Dynamic Map Building and Localization: New Theoretical Foundations

Jeffrey K. Uhlmann

Robotics Research Group

Department of Engineering Science

University of Oxford

Trinity Term 1995

Dynamic Map Building and Localization: New Theoretical Foundations

Jeffrey K. Uhlmann
University of Oxford

Doctor of Philosophy
Trinity Term 1995

Abstract

This thesis examines the theoretical and computational problems associated with map building and localization for autonomous vehicles. In particular, components of a system are described for performing terrain-aided navigation in real time for high speed vehicles or aircraft. Such a system would be able to dynamically construct a map of distinctive naturally-occurring environmental features while simultaneously using those features as landmarks to estimate the position of the vehicle.

In order to develop such a system, a variety of challenges are addressed. Specifically:

1. A new approach for nonlinear filtering is described that is not only easier to implement, but substantially more accurate than the conventional methods.
2. A new approach is developed for avoiding problems associated with correlations among the position estimates of mapped features. Such correlations prevent the application of standard real time filtering methods and constitute the key challenge in the area of large scale map building. A byproduct of this development is a new general-purpose filtering and data fusion technique.
3. A new data structure is developed for storing the map so that sensor observations can be associated with candidate features in the map in real time. This data structure is shown to be capable of supporting real time performance for maps having many thousands of features.
4. A new combinatorial result is derived that facilitates the decision process for determining which mapped feature is most likely to have produced a given sensor observation.

Applications of the above results to other more general engineering problems are also discussed.

Extended Technical Abstract

This thesis considers technical issues associated with the estimation of the state of a physical system or systems based on a sequence of sensor measurements. Although autonomous map building is the primary application of interest, the technical contributions of this thesis are generally applicable to a wide variety of filtering, control, and data association problems. This extended abstract is intended for persons who are already familiar with these areas, and in particular for persons interested in Kalman filtering.

Definitions, Assumptions, and Other Details

We treat a system $\mathcal{X}$ as a random vector $\mathbf{x}(\cdot)$ with an unknown distribution function. We define an estimate of $\mathcal{X}$ to be a pair $(\mathbf{a}, \mathbf{A})$ in which the vector $\mathbf{a}$ is a purported mean vector of the distribution function associated with the random vector $\varphi[\mathbf{x}(\cdot)]$, where $\varphi[\cdot]$ is some known/modeled transformation of the true state of the system. The matrix $\mathbf{A}$ is defined to be a *conservative* estimate of the covariance of the distribution function of the transformed state about the mean vector $\mathbf{a}$. Specifically,

$$\mathbf{A} \geq \mathrm{E}\left[(\mathbf{a} - \varphi[\mathbf{x}(\cdot)])(\mathbf{a} - \varphi[\mathbf{x}(\cdot)])^T\right],$$

where a matrix inequality of the form $\mathbf{A} \geq \mathbf{X}$ for positive definite or semidefinite matrices $\mathbf{A}$ and $\mathbf{X}$ holds if and only if $\mathbf{A} - \mathbf{X}$ is positive definite or semidefinite. In other words, the estimated covariance matrix $\mathbf{A}$ is always an overestimate of the expected squared difference between the true mean of the unknown distribution function and the estimated mean $\mathbf{a}$.

The function $\varphi[\cdot]$ defines the subspace of interest out of the potentially infinite state space within which the system $\mathcal{X}$ can be described. For example, the state of a car could be defined by its make, model, color, year of manufacture, weight, materials, price, etc., but it is typically defined in engineering applications in terms of a small number of variables such as position and kinematics. Implicitly, therefore, $\varphi[\cdot]$ projects the full state of the vehicle down to the subspace defined by the relevant variables and choice of units, e.g., meters/second versus kilometers/hour.

Given some fixed choice of $\varphi[\cdot]$, the data fusion problem of interest is the following: Given two conservative estimates of the state of a system, $(\mathbf{a}, \mathbf{A})$ and $(\mathbf{b}, \mathbf{B})$, how can an improved fused estimate $(\mathbf{c}, \mathbf{C})$ be formed from the information provided by the two estimates? More specifically, the goal is to obtain a fused estimate in which $\mathbf{C} \leq \mathbf{A}$ and $\mathbf{C} \leq \mathbf{B}$, and preferably the inequalities are strict. If $\mathbf{A} \leq \mathbf{B}$, then letting the fused estimate be $(\mathbf{a}, \mathbf{A})$ achieves this goal, but it ignores any information provided by the other estimate. If the fused estimate is to improve the state of knowledge about the system, i.e., has covariance strictly less than either of the prior estimates, then information from both estimates must be exploited.

If the errors associated with the prior estimates can be assumed independent, then the Kalman filter update:

$$\begin{aligned} \mathbf{C} &= \left[\mathbf{A}^{-1} + \mathbf{B}^{-1}\right]^{-1} \\ \mathbf{c} &= \mathbf{C}\left[\mathbf{A}^{-1}\mathbf{a} + \mathbf{B}^{-1}\mathbf{b}\right] \end{aligned}$$

guarantees a conservative estimate such that $\mathbf{C} < \mathbf{A}$ and $\mathbf{C} < \mathbf{B}$ for any conservative estimates $(\mathbf{a}, \mathbf{A})$ and $(\mathbf{b}, \mathbf{B})$ with finite, nonzero covariances.

If the given estimates are (1) conservative, (2) in the same state space, and (3) have independent errors, then the Kalman filter provides an optimal mechanism for solving the data fusion problem. The applicability of the Kalman filter, however, depends on the ability to satisfy these conditions in practice. We will explain the implications of these conditions and describe additional mechanisms and generalizations which avoid their most important practical limitations. We will then consider an additional condition, (4), that the two given estimates are of the same system. This latter condition is an issue when the states of multiple systems are being estimated and it is not known a priori from which system any given measurement was taken.

Condition (1) - Estimates must be conservative

If the data fusion mechanism is guaranteed to yield conservative fused estimates from conservative prior estimates, then it is inductively critical to ensure that all estimates used in the process are conservative. In most cases, the only source of direct information about the system comes from sensor measurements.

Although most measuring devices/processes are based on well understood physical principles, it is not generally possible to analytically characterize all possible sources of measurement error. For example, a measuring device can be affected by random variables associated with the manufacturing process which produced it. Even if statistics about the manufacturing variables can be determined, that is not enough to allow conservative covariances to be determined for measurements taken from any particular device.

The typical way to quantify the measurement accuracy of a sensor is to take many sample measurements of systems whose true states are known. By comparing the known states to the measurements, it is possible to estimate an error covariance matrix. By taking more and more measurements, it is possible to obtain increasingly more accurate covariance estimates. It is also possible, with some weak additional assumptions, to estimate from the number of samples tested how much larger to scale the empirically determined covariance to ensure with high confidence that it is truly conservative.

In summary, it is possible to empirically determine the covariance of the error distribution associated with a measurement process to almost any level of precision. It is not possible to determine it exactly from any finite number of samples because there may be extremely infrequent occasions when the sensor yields measurements with enormous errors. However, the covariance of an unknown distribution is much more amenable to empirical determination than are other distribution statistics, e.g., maximum bounds, which often do not exist (as opposed to covariance, which exists for almost any distribution associated with a real-world measurement process).

Condition (2) - Estimates must be defined in the same state space

Most measurement processes provide information in a local coordinate frame different from the state space of interest. For example, a radar typically provides measurements in a local spherical coordinate frame centered at the position of the radar. Thus its estimates are defined in terms of the target's range and bearing from the radar. If the target's state is being maintained in a global Cartesian coordinate frame, then the measurement estimates cannot be used directly. An analogous situation arises with the time-indexed coordinate frame of a dynamical system in which an estimate of the state of the system at time t_k must be fused with information from a measurement taken at a subsequent time t_{k+1}.

If an estimate in one coordinate frame can be transformed to the coordinate frame of another estimate, then the two estimates can be fused in the common coordinate frame. For example, if a matrix $\mathbf{H}$ transforms an estimate $(\mathbf{a}, \mathbf{A})$ into the coordinate frame of an estimate $(\mathbf{b}, \mathbf{B})$, then a Kalman fused estimate can be generated directly in the coordinate frame of $(\mathbf{b}, \mathbf{B})$. It is also straightforward to show that a Kalman estimate in the coordinate frame of $(\mathbf{a}, \mathbf{A})$ can be generated as:

$$\begin{aligned} \mathbf{C} &= \left[\mathbf{A}^{-1} + \mathbf{H}^T\mathbf{B}^{-1}\mathbf{H}\right]^{-1} \\ \mathbf{c} &= \mathbf{C}\left[\mathbf{A}^{-1}\mathbf{a} + \mathbf{H}^T\mathbf{B}^{-1}\mathbf{b}\right]. \end{aligned}$$

However, this straightforward application of the Kalman filter depends on the fact that a linear transformation $\mathbf{H}$ applied to any distribution with mean and covariance $(\mathbf{a}, \mathbf{A})$ will produce a distribution with a mean and covariance $(\mathbf{Ha}, \mathbf{HAH}^T)$. In other words, the first two central moments of a distribution are linear statistics.

The problem that arises in most practical situations is that the transformations of interest are nonlinear. For example, the transformation from the spherical coordinate frame of a sensor to a global cartesian coordinate frame is highly nonlinear. The same is usually true for estimating the future state of real-world dynamical systems. Without knowledge of the exact underlying distribution associated with an estimate, it is not generally possible to determine a nonlinearly transformed mean and provably conservative covariance for an arbitrary nonlinear transformation. Because the underlying distribution is not generally known, approximate estimates must be made based only on mean and covariance information.

The traditional method for obtaining approximate nonlinearly transformed estimates is to generate an approximate transformation matrix $\mathbf{H}$ by linearizing the true nonlinear transformation $\mathbf{h}[\cdot]$. More specifically, the estimated transformation of an estimate $(\mathbf{a}, \mathbf{A})$ would be $(\mathbf{h}[\mathbf{a}], \mathbf{HAH}^T)$, where $\mathbf{H}$ is the Jacobian $\nabla\mathbf{h}[\cdot]$. The Kalman filter using linearized transformations is referred to as the extended Kalman filter (EKF). It can be shown empirically that linearization often produces very poor estimates with nonconservative covariances. The poor estimates are due in part to the fact that the transformation of the mean does not in any way exploit covariance information.

An intuitively (and empirically verifiable) better mechanism for applying nonlinear transformations is to transform a proxy distribution having the same mean and covariance as the estimate. Specifically, a discrete distribution of points/vectors $\{\mathbf{x}_1, ..., \mathbf{x}_n\}$ can be generated so as to have mean $\mathbf{a}$ and covariance $\mathbf{A}$, and then can be directly transformed as $\{\mathbf{h}[\mathbf{x}_1], ..., \mathbf{h}[\mathbf{x}_n]\}$. Assuming that the proxy distribution transforms similarly to the true distribution, the mean and covariance of the transformed set of points should be a good approximation to the mean and covariance of the true transformed mean and covariance.

It is easy to verify that the set of vectors $\mathbf{a} \pm \sqrt{n\mathbf{A}_i}$, generated from the n columns (or rows, depending on the chosen root) of the square root of the $n \times n$ covariance matrix $\mathbf{A}$, does in fact represent a distribution with the desired first and second moments. The fact that such a distribution consisting of $O(n)$ points exists is important because it allows the transformation to be computed with same order of calculations as is required for an ordinary linear transformation. In other words, the improved estimates of the method are obtained at little or no computational cost beyond that of linearization.

The improved accuracy of this new method, which is referred to as the Unscented Transformation, has been demonstrated in a variety of applications. Analysis has also been performed to explain why this improved accuracy is to be expected. However, more accurate estimates do not necessarily imply that the covariance estimates are conservative. When dealing with nonlinearly transformed estimates, it is necessary to enlarge the estimated covariance to account for errors associated with the transformation process. In general this is a tuning process that requires empirical analysis analogous to that required to determine the noise characteristics of a measuring device.

Condition (3) - Estimate errors must be independent

The independence assumptions associated with the Kalman filter are usually considered to be relatively innocuous for practical applications because of the following misconceptions:

- It is widely believed that in most applications the error associated with each new sensor measurement is independent of the error in the current system estimate, which is derived from previous measurements. The problem is that any nonlinear transformation will introduce time-correlated errors. For example, if each update of the system estimate involves the nonlinear transformation of a measurement from the coordinate frame of the sensor, the errors associated with that transformation will be time-correlated. Specifically, while the errors in each raw measurement may be independent of those of previous measurements, the errors in the transformed measurements are *not* independent. Because the system estimate is derived from previously transformed measurements, its errors are also not independent of those of the current transformed measurement. There are many other even more subtle avenues by which correlated errors may enter the estimation process.

- If estimates are not independent, it is widely believed that augmented Kalman filter equations can be applied to incorporate cross covariance information to yield conservative estimates. The problem is that while such augmented equations have been derived, they require *exact* knowledge of estimate cross covariances. Unlike covariances which can be conservatively over-estimated, any deviation from the use of true cross covariance information in the augmented Kalman filter equations can lead to erroneous (nonconservative) estimates. In other words, not only is it generally impossible to determine the degree of cross covariance between two estimates, the augmented Kalman equations do not admit approximations to be used. Texts on the Kalman filter typically assume independence and casually suggest that all other cases can be addressed with the augmented equations. This is a significant misrepresentation of the situation.

It is possible to re-examine much of the applied literature on the Kalman filter and explain observed poor results (though often purported to be good results by the authors) in terms of violations of the strict Kalman independence assumptions. In fact, the performance of the Kalman filter is often observed to be relatively unstable, and very sensitive to the effects of tuning, despite theoretical analysis suggesting that it should exhibit strong stability. The problem is that a subtle violation of independence assumptions can severely undermine the integrity of a Kalman filter. Because there is no general way to avoid such violations, an alternative to the Kalman filter is required to fuse estimates with unknown degrees of correlation.

It turns out that it is possible to derive an alternative to the Kalman filter that avoids independence assumptions. In fact, the equations are very similar:

$$\begin{aligned}
\mathbf{C} &= \left[\omega \mathbf{A}^{-1} + (1-\omega)\mathbf{B}^{-1}\right]^{-1} \\
\mathbf{c} &= \mathbf{C}\left[\omega \mathbf{A}^{-1}\mathbf{a} + (1-\omega)\mathbf{B}^{-1}\mathbf{b}\right],
\end{aligned}$$

where $0 \le \omega \le 1$ is a scalar parameter selected to minimize any chosen measure of the size of the covariance matrix $\mathbf{C}$. This update mechanism (and its various algebraic formulations and generalizations) is referred to as Covariance Intersection. It can be shown that the resulting covariance $\mathbf{C}$ is guaranteed to be conservative with respect to the estimated mean $\mathbf{c}$ for any choice of ω regardless of the degree of correlation between the prior estimates. This is trivially verifiable for the limiting values $\omega = 0$ or $\omega = 1$ because the result is one of the prior estimates.

In order to ensure that the system estimate never degrades after a measurement update, i.e., to ensure nondivergence, it is necessary at each update to select a value for ω that minimizes a fixed measure of covariance size. In other words, it is necessary to ensure that the updated system covariance estimate is always less than or equal to, according to the chosen measure, the prior system covariance. For a variety or reasons, it is usually best to choose ω so as to minimize the determinant of $\mathbf{C}$. Fortuitously, the optimum value of ω for minimizing the determinant can be computed very efficiently.

Covariance Intersection provides a mechanism for solving a much larger class of estimation/filtering problems than the Kalman filter. General nonlinear filtering can be performed rigorously using Covariance Intersection as long as conservative covariances can be generated for nonlinearly transformed estimates. The Kalman filter, as has been discussed, cannot be applied rigorously because correlated errors of unknown magnitude are introduced and propagated by nonlinear transformations even if the estimated covariances are conservative.

A problem that illustrates the power of Covariance Intersection is simultaneous map building and localization. In this problem a vehicle with an onboard sensor, whose initial position is known in a global coordinate frame, must dynamically map the positions of observed features in its environment while simultaneously using the estimated positions of re-observed features to update its own position estimate. The difficulty is that the position estimates of mapped features inherit errors due to the positional uncertainty of the vehicle. Consequently, the vehicle and mapped feature estimates are all correlated to an unknown degree unless cross covariances are maintained between every feature estimate and the vehicle estimate, and between every feature estimate and every other feature estimate.

Maintaining cross covariance information, which grows quadratically in the number of mapped features, is impractical for most real-time applications. However, the use of Covariance Intersection avoids the need to maintain cross covariances because it can fuse correlated estimates directly. As an example, if a stationary vehicle observes the position of a mapped feature, successive observations will not improve the estimates of the positions of the vehicle or feature because no new information is being obtained. The application of a Kalman filter to fuse these correlated estimates, however, will lead to spurious improvements as the vehicle and feature covariance estimates rapidly go to zero. The problem is that the Kalman filter assumes that each new measurement is an independent piece of information that can be exploited to yield an improved updated estimate.

Condition (4) - Estimates must be of the same system

The simplest class of filtering problems involves the successive update of a system estimate with a sequence of measurements of that system. A more general class of problems involves the simultaneous filtering of multiple system estimates when it is not known from which system each measurement originated. Simultaneous map building and localization is an example of such a problem because it is not known to which feature each observation corresponds. Thus, in addition to the update/fusion step, a preliminary step must be performed to determine which estimate to update. This preliminary step is referred to as data association.

There are two aspects of the data association problem: One is algorithmic, the other statistical. The algorithmic aspect of the problem is often referred to as gating. Gating is performed in order to quickly identify the subset of systems from which a particular measurement might have originated. This is typically done by determining a region in the system state space that includes, with some sufficiently high probability, all feasible positions of the measured state. This allows all systems whose state estimates are outside of this region to be excluded as possible sources of the measurement.

It turns out that if convex bounded regions can be associated with the n system estimates, then these regions can be represented in a data structure so that the regions intersecting the region associated with a given measurement can be identified in $O(n^{1-1/d} + m)$ time, where d is the dimensionality of the state space and m is the number of identified estimate regions. If the bounded regions are defined as coordinate-aligned boxes, then the same worst case retrieval complexity can be achieved, but the average case efficiency can be substantially improved. The use of boxes also allows the data structure to be efficiently updated if system states change. This is necessary for the real-time estimation of dynamical systems.

The use of bounded regions to reflect uncertainty estimated in terms of covariance introduces the possibility that the true source of a highly deviant measurement may be erroneously excluded. There is no rigorous way to avoid this problem because lack of knowledge about the complete distribution associated with each mean/covariance estimate makes it impossible to determine guaranteed bounds on the error associated with each state variable. Thus, a tradeoff must be made between computational efficiency and gating accuracy, i.e., to maximize accuracy within given computational constraints.

After gating, the other component of the data association problem is the determination of which candidate system estimate (or estimates) should be updated with the current measurement. This is referred to as the assignment problem. If there are n measurements of n systems, then it is possible to identify all feasible one-to-one mappings that are consistent with the coarse associations determined by the gating step. For example, if measurement M_1 gated only with the two system estimates S_1

and S_2, and measurement M_2 only gated with S_1, then the one-to-one mapping constraint implies that M_1 can only be assigned to S_2.

The difficulty associated with the assignment problem is that there may be multiple different feasible assignments. One way to resolve this ambiguity is to associate an assignment likelihood with each pair of system and measurement estimates by counting the number of feasible assignments containing the pair. It turns out that this counting of assignments involves the calculation of a combinatorial quantity called the permanent. Unfortunately, there is no known efficient algorithm for computing the permanent, nor is there even an efficient algorithm for computing accurate approximations to the permanent. However, it has been shown empirically that highly crude, but efficiently computable approximations to the permanent are sufficient to yield good approximations to the assignment problem.

Summary

The Unscented Transformation, Covariance Intersection filter, Priority kd-Tree for gating, and the permanent approximations for data association are all novel contributions of this thesis. They address problems associated with simultaneous map building and localization as well as many other general filtering and data fusion applications.

Contents