncertainty in a continuously changing quantity that is computed in an arbitrary way from a continuous sequence of measurements.

• This is a more general combination of measurements than a continuous sum. The measurements may be added or combined in more complicated ways.

• We can determine the uncertainty in a general combination of sequential measurements by considering the result at any point to be a function of:

- The last answer and its uncertainty
- The current measurement and its uncertainty.

2.1 General Formulation

• Generally, let a function depend on a parameter vector $\underline{x}$ that we have decided to partition into two smaller parameter vectors of possibly different length:

$$\underline{y} = f(\underline{x}) \qquad \underline{x} = \begin{bmatrix} \underline{x}_1 & \underline{x}_2 \end{bmatrix}^T$$

• Let us also partition the covariance C_{xx} and Jacobian J_x of the inputs:

$$J_x = \begin{bmatrix} J_1 & J_2 \end{bmatrix} \qquad C_{xx} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

• where:

$$C_{ij} = Exp(\underline{x}_i \underline{x}_j^T)$$

• Then, the covariance of y is:

$$C_{yy} = \begin{bmatrix} J_1 & J_2 \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \begin{bmatrix} J_1^T \\ J_2^T \end{bmatrix}$$

• This can be written out in gory detail, but to save work, let us make the assumption that $\underline{x}_1$ and $\underline{x}_2$ are uncorrelated:

$$C_{12} = C_{21} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \quad \therefore \quad C_{yy} = \begin{bmatrix} J_1 & J_2 \end{bmatrix} \begin{bmatrix} C_{11} & 0 \\ 0 & C_{22} \end{bmatrix} \begin{bmatrix} J_1^T \\ J_2^T \end{bmatrix}$$

• Note that the size of the input correlation matrix is the length of the parameters $\underline{x}$, but the output is determined by the size of $\underline{y}$.

• Intuitively, two variables are uncorrelated when knowing the value of one tells you nothing at all about the value of the second, on average. That is, they are not co - related.

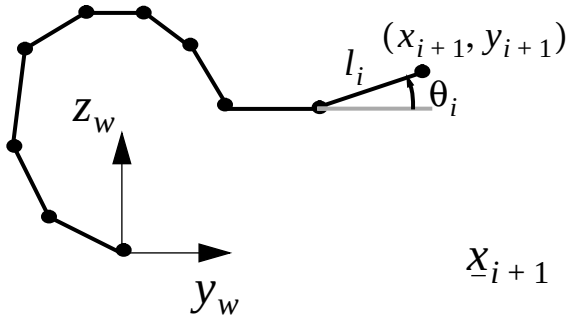
• This assumption then allows us to simply write the result as the sum of the individual transformed uncertainties.

$$C_y = J_1 C_{11} J_1^T + J_2 C_{22} J_2^T$$

2.2 Example: Dead Reckoning with Odometer Error Only

• Consider a case where vehicle position is generated from dead reckoning and a Gaussian error exists on the odometer. The heading is assumed determined from a perfect compass.

• This case can be formulated in terms of two vectors, the current position $\underline{x}_i$ and the current measurements $\underline{z}_i$, which when combined generate the new position $\underline{x}_{i+1}$:



$$\underline{x}_i = \begin{bmatrix} x_i & y_i \end{bmatrix}^T \qquad \underline{z}_i = \begin{bmatrix} l_i & \theta_i \end{bmatrix}^T$$

$$\underline{x}_{i+1} = \underline{f}(\underline{x}_i, \underline{z}_i) = \begin{bmatrix} x_i + l_i \cos(\theta_i) \\ y_i + l_i \sin(\theta_i) \end{bmatrix}$$

• The Jacobians are:

$$J_x = \frac{\partial \underline{x}_{i+1}}{\partial \underline{x}_i} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad J_z = \frac{\partial \underline{x}_{i+1}}{\partial \underline{z}_i} = \begin{bmatrix} c_i & -l_I s_i \\ s_i & l_I c_i \end{bmatrix}$$

• Let the uncertainties in the current position and the measurements (for a perfect compass) be:

$$C_i = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{bmatrix}_i \qquad C_z = \begin{bmatrix} \sigma_l^2 & 0 \\ 0 & 0 \end{bmatrix}$$

• So, for uncorrelated error (i.e. if the error in the new measurements has nothing to do with the current position), the error in the new position can be written as:

$$C_{i+1} = J_x C_i J_x^T + J_z C_z J_z^T$$

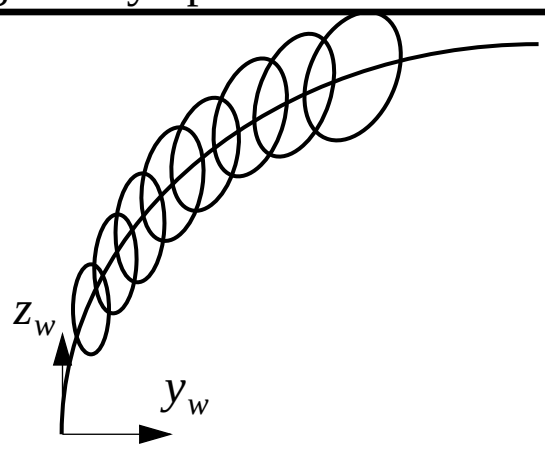
$$C_{i+1} = C_i + \begin{bmatrix} c_i^2\sigma_l^2 & c_i s_i \sigma_l^2 \\ c_i s_i \sigma_l^2 & s_i^2 \sigma_l^2 \end{bmatrix}$$

• This gives the uncertainty of the new position estimate. Note in particular that:

  • The cross terms will remain zero for motion along the axes.

  • The diaginal terms are always positive so the uncertainty increases monotonically.

• For example, in the following run of such a system, along a slowly turning path, the uncertainty grows without bound. Also, the ellipses slowly turn to the right but because they remember the contributions of all past measurments, they remain pointing mostly upward.



3 Merging Redundant Measurements

• Sometimes, we have several measurements that are directly related to the same thing. These can be used to get the best overall estimate. This process of getting the best overall estimate is called **estimation** and the tools used are called **estimators**.

• Most of robotics currently employs the Kalman filter to estimate **static** quantities - though it was originally developed to estimate dynamically changing quantities. This section deals with the static case and the more general dynamic case is covered later.

• Unfortunately there are at least a half dozen forms of Kalman Filter equations out there. This section will walk through some in order from simple to complex. All Kalman Filters have these characteristics:

- They estimate recursively
- They combine multiple measurements

• Other options in the formulation include:

- Weighted estimation or not (R or I matrix?)
- Direct and indirect measurement (H or I matrix?)
- Linear and Nonlinear versions (Hx or h(x))
- System Dynamics Model (F matrix?)

3.1 Batch Estimation

3.1.1 Sample Mean

• The most straightforward example of merging data is computing the mean of a set of measurements. Let a set of n measurements $\underline{z}_i$ exist of the (unknown) constant $\underline{x}$ called the

state vector. The sample mean can be computed from all measurements in one shot as follows:

$$\hat{\underline{x}}_k = \frac{1}{k}\sum_{i=1}^{k} \underline{z}_i$$

3.1.2 Weighted Batch Estimator

• By computing an average for the best estimate we have implicitly assumed that the measurements all have the same uncertainty. If they do not, a weighted average is the best estimate. One technique for computing the best estimate of measurements of varying uncertainty is weighted least squares.

• The problem is now generalized as follows:

  • The **state** $\underline{x}$ and **measurements** $\underline{z}$ are now vectors.

  • The measurements are indirectly related to the state vector through a measurement matrix H.

  • The measurements are assumed to be corrupted by a random noise vector $\underline{v}$ of known covariance R:

$$\underline{z} = H\underline{x} + \underline{v} \qquad R = Exp(v\underline{v}^T)$$

• Here, we attempt to minimize the total weighted squared residual of all of the measurements[1]:

$$\Gamma = (\underline{z} - H\hat{\underline{x}})^T R^{-1} (\underline{z} - H\hat{\underline{x}})$$

1. Note that this is a matrix quadratic form - which results in a scalar. It is not like the outer product used to form a covariance matrix.

where we have weighted the elements of the residual by their reciprocal variances (i.e. those measurements which are very uncertain do not contribute much to the estimate).

• Setting the derivative[1] of this with respect to the state estimate to zero, there results:

$$2HR^{-1}(\underline{z} - H\underline{\hat{x}}) = 0$$

$$(H^T R^{-1} H)\underline{\hat{x}} = H^T R^{-1}\underline{z}$$

$$\underline{\hat{x}} = (H^T R^{-1} H)^{-1} H^T R^{-1}\underline{z}$$

• Note that setting R to the identity matrix results in the usual **normal equations** of (unweighted) linear least squares.

3.2 Recursive Estimation

3.2.1 Simple Merging - Recursive Sample Mean

• More generally, the measurements may not all be available at once. In this case, a recursive formulation for the estimate is possible by rewriting the above scalar mean as:

$$\underline{\hat{x}}_{k+1} = \frac{k}{(k+1)}\underline{\hat{x}}_k + \frac{1}{(k+1)}\underline{z}_k$$

• or as:

$$\underline{\hat{x}}_{k+1} = \underline{\hat{x}}_k + \frac{1}{(k+1)}[\underline{z}_k - \underline{\hat{x}}_k]$$

• When k is large $\underline{z}_k$ has less and less effect on the answer.

1. Note that the derivative of $\underline{x}^T A\underline{x}$ wrt x is $2A\underline{x}$. Don't forget the 2.

• Where the difference between the current estimate and the current measurement $z_k - \hat{x}_k$ is known commonly as the **measurement residual**.

• Let σ^2 be the variance of a single measurement and $\sigma_k^{2'}$ be the total variance of the current estimate at step k.

• The equation for updating the uncertainty is, from the Central Limit Theorem:

$$\frac{1}{\sigma_{k+1}^2} = \frac{1}{\sigma_k^2} + \frac{1}{\sigma^2}$$

• One straightforward way to interpret this formulation is that the old estimate votes 'k' times for its value and the new measurement votes once for its value. It is instructive to rewrite this result as:

$$\hat{x}_{k+1} = \hat{x}_k + K[z_k - \hat{x}_k]$$

where K is called the **Kalman gain**. Note that, in this simple case, it is simply the reciprocal of the time step and it could be precomputed.

• The uncertainty propagation can be rewritten as:

$$\sigma_{k+1} = \left(1 - \frac{1}{1+k}\right)\sigma_k^2 = (1-K)\sigma_k^2$$

• These last two eqiuations are the "Maybeck form" of the Kalman Filter.

• This form of estimator is called a recursive filter. The major advantages of recursive filters are:

• They require virtually no memory because there is no longer a need to store all k previous measurements.

• They spread the computation out over time so that each new estimate requires only a few operations.

3.2.2 Interpretation as Weighted Average

• An as alternate view of what is happening here (see Maybeck), suppose that the measurement has an variance R_k and the state estimate has a variance P_k.

• We might further suppose that it would be reasonable to combine the current estimate and measurement in amounts inversely proportional to their own uncertainties thus (each has a weight proportional to the other element's uncertainty):

$$\hat{\underline{x}}_{k+1} = \frac{R_i \hat{\underline{x}}_k}{P_k + R_k} + \frac{P_k \underline{z}_k}{P_k + R_k}$$

• Putting this in recursive form:

$$\hat{\underline{x}}_{k+1} = \hat{\underline{x}}_k + \left[\frac{P_k}{P_k + R_k}\right](\underline{z}_k - \hat{\underline{x}}_k)$$

• This is equivalent to our earlier form where if the current estimate has one unit of uncertainty, $P_k = 1$, then the new measurement has proportionally more $R_k = k$. Also, the Kalman gain is:

$$K_k = \left[\frac{P_k}{P_k + R_k}\right]$$

• When the measurement uncertainty is large, the gain is small and the state is changed very little.

• The central limit theorem suggests that we can compute the new uncertainty from adding reciprocals:

$$\frac{1}{P_{k+1}} = \frac{1}{P_k} + \frac{1}{R_k}$$

• This can easily be rewritten as:

$$P_{k+1} = (1 - K_k)P_k$$

3.2.3 Matrix Form with Direct Measurements

• The matrix form of these equations is the "Smith and Cheeseman version" of the Kalman Filter:

$$K = P_k[P_k + R_k]^{-1}$$
$$\hat{\underline{x}}_{k+1} = \hat{\underline{x}}_k + K(\underline{z}_k - \hat{\underline{x}}_k)$$
$$P_{k+1} = (1 - K)P_k$$

• Smith and Cheeseman use a direct measurement of the state so that the interpretation is $\hat{\underline{x}}_k$ is the result of all merging so far and $\underline{z}_k$ is the latest "approximate transform" to be merged in.

3.2.4 Matrix Form with Indirect Measurements

• Let the measurements $\underline{z}$ be indirectly related to the state vector $\underline{x}$. Let the measurement vector be corrupted by a random noise vector $\underline{v}$ of known covariance R:

$$\underline{z} = H\underline{x} + \underline{v} \qquad R = Exp(\underline{v}\underline{v}^T)$$

• Identical derivation from the central limit theorem (or, from optimization as in Gelb) leads to the "Gelb form" of the Kalman Filter:

$$K_k = P_k H_k^T [H_k P_k H_k^T + R_k]^{-1}$$
$$\hat{\underline{x}}_{k+1} = \hat{\underline{x}}_k + K_k(\underline{z}_k - H_k \hat{\underline{x}}_k)$$
$$P_{k+1} = (1 - K_k H_k) P_k$$

• The form $\underline{z}_k - H_k \hat{\underline{x}}_k$ is the **measurement residual** - the difference between the current measurement and the measurement predicted based on the current state estimate.

3.2.5 Matrix Form with Indirect Nonlinear Measurements

• When the measurement relationship (or other things we have not covered) are nonlinear, the filter is called an Extended Kalman Filter (EKF).

• Let the measurements $\underline{z}$ be indirectly related to the state vector $\underline{x}$.by a nonlinear function and let the measurement vector be corrupted by a random noise vector $\underline{v}$ of known covariance R:

$$\underline{z} = h(\underline{x}) + \underline{v} \qquad R = Exp(v\underline{v}^T)$$

• We define the Jacobian matrix of the measurement relationship evaluated at the current state estimate:

$$H_k = \frac{\partial}{\partial \underline{x}}[h(\underline{x})]\Big|_{\underline{x} = \underline{x}_k}$$

• of the Kalman Filter:

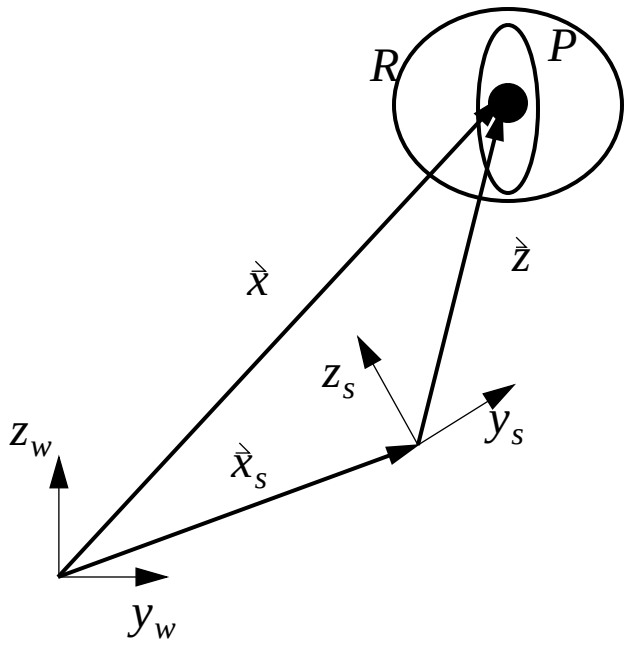
$$K_k = P_k H_k^T [H_k P_k H_k^T + R_k]^{-1}$$
$$\hat{\underline{x}}_{k+1} = \hat{\underline{x}}_k + K_k(\underline{z}_k - h(\hat{\underline{x}}_k))$$
$$P_{k+1} = (1 - K_k H_k) P_k$$

3.3 Example: Building Maps of Objects with Perfect Positioning

• Consider a situation where a mobile robot observes the positions of various objects as it roams around. As it moves, it sees the same objects over and over, so an opportunity exists to fix their position accurately through merging observations.

• While the sensor position in the world, $\underline{x}_s$, is known perfectly, the position of the object $\underline{z}$ relative to the sensor is computed from a noisy sensor. Based on techniques presented earlier and a model of sensor uncertainty, the covariance of the object position relative to the sensor is R.

• The position of the object relative to the world is called $\underline{x}$ and is given by:



$$\hat{x} = \hat{x}_s + \hat{z}$$

Express in world coordinates:

$$\underline{x} = \underline{x}_s + R_s^w \underline{z}$$

In standard form:

$$\underline{z} = R_w^s(\underline{x} - \underline{x}_s)$$

• This is of the form $\underline{z} = h(\underline{x})$. Given an initial estimate of the position and uncertainty of the object position P, a new measurement of uncertainty R can be merged with it.

3.4 Example: Building Maps of Objects with Imperfect Positioning

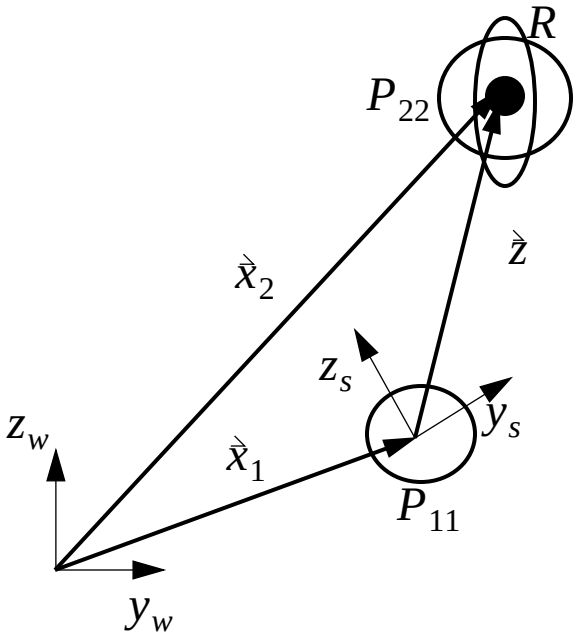
• It is possible to do exactly the same operations when both the sensor position and the object position are updated and considered noisy.

• Let the sensor position in the world be $\underline{x}_1$ and the object position in the world be $\underline{x}_2$. The composite state vector and its uncertainty is:

$$\underline{x} = \begin{bmatrix} \underline{x}_1 & \underline{x}_2 \end{bmatrix} \qquad P = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$

• Let the object position wrt the sensor be denoted $\underline{z}$ with covariance R.

• The position of the object relative to the world is given by:



$$\hat{x}_2 = \hat{x}_1 + \hat{z}$$

Express in world coordinates:

$$\underline{x}_2 = \underline{x}_1 + R_s^w \underline{z}$$

In standard form:

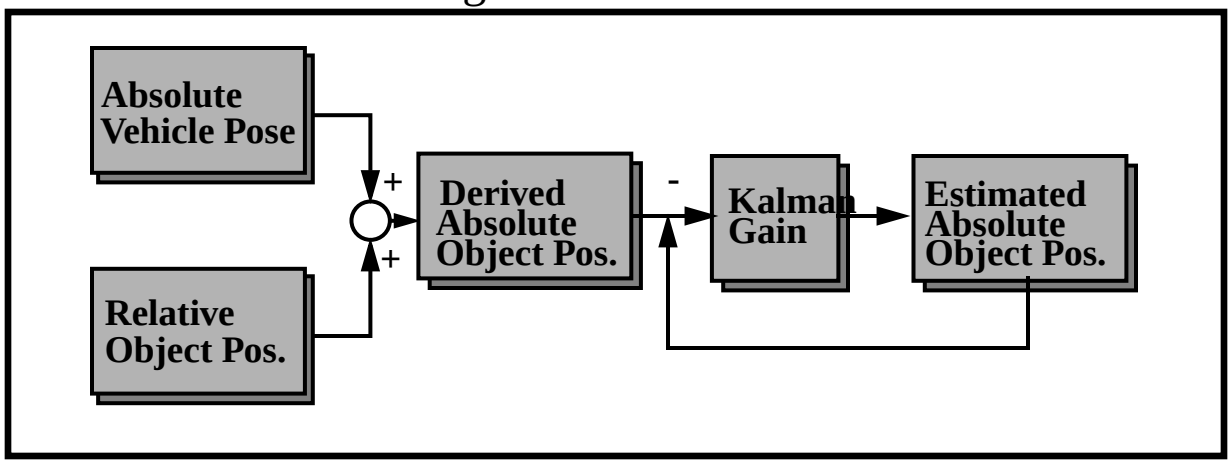
$$\underline{z} = R_w^s(\underline{x}_2 - \underline{x}_1)$$

• This is of the form $\underline{z} = h(\underline{x})$. Given an initial estimate of the position and uncertainty of the sensor position P_{11} and the object position P_{22}, a new measurement of uncertainty R can be merged with it.

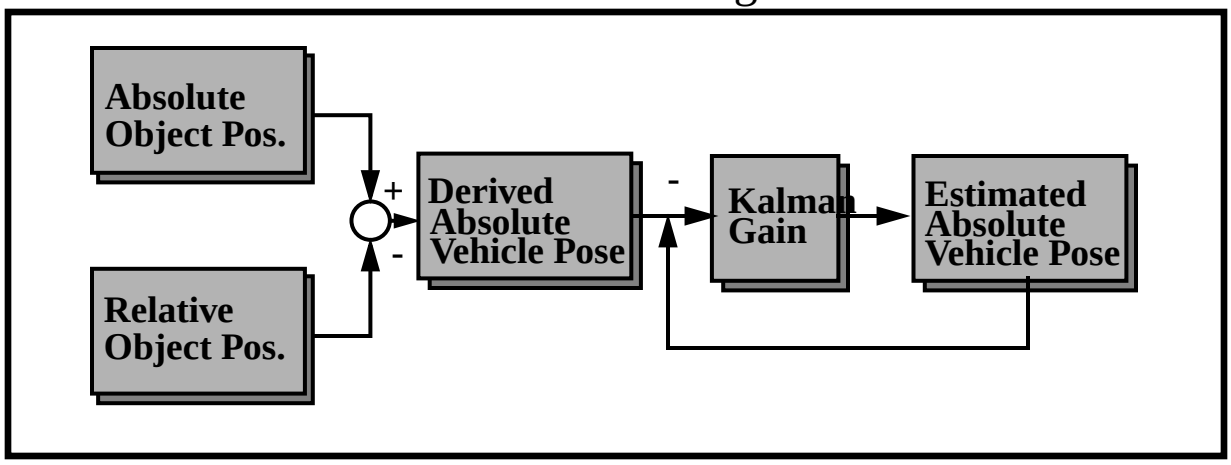
3.5 Duality of Mapping and Position Estimation

• A very general duality exits between position estimation and mapping in mobile robotics.

• In mapping, we take the vehicle pose and relative object positions as known and update the positions of objects based on it and sensor readings:



• In position estimation from landmarks, we take the absolute object positions as known and update the position of the vehicle based on it and sensor readings:



• And we have seen that its possible to do both at once.

4 Summary

• Combining noisy measurements leads to a noisier result. However, merging redundant measurements (filtering) leads to a less noisier result.

• In a compounding situation, a continuously changing quantity is computed from a sequence of measurements.

• In a merging situation, redundant measurements related even indirectly to the quantity of interest are combined.

• Kalman Filters can be used to estimate the positions of objects and the position of the robot simultaneously.