The 2016 AAAI Fall Symposium Series:
Cross-Disciplinary Challenges for Autonomous Systems
Technical Report FS-16-03

Reduction of the State Observation Problem to an Identifiability Problem

Masataka Nishi
Center for Technology Innovation, Hitachi Research Laboratory, Hitachi Ltd, Japan
masataka.nishi.en@hitachi.com

Abstract

Data integrity is a property which a world state interpreted with a world model is consistent with the real operating environment. Even a formally verified safety claim of an autonomous system is prone to a malfunction caused by loss of data integrity. From a first-person viewpoint in a congested environment, some components of measurable part of the world state may become transiently deficient or unavailable because of the limited capability of sensor devices. If the system could get into a situation where the world state becomes suddenly unobservable, existing estimation methods may get unstable. These methods can hardly detect the loss of data integrity and produce an incorrect estimate without any notice. Our insight is that we can merge the original concept of observer theory with that of automated reasoning. Firstly, we propose a new way of unifying them into a problem of checking satisfiability of a formula that consists of predicates regarding the world model and decision variables regarding unmeasurable part of the world state. We can detect a loss of data integrity by checking if the problem is unsatisfiable. Secondly, we replace the idea of observability in control theory with identifiability with respect to a measure of tolerance and a world model. We show a procedure of estimating the world state with a bounded uncertainty specified by the measure of tolerance. Third, we show that a problem of sensor fusion, a problem of reasoning a world state of discrete and enumerated type, and a decision problem under uncertainty in the world state are formulated as an identifiability problem. The proposal presents a constructive basis for supporting the degree of confidence in the estimated world state.

Introduction

A world state consists of observable and unobservable part. Observable part of the world state consists of measurable and directly unmeasurable part. A world model consists of constraints on the world state. Data integrity is a property which the world state interpreted using the world model is consistent with the real operating environment. Without a guarantee of data integrity, a safety claim of an autonomous system that operates in an open and dynamic environment may become invalid. The system may make a false decision on the risk of hazard and may trouble determining a safe motion in response to the decision. We judge that presuming data integrity is unrealistic because of five basic issues.

First, measurable part of the world state may be corrupt or transiently unavailable from the system's first-person viewpoint because of physical limits of sensor devices. The reliability of a machine vision system using a stereo camera would be impaired by poor visibility and a limited line-of-sight. GPS signal would be blocked in a shielded area.

Second, some components of the world state are directly unmeasurable because it lacks a valid physical mechanism. We overcome this problem by developing a state observation system that uses a world model and produces an estimate of the world state. Such unmeasurable part of the world state could be produced by, e.g., the machine vision system that uses an artificially designed estimator called a classifier. The correctness of the classifier is prone to bias in the sample data set and is affected by the quality of the data supplied to the classifier. It also depends on a way of handling the impact of voids in the sample data set on the accuracy of the estimate (Szegedy 2013). We hence judge that a comprehensive framework for robustly improving the quality of the classifier has not been established yet.

Third, assuming that all components of the world state are always observable is unrealistic. This is an infrequently discussed concern that it may result from that part of the world state become transiently unobservable or that the world model is flawed or in a transitory change. In control theory, it means that we cannot find any impact of a change in unmeasurable part of the world state on a change in measurable part. In automated reasoning theory, it means that we cannot determine a unique estimate of the world state. In principle, we cannot reconstruct unobservable part of the world state with a bound on its precision, regardless of high precision of measurable part of the world states. Because the problem of inferring the unobservable part of the world state is under-constrained. Unfortunately, the autonomous system needs to maintain operation in such a situation for validating a statistical requirement on the accident frequency. We are not sure whether it is possible to define the safety claim by using only observable part of the world state. Thus, one possible direction is to find a good world model that consists of reliable, empirically and statistically supported constraints on the world state and the components of the world state that constitute the safety claim keep observable even if either one

Copyright © 2016, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

of measurable part of the world state could be unavailable.

Fourth, it is unlikely that the quality of unmeasurable part of the world state could be improved by using redundant measurement devices and an information fusion method. Aside from the inherent inaccuracy of the measurement devices, the estimation method needs to reconstruct the world state by using a single imperfect world model for interpretation of the raw data from the redundant devices. Yet, the statistical trend of the measurement error can hardly be stationary and the world model used for removing the error may be inaccurate or change over time or from one location to another. Thus, redundancy does not help avoid deviations in the estimate of the world state from the actual state and there are no means to detect such deviations.

Fifth, few studies exist on decision and control algorithms for the autonomous systems that can tolerate a partially deficient world state. Even a formally verified safety claim of a decision and control logic of an autonomous system is prone to the mismatch between the deficient world state and the real environment. This is a single point of failure.

The real environment is inherently uncertain at the design stage. As we can hardly predict which issue could result in loss of data integrity, this poses a challenge to certification (Rushby 2008). The autonomous system should be capable of detecting the loss of data integrity at runtime, and reasoning which components of the world state are unreliable. They are essential for adaptively changing the world model without demanding an unrealistic degree of precision.

The idea of inferring unknown part of the world state using the world model together with known part of the world state is called a state observer in control theory and automated reasoning in logic theory. Our insight is that both techniques can be unified in a problem of checking satisfiability of a formula that consists of the world state and the world model that encodes imperfect knowledge about the environment. Given measurable part of the world state, we check if there exists a satisfiable assignment of the unmeasurable part of the world states. Loss of data integrity is detected if the problem is unsatisfiable. We can examine if the world model is inconsistent with the real environment or the measurable part of the world state is deficient. Such a formulation removes the need for a tedious convergence analysis of the state observer and enables a feature of adaptively updating the world model without any re-design of the state observer. The formulation enables a Boolean structure to be embedded in the original observer theory and also opens a new way of encoding imperfect knowledge about the world.

As an alternative to the notion of observability, we propose a notion of identifiability by using a measure of tolerance and judge the degree of confidence in the estimate. If the unmeasurable part of the world state is identifiable even when part of the world state is deficient or unavailable, such a capability is a solution to the problem of making a decision under limited observability. We can monitor the validity of the safety claim subject to limited observability of the world states by adding a predicate of the safety claim and by checking if Boolean value of the predicate is identifiable.

Formulation of State Observation Problem

Conventional State Observer

In control theory, we split the world model into a process model (1) and an observation model (2).

$$\boldsymbol{x}_{t+1} = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_{t+1}) + \bar{\boldsymbol{e}}_t \tag{1}$$

$$\boldsymbol{y}_t = \boldsymbol{g}(\boldsymbol{x}_t) + \bar{\boldsymbol{w}}_t \tag{2}$$

The world state at the time t is $\{\boldsymbol{x}_t, \boldsymbol{u}_t, \boldsymbol{y}_t\}$ where $\boldsymbol{x}_t \in \mathbb{R}^n$ is a state vector, an input vector $\boldsymbol{u}_t \in \mathbb{R}^m$ and an output vector $\boldsymbol{y}_t \in \mathbb{R}^r$. $\boldsymbol{u}_t$ and $\boldsymbol{y}_t$ belong to measurable part of the world state, while $\boldsymbol{x}_t$ is unmeasurable part of the world state. We denote statistical modeling errors in (1) and (2) as $\bar{\boldsymbol{e}}_t \in \mathbb{R}^n$ and $\bar{\boldsymbol{w}}_t \in \mathbb{R}^r$, respectively. We assume that they are unmeasurable, additive, zero mean, uncorrelated Gaussian processes with an error covariance matrix $\boldsymbol{Q}_t \in \mathbb{R}^{n\times n}$ and $\boldsymbol{R}_t \in \mathbb{R}^{r\times r}$, respectively. We denote a temporal series of the world states as $\boldsymbol{X}_T := \{\boldsymbol{x}_t | 0 \le t \le T\}$, $\boldsymbol{U}_T := \{\boldsymbol{u}_t | 0 \le t \le T\}$, and $\boldsymbol{Y}_T := \{\boldsymbol{y}_t | 0 \le t \le T\}$. We assume that $\boldsymbol{f} : \mathbb{R}^{n+m} \to \mathbb{R}^n$ and $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^r$ are Lipschitz continuous and thus $\{\boldsymbol{X}_T, \boldsymbol{U}_T, \boldsymbol{Y}_T\}$ is uniquely determined from the initial state $\{\boldsymbol{X}_0, \boldsymbol{U}_0, \boldsymbol{Y}_0\}$. A state observer produces an estimate $\hat{\boldsymbol{x}}_T \in \mathbb{R}^n$ of the state vector $\boldsymbol{x}_T$ by using $\{\hat{\boldsymbol{X}}_{T-1}, \boldsymbol{U}_{T-1}, \boldsymbol{Y}_{T-1}\}$ where $\hat{\boldsymbol{X}}_T := \{\hat{\boldsymbol{x}}_t | 0 \le t \le T\}$ is an estimate of $\boldsymbol{X}_T$. A conventional state observer consists of an update rule (3), an observer gain $\boldsymbol{h} : \mathbb{R}^r \to \mathbb{R}^n$, and a predictor $\hat{\boldsymbol{y}}_t \in \mathbb{R}^r$.

$$\hat{\boldsymbol{x}}_{t+1} = \boldsymbol{f}(\hat{\boldsymbol{x}}_t, \boldsymbol{u}_{t+1}) + \boldsymbol{h}(\boldsymbol{y}_{t+1} - \hat{\boldsymbol{y}}_{t+1}) \wedge \hat{\boldsymbol{y}}_t = \boldsymbol{g}(\hat{\boldsymbol{x}}_t) \tag{3}$$

We can test the quality of the prediction $\hat{\boldsymbol{y}}_t$ by computing an estimation error $\boldsymbol{e}_t \equiv \boldsymbol{x}_t - \hat{\boldsymbol{x}}_t$ and $\boldsymbol{w}^y_t := \boldsymbol{y}_t - \hat{\boldsymbol{y}}_t$. Given $\{\boldsymbol{U}_T, \boldsymbol{Y}_T\}$, we design the observer gain $\boldsymbol{h}(:)$ such that a system of equations (3) (4) is stable and $\underset{t\to\infty}{lim} |\boldsymbol{e}_t| = 0$.

$$\boldsymbol{e}_{t+1} = \boldsymbol{f}(\boldsymbol{e}_t + \hat{\boldsymbol{x}}_t, \boldsymbol{u}_{t+1}) - \boldsymbol{f}(\hat{\boldsymbol{x}}_t, \boldsymbol{u}_{t+1}) - \boldsymbol{h}(\boldsymbol{w}^y_{t+1}) \tag{4}$$

If the world model (1) (2) is a linear time invariant (LTI) system, where $\nabla_{\boldsymbol{x}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) := \boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\nabla_{\boldsymbol{u}} \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) := \boldsymbol{B} \in \mathbb{R}^{n\times m}$, and $\nabla_{\boldsymbol{x}} \boldsymbol{g}(\boldsymbol{x}) := \boldsymbol{C} \in \mathbb{R}^{r\times n}$, then we must assume that $[\boldsymbol{A}, \boldsymbol{C}]$ is observable. The combined system (4) gets convergent by selecting a stable eigenstructure. The condition of observability is required to guarantee that there is an observer gain $\boldsymbol{h}(:)$ such that $\underset{t\to\infty}{lim} |\boldsymbol{e}_t| = \boldsymbol{0}$ and a stable estimate $\hat{\boldsymbol{X}}_T$ is determined. The Extended Kalman filter (EKF) uses a state observer (5) and a rule (6) for updating two matrix $\boldsymbol{P}_{t|t}, \boldsymbol{P}_{t+1|t} \in \mathbb{R}^{n\times n}$.

$$\boldsymbol{h}(\boldsymbol{w}^y_{t+1}) := \boldsymbol{W}_{t+1} \boldsymbol{w}^y_{t+1} \tag{5}$$

$$\begin{gathered}
\boldsymbol{P}_{t+1|t} \equiv \nabla \boldsymbol{f}(\hat{\boldsymbol{x}}_t, \boldsymbol{u}_{t+1}) \boldsymbol{P}_{t|t} \nabla \boldsymbol{f}(\hat{\boldsymbol{x}}_t, \boldsymbol{u}_{t+1})^T + \boldsymbol{Q}_{t+1} \\
\boldsymbol{S}_{t+1} \equiv \nabla \boldsymbol{g}(\hat{\boldsymbol{x}}_{t+1}) \boldsymbol{P}_{t+1|t} \nabla \boldsymbol{g}(\hat{\boldsymbol{x}}_{t+1})^T + \boldsymbol{R}_t \\
\boldsymbol{W}_{t+1} \equiv \boldsymbol{P}_{t+1|t} \nabla \boldsymbol{g}(\hat{\boldsymbol{x}}_{t+1})^T \boldsymbol{S}^{-1}_{t+1} \\
\boldsymbol{P}_{t+1|t+1} \equiv \boldsymbol{P}_{t+1|t} - \boldsymbol{W}_{t+1} \boldsymbol{S}_{t+1} \boldsymbol{W}^T_{t+1}
\end{gathered} \tag{6}$$

The state observer using (5) gives the estimate $\hat{\boldsymbol{X}}_T$ that becomes an optimal solution to a quadratic error (7).

$$\min_{\hat{\boldsymbol{X}}_T} \sum_{t=0}^{T-1} \boldsymbol{e}_t^T \boldsymbol{Q}_t \boldsymbol{e}_t + \boldsymbol{w}_t^{yT} \boldsymbol{R}_t \boldsymbol{w}_t^y \tag{7}$$

When $\boldsymbol{y}_t$ quickly changes but convergence rate of the state observer is slow, a large transient deviation of the estimation error $\boldsymbol{w}_t^y$, $\boldsymbol{e}_t$ from zero is inevitable by design. If either (1) or (2) is non-linear and the EKF cannot ensure monotonic convergence, the large deviation becomes a practical issue. The convergence analysis of non-linear systems when either one of $\boldsymbol{f}, \boldsymbol{g}$ and $\boldsymbol{h}$ in (4) is non-linear and synthesizing the observer gain have been topics of active research for decades. Backstepping method (Smyshlyaev 2005) presents a constructive way of building a stable Lyapunov function, yet only when (1) satisfies an assumption of passivity.

State Observation by Solving Satisfiability Problem

An analytical convergence analysis of (4) is difficult in general. Instead, we will check numerically if the system of simultaneous equations (8) is satisfiable, given $\{\boldsymbol{Y}_T, \boldsymbol{U}_T\}$. The problem is numerically solved using a dedicated reformulation techniques (Nishi 2016) and the nonlinear programming (NLP) solvers IPOPT (Wachter 2006) and ANTIGONE (Misener 2014), which guarantee global convergence from an arbitrary initial search point.

$$\begin{gathered} \bigwedge_{t=0}^{T-1} \begin{aligned} \hat{\boldsymbol{x}}_{t+1} - \boldsymbol{f}(\hat{\boldsymbol{x}}_t, \boldsymbol{u}_{t+1}) + \boldsymbol{v}_t^x = \boldsymbol{0} \wedge \\ \boldsymbol{y}_t - \boldsymbol{g}(\hat{\boldsymbol{x}}_t) + \boldsymbol{v}_t^y = \boldsymbol{0} \wedge \end{aligned} \\ \textstyle\sum_{t=0}^{T-1} \left[\boldsymbol{v}_t^{xT} \boldsymbol{Q}_t \boldsymbol{v}_t^x + \boldsymbol{v}_t^{yT} \boldsymbol{R}_t \boldsymbol{v}_t^y \right] = s \wedge \\ s \le \epsilon_0 \wedge \textstyle\sum_{t=0}^{T-1} \boldsymbol{v}_t^x = \boldsymbol{0} \wedge \sum_{t=0}^{T-1} \boldsymbol{v}_t^y = \boldsymbol{0} \end{gathered} \tag{8}$$

Unmeasurable residual vectors $\boldsymbol{v}_t^x \in \mathbb{R}^n, \boldsymbol{v}_t^y \in \mathbb{R}^r$ are comparable with $\boldsymbol{e}_t$ and $\boldsymbol{w}_t^y$, respectively. Contrary to $\boldsymbol{e}_t$ which contains the both of an intrinsic modeling error $\bar{\boldsymbol{e}}_t$ and a transient overshoot $\boldsymbol{h}(\boldsymbol{w}_{t+1}^y)$ in (4) artificially caused by design, we can remove the latter one from $\boldsymbol{v}_t^{x,y}$. We reuse the same error covariance matrix $\boldsymbol{Q}_t, \boldsymbol{R}_t$. This is one of the assumptions placed on the world model. We compute $\hat{\boldsymbol{X}}_T$ and $\{\boldsymbol{v}_t^{x,y} | 0 \le t < T\}$ such that the quadratic residual error s in (8) is not greater than a given error tolerance $\epsilon_0 > 0$. We can optimize s by iteratively lowering ϵ_0, in return for longer computation time. As long as the observability condition holds, We can reconstruct a stable $\hat{\boldsymbol{X}}_T$ in a conventional sense that $\boldsymbol{X}_T$ resides near $\hat{\boldsymbol{X}}_T$. Indeed, if the world model (1) (2) is an LTI system, the condition of observability corresponds to that the matrix on LHS of (9) is invertible.

$$\begin{bmatrix} -\boldsymbol{A} & \boldsymbol{I} & \boldsymbol{0} & . \\ \boldsymbol{C} & \boldsymbol{0} & \boldsymbol{0} & . \\ \boldsymbol{0} & -\boldsymbol{A} & \boldsymbol{I} & . \\ \boldsymbol{0} & \boldsymbol{C} & \boldsymbol{0} & . \\ . & . & . & . \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{x}}_0 \\ \hat{\boldsymbol{x}}_1 \\ \cdots \\ \hat{\boldsymbol{x}}_T \end{bmatrix} = \begin{bmatrix} -\boldsymbol{v}_0^x + \boldsymbol{B}\boldsymbol{u}_1 \\ \boldsymbol{v}_0^y + \boldsymbol{y}_0 \\ -\boldsymbol{v}_1^x + \boldsymbol{B}\boldsymbol{u}_2 \\ \cdots \\ \boldsymbol{v}_{T-1}^y + \boldsymbol{y}_{T-1} \end{bmatrix} \tag{9}$$

Here, we point out that the condition of observability can be transiently violated, only if part of $\boldsymbol{Y}_T$ gets deficient or unavailable. If $y_{j,k}$ is unavailable and a correspondent component $g_j(\boldsymbol{x}_k)$ becomes invalid, the conventional observer may become unstable and produce an unreliable $\hat{\boldsymbol{X}}_T$. A benefit of formula (8) is that it enables us to detect situations where the world model suddenly changes in a way that the condition of observability is violated and the observer becomes unstable unless a detectability condition regarding the world model holds. It results from that the formula (8) in the situations is under-constrained, and there are several satisfiable solutions that cannot be bounded with a small tolerance. Since formula (8) without $s < \epsilon_0$ is guaranteed to be satisfiable, satisfiability of (8) depends on that of the quadratic residual error s. The residual error suggests a degree of confidence in the data integrity. If the formula (8) is unsatisfiable, then we can detect the loss of data integrity and report that we cannot reconstruct a stable integrity-preserved estimate $\hat{\boldsymbol{X}}_T$. On the other hand, the conventional observer may provide a bad estimate $\hat{\boldsymbol{X}}_T$ without any noticeable warning on the loss of data integrity, even when the quadratic error (7) is optimal but large.

Another benefit of the formula (8) is that it allows us to skip the convergence analysis of (4). Instead, the NLP solver internally reproduces the recursive steps (5)(6) of computing a series of search iterates that converge to an optimal solution of (7). The transient deviation $\boldsymbol{h}(\boldsymbol{w}_{t+1}^y)$ due to the slow convergence rate of (4) is diminished within the NLP solver and is not added to the estimation error of $\hat{\boldsymbol{X}}_T$.

Yet, we can infer at most that the loss of data integrity results from that the formula (8) is over-constrained. It can result from either that a transient measurement error in $\boldsymbol{Y}_T$, that the assumption on the modeling uncertainty $\bar{\boldsymbol{e}}_t, \bar{\boldsymbol{w}}_t$ in the presumed world model (1) and (2) from the actual environment is violated, or that the assumption that the world state is observable is violated. We need more knowledge to identify which factor resulted in the loss of data integrity. While the first two factors are inseparable, the third one is not. We will define a notion of identifiability of the world state with respect to a given measure of tolerance.

Checking Identifiability

The idea of reformulating a state observation problem as a satisfiability problem of (8) leads to a general way toward describing the world model with imperfect knowledge. As mentioned in regard to (9), the condition of observability is necessary for reconstructing a stable integrity-preserved estimate. Yet, we can hardly guarantee that the world state is always fully observable. We need to check the world model and to check if the condition of observability still holds, even if part of the world state can be transiently unavailable and we can hardly presume perfect knowledge about the world model. Instead of hoping for a guarantee of observability, we propose checking identifiability subject to a measure of tolerance. Let $\{\boldsymbol{X}, \boldsymbol{Y}\}$ be directly unmeasurable part and measurable part of the world state, respectively. $\boldsymbol{Y}$ can contain a component of an action. Let $\boldsymbol{M}(\boldsymbol{X}, \boldsymbol{Y})$ be a formula of the world model that consists of a collection of constraints on the world state. The pair of the process model (1) and the observation model (2) corresponds to this world model. Let

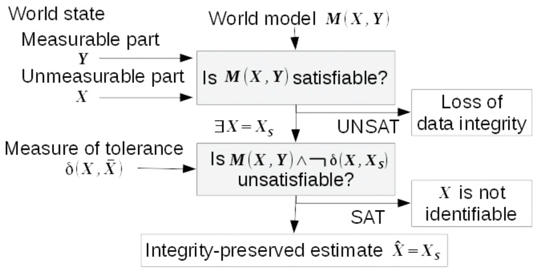


Figure 1: Computing integrity-preserved estimate.

Table 1: Classification of error modes.

Correct world model	Observable	Correct measurement	loss of data integrity(L), or not identifiable(NI)?
YES	YES	YES	-
YES	YES	NO	L,NI
YES	NO	YES	NI
YES	NO	NO	L, NI
NO	YES	YES	L
NO	YES	NO	L,NI, or false negative
NO	NO	YES	L,NI
NO	NO	NO	L,NI, or false negative

the predicate $\delta(\boldsymbol{X}, \bar{\boldsymbol{X}})$ be a measure of tolerance on a deviation between $\boldsymbol{X}$ and a reference state $\bar{\boldsymbol{X}}$.

Definition: $\boldsymbol{X}$ is identifiable with respect to the measure of tolerance $\delta(\boldsymbol{X}, \bar{\boldsymbol{X}})$ and given $\boldsymbol{Y}$, if and only if there exists a satisfiable solution $\boldsymbol{X} = \boldsymbol{X}_S$ to $\boldsymbol{M}(\boldsymbol{X}, \boldsymbol{Y})$ and also a formula $\boldsymbol{M}(\boldsymbol{X}, \boldsymbol{Y}) \wedge \neg\delta(\boldsymbol{X}, \boldsymbol{X}_S)$ is unsatisfiable.

Figure 1 shows the procedure of computing an integrity-preserved estimate $\hat{\boldsymbol{X}}$, that of detecting a loss of data integrity, or that of detecting a violation of the assumption that the world state is observable. The constraint on the quadratic error regarding $\{\boldsymbol{v}_t^x, \boldsymbol{v}_t^y\}$ in (8) corresponds to the measure of tolerance. Even part of $\boldsymbol{Y}$ is transiently unavailable, we can reconstruct part of $\boldsymbol{X}$ unless the world model subject to $\boldsymbol{Y}$ is under-constrained. It is a notable benefit of removing the convergence analysis of the state observer. In case the world state is not identifiable and thus unreliable, then we can either relax the measure of tolerance and reiterate the procedure, or instruct the subsequent decision system to set a quantified predicate which encodes uncertainty in the unidentifiable world state. Table 1 shows how the procedure can correctly detect a loss of data integrity.

There are four cases of interest in the tables. The first case is on rows 2, 4, 5 and 7, wherein either one of the world model or measurable part of the world state $\boldsymbol{Y}$ is incorrect. If we cannot compute a satisfiable solution to $\boldsymbol{M}(\boldsymbol{X}, \boldsymbol{Y})$, we detect a loss of data integrity. The second case is on rows 3 and 4, wherein the world model is correct but part of the world state $\boldsymbol{X}$ is not observable. Here, although we can find a satisfiable solution $\boldsymbol{X} = \boldsymbol{X}_S$ to $\boldsymbol{M}(\boldsymbol{X}, \boldsymbol{Y})$, there may be another one $\boldsymbol{X} = \widetilde{\boldsymbol{X}}_S$ such that $\delta(\widetilde{\boldsymbol{X}}_S, \boldsymbol{X}_S)$ is false; thus we conclude that the estimate of $\boldsymbol{X}$ is not identifiable. The third case is on row 2 and 7, wherein the world model is imperfect or the measurement $\boldsymbol{Y}$ is noisy but we try to raise the degree of confidence on the estimate by setting a too small ϵ_0. Yet the estimate is not identifiable. The fourth case is on rows 6 and 8 wherein both of the world model and measurable part of the world state are incorrect. In this case, a risk of false negative arises, because there is an evident counter-example in which data integrity of the world model and the world state is preserved but they are greatly different from the actual environment. Thus, we need to guarantee that at least either the world model or the measurable part of the world state is correct.

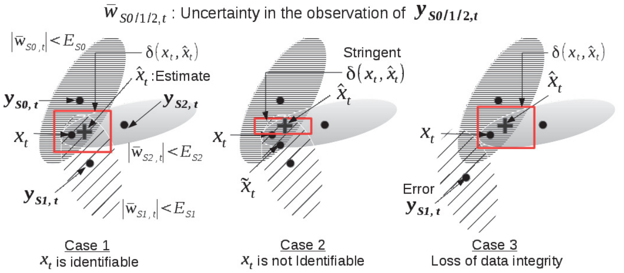


Figure 2: Three results of the estimation.

Observation Using Weak World Model

Sensor Fusion

Sensor fusion is a technique of computing an estimate $\hat{\boldsymbol{X}}_T$ of unmeasurable part of the world state $\boldsymbol{X}_T$ by using disparate sensor devices. Each device produces a component of measurable part of the world state $\boldsymbol{Y}_T$ which follows a different observation model $\boldsymbol{g}(\boldsymbol{x}_t)$ subject to a limit of observability. The devices are selected in a way that the condition of observability is maintained and the modeling error $\bar{\boldsymbol{w}}_t$ in the observation model (2) follows an independent or weakly coupled stochastic process such that the error covariance matrix $\boldsymbol{R}_t$ is sparse, or preferably, diagonal. The technique helps to reduce the joint probability of the estimation error $\bar{\boldsymbol{e}}_t$ under the assumption of the stated stochastic process. Figure 2 shows an illustrative classification of three estimation results. Here, we added an assumption to the world model that the modeling error in each observation model $\bar{\boldsymbol{w}}_{S0/S1/S2,t}$ is bounded with a certain threshold, which is formulated as $|\bar{\boldsymbol{w}}_{S0/S1/S2,t}| < E_{S0/S1/S2} \in \mathbb{R}$. This is a problem of checking identifiability of $\hat{\boldsymbol{X}}_T$ where a combined observation model (10) is the world model $M(x_t, \boldsymbol{Y})$ and measurable part $\boldsymbol{Y} \equiv \{\boldsymbol{y}_{S0/S1/S2,t}\}$ is known.

$$\bigwedge_{i=S0,S1,S2} \boldsymbol{y}_{i,t} = \boldsymbol{x}_t + \bar{\boldsymbol{w}}_{i,t} \wedge |\bar{\boldsymbol{w}}_{i,t}| < E_i \qquad (10)$$

Case-2 in Fig. 2 suggests that if we set a more stringent measure of tolerance $\delta(:)$ than that in Case-1, another state $\tilde{\boldsymbol{x}}_t$ exists such that (10) is true but $\delta(\tilde{\boldsymbol{x}}_t, \hat{\boldsymbol{x}}_t)$ is false. Case-3 suggests that if the measurement $\boldsymbol{y}_{S1,t}$ is wrong or the assumption on the boundedness of the error distribution of $\bar{\boldsymbol{w}}_{S1,t}$ is violated, then a loss of data integrity is detected.

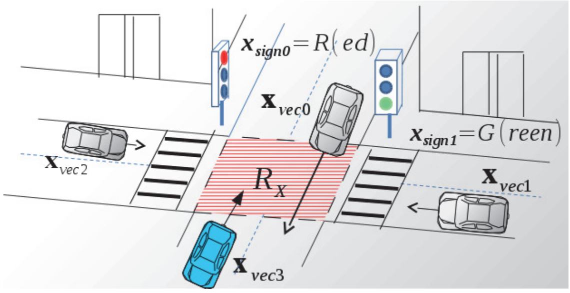


Figure 3: Estimating a world state of an enumerated type.

Reasoning Discrete and Enumerated Variables

We presumed that the world state is quantifiable, the world model (1) (2) is given as a formula of equality, and the inherently unmeasurable modeling error is either quantitatively bounded or follows the stated stochastic process. Quantitative inaccuracy of the world is of practical concern. Once we add a world state of discrete or enumerated type, the formulation of the world model is inadequate. The idea of identifiability makes it possible to describe a world model using partial knowledge about the world, which is formulated as weak constraints of inequality on the world state.

Figure 3 shows an illustrative example of estimating the state of a traffic signal $x_{sign0/1} \in \{G(reen), R(ed)\}$ at an intersection where the region is denoted as R_X in $\mathbb{R}^2$. The world state $\boldsymbol{X}$ consists of $\{\boldsymbol{x}_{vec0/1/2/3}, x_{sign0/1}\}$ and we assume it is observable. Three other vehicles and two traffic signs that control the right-of-way are in the scope of a vehicle whose position vector is $\boldsymbol{x}_{vec3} \in \mathbb{R}^2$. The position vectors $\boldsymbol{x}_{vec0/1/2/3}$ are produced by devices with physics-based measurement mechanism and are thus reliable. The world state $x_{sign0/1}$ is directly unmeasurable part of the world state produced by a machine vision system and is often unreliable. A serious false negative results from that the machine vision system fails to clip out a picture in which they are in scope and fails to even discover the world state $x_{sign0/1}$.

We use a world model that consists of a knowledge-oriented predicate (11) and empirically verified predicate (12) using a measure of tolerance (13). We compute a satisfiable assignment $\boldsymbol{X}_S \equiv (x^{SAT}_{sign0/1}\}$.

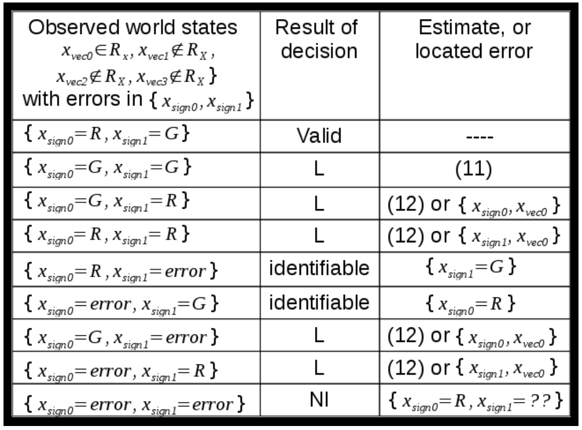
$$\begin{aligned} &\neg\left[x_{sign0} = G \wedge x_{sign1} = G\right] \wedge \\ &\left[x_{sign0/1} \in \{G(reen), R(ed)\}\right] \end{aligned} \tag{11}$$

$$\begin{aligned} &\left[x_{sign0} = R \to x_{vec1/2} \notin R_X\right] \wedge \\ &\left[x_{sign1} = R \to \boldsymbol{x}_{vec0} \notin R_X\right] \wedge \\ &\left[x_{sign0} = G \to x_{vec0} \notin R_X\right] \wedge \\ &\left[x_{sign1} = G \to \boldsymbol{x}_{vec1/2} \notin R_X\right] \end{aligned} \tag{12}$$

$$h(\boldsymbol{X}, \boldsymbol{X}_S) \equiv \bigwedge_{i=sign0,1} x_i^{SAT} = x_i \tag{13}$$

We do not need to assume a perfect precision on the knowledge about the world model and the world state. The predicate (12) is a statistically invariant partial knowledge about the world. The predicates (11) (12) are weak in a sense that the world state in the predicate is under-constrained. As we can add or refine the predicate of the partial knowledge, the world model is compostitional by design. This is a better alternative to waiting for a constructive way of building a reliable classifier. Rather, we need to describe the world model in an analytically redundant way in a sense that the world state remains identifiable even if some components in measurable part of the world state are corrupt. The degree of confidence is a quantifiable measure of probability in a sense that the correctness of the estimate is supported by the probability that the world model (11) (12) is statistically valid.

Table 2: Results of decisions and estimated world states.



Observed world states $x_{vec0} \in R_x, x_{vec1} \notin R_X,$ $x_{vec2} \notin R_X, x_{vec3} \notin R_X\}$ with errors in $\{x_{sign0}, x_{sign1}\}$	Result of decision	Estimate, or located error
$\{x_{sign0}=R, x_{sign1}=G\}$	Valid	----
$\{x_{sign0}=G, x_{sign1}=G\}$	L	(11)
$\{x_{sign0}=G, x_{sign1}=R\}$	L	(12) or $\{x_{sign0}, x_{vec0}\}$
$\{x_{sign0}=R, x_{sign1}=R\}$	L	(12) or $\{x_{sign1}, x_{vec0}\}$
$\{x_{sign0}=R, x_{sign1}=error\}$	identifiable	$\{x_{sign1}=G\}$
$\{x_{sign0}=error, x_{sign1}=G\}$	identifiable	$\{x_{sign0}=R\}$
$\{x_{sign0}=G, x_{sign1}=error\}$	L	(12) or $\{x_{sign0}, x_{vec0}\}$
$\{x_{sign0}=error, x_{sign1}=R\}$	L	(12) or $\{x_{sign1}, x_{vec0}\}$
$\{x_{sign0}=error, x_{sign1}=error\}$	NI	$\{x_{sign0}=R, x_{sign1}=??\}$

Table 2 shows the results of the estimation of the case on row 2 in Table 1 using the proposed procedure shown in Fig. 1. The problem of automatic error localization is formulated as MaxSAT (Fey 2008; Manu 2011) and solved by computing the unsatisfiable core of the formula $\boldsymbol{M}(\boldsymbol{X}, \boldsymbol{Y})$. Yet we may hardly determine uniquely that an error is in the world model or in the measurement, when the world state is not identifiable. We explore a sound resolution procedure of recovering the world state identifiable while adaptively updating the world model.

Decision of Safety Subject to Partial Observation

A safety criterion that represents a precursor of hazard is a component in $\boldsymbol{X}$. An autonomous system must monitor violations of the criterion correctly regardless of that part of $\boldsymbol{Y}$ can be transiently unavailable or deficient. A practical burden is that we need to enumerate a variety of errors in $\boldsymbol{Y}$ and to implement exhaustively one error handling procedure for each. Insufficient coverage of them or incorrect handling of the error in the implementation could result in a false decision. We can remove the burden by combining the world model with a predicate of the criterion and by checking if Boolean value of the predicate is identifiable and true in face of an imminent risk of hazard.

The degree of confidence in the Boolean value of the predicate is vital at the system level. In negligence theory (Villasenor 2014), the system must foresee at least reasonably known risk of hazard that the system is at least partially liable for and must act responsibly to avoid the risk of hazard. The ability to foresee the risk is essentially the same as the ability of deciding identifiability of the Boolean value. A loss of data integrity can impair confidence as to the risk of hazard. The Boolean value is not identifiable if part of $\boldsymbol{Y}$ is unavailable or if the predicates (11)(12) cannot be perfect. If the hazard actually could result from such limits and technically justifiable, the negligence claim against the manufacturer should be dismissed. An economic loss resulting from the limits should be covered by a compulsory insurance.

Related Work

In robotics field, a problem of reconstructing the world state is called simultaneous localization and mapping (SLAM) (Dissanayake 2001). SLAM is built on Kalman filter theory. A Kalman filter attached to a LTI system is a linear quadratic estimator. The effectiveness of the Kalman filter (Kalman 1960) depends on the accuracy of the world model and assumes only additive uncertainty. A robust Kalman filtering for a LTI system with an assumption of additive, time-varying and norm-bounded parametric uncertainty in the world model is presented in (Xie 1994). While it suggests a modified Kalman Filter with a guaranteed bound on the quadratic error (7), it cannot detect a loss of data integrity that results from a violation of the assumption about the world model and leaves the risk of producing a corrupt estimate going unnoticed. We can build an estimator using an imperfect world model with a collection of partial and imprecise knowledge about the world. A recent work on decision and control systems (Wolff 2012) opened a way toward handling the estimate $\hat{\boldsymbol{X}}$ with an uncertainty bounded by the measure of tolerance $\delta(\hat{\boldsymbol{X}}, \boldsymbol{X}_S)$.

A motivation of assuming the linear additive uncertainty originates from analytical burden of convergence analysis of the Kalman filter. The stability of EKF depends on local linearity and the filter becomes unstable beyond the local domain (Huang 2007). A work (Ghaoui 2001) addressed to reliably handling the impact of an unstable EKF on SLAM. The under-estimated mean of $\bar{\boldsymbol{e}}_t, \bar{\boldsymbol{w}}_t$ and the error covariance matrix $\boldsymbol{Q}_t, \boldsymbol{R}_t$ can also result in an unstable EKF. Unscented Kalman filter (UKF) addresses the issue (Uhlmann 2000) and thus, UKF is a better alternative to EKF (Julier 1997). Instead, we rely on the NLP solver that produces a globally convergent series of the search iterates and removed the burden of the convergence analysis. Now we can adaptively update the world model at runtime without any change in the state observation system.

To the best of our knowledge, the bounded quantifiable uncertainty has been presumed for the theoretical soundness and stability of existing estimation methods. To support the observability condition, it is presumed that measurable part of the world state $\boldsymbol{y}_t$ is available and reliable. Once the presumptions are violated, the existing pre-designed estimators get unstable but a consequent loss of data integrity goes unnoticed. Thus, the idea of checking identifiability subject to a measure of tolerance is a safer alternative. We need to make more effort in developing a formal technique for reliably reconstructing the world state subject to an imperfect world model and the mentioned limit of observability.

Conclusions

Our insight is that we can merge the original concept of observer theory with that of automated reasoning. We proposed a new way of formulating the problem of estimating the world state as satisfiability of the world state in the world model. A loss of data integrity is detected if the problem is unsatisfiable. Checking identifiability and the measure of tolerance is a better alternative to presuming observability. We showed a procedure of reliably reasoning the world state subject to the imperfect world model and an imprecise measurable part of the world state. We can constructively support the degree of confidence in the estimated world state.

References

Dissanayake, G. 2001. A solution to the simultaneous localization and map building (slam) problem. *IEEE Trans. on Robotics and Automation* 17(3):229–241.

Fey, G. 2008. Automatic fault localization for property checking. *a* 27(6):1138–1149.

Ghaoui, L. E. 2001. Robust filtering for discrete-time systems with bounded noise and parametric uncertainty. *IEEE Trans. on Automatic Control* 46(7):1084–1089.

Huang, S. 2007. Convergence and consistency analysis for extended kalman filter based slam. *IEEE Trans. on Robotics* 23(5):1036–1049.

Julier, S. J. 1997. New extension of the kalman filter to nonlinear systems. *Proc. of SPIE* 3068-VI:182.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *Trans. of the ASME* 82:35–45.

Manu, J. 2011. Cause clue clauses: error localization using maximum satisfiability. *ACM SIGPLAN Notices* 46(6):437–446.

Misener, R. 2014. Antigone: algorithms for continuous/integer global optimization of nonlinear equations. *Journal of Global Optimization* 59(2-3):503–526.

Nishi, M. 2016. Towards bounded model checking using nonlinear programming solver. *Proc. of 31st Automated Software Engineering* 560–565.

Rushby, J. 2008. Runtime certification. *Proc. of Runtime Verification* 21–35.

Smyshlyaev, A. 2005. Backstepping observers for a class of parabolic pdes. *Systems and Control Letters* 54(7):477–482.

Szegedy, Z. 2013. Intriguing properties of neural networks. *arXiv preprint* 1312.6199.

Uhlmann, J. 2000. A new method for the non linear transformation of means and covariances in filters and estimations. *IEEE Trans. on Automatic Control* 46(3):477–482.

Villasenor, J. 2014. Products liability and driverless cars. *Issues and Guiding Principles for Legislation*. Brookings.

Wachter, A. 2006. The implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming* 106(1):25–57.

Wolff, E. 2012. Robust control of uncertain markov decision processes with temporal logic specifications. *Proc. of 51st IEEE Conf. on Decision and Control* 3372–3379.

Xie, L. 1994. Robust kalman filtering for uncertain discrete-time systems. *IEEE Trans. on Automatic Control* 39(6):1310–1314.