Initialization of the Kalman Filter without Assumptions on the Initial State

Linderoth, Magnus; Soltesz, Kristian; Robertsson, Anders; Johansson, Rolf

Published in:
2011 IEEE International Conference on Robotics and Automation

DOI:
10.1109/ICRA.2011.5979684

2011

Document Version:
Peer reviewed version (aka post-print)

Initialization of the Kalman Filter without Assumptions on the Initial State

Magnus Linderoth, Kristian Soltesz, Anders Robertsson, and Rolf Johansson

***Abstract*— In absence of covariance data, Kalman filters are usually initialized by guessing the initial state. Making the variance of the initial state estimate large makes sure that the estimate converges quickly and that the influence of the initial guess soon will be negligible. If, however, only very few measurements are available during the estimation process and an estimate is wanted as soon as possible, this might not be enough.**

This paper presents a method to initialize the Kalman filter without any knowledge about the distribution of the initial state and without making any guesses.

I. Introduction

When performing state estimation on dynamical systems, the Kalman filter [1] is a very commonly used tool. Just as for other recursive algorithms, initialization is a necessary computational step and such initialization may be accomplished in a variety of different approaches, *e.g.*, probabilistic (Gaussian, Bayesian), geometric and information-theoretical approaches. In the original formulation of the Kalman filter it is assumed that the initial value of the state has a known mean value and variance. If no such data is available, the estimate will have a transient in the initial phase of the filtering. If it is possible to start the estimation well before the estimate is to be used, this causes no problem, since the estimate will have time to converge. The transient can also be reduced by letting the initial covariance matrix of the estimate have very large eigenvalues.

However, if the estimate is needed as soon as possible after the start of the estimation, it is desirable that not even the first estimates are affected by the guess of the initial state. One such example is a ball-catching robot, which has been treated in, *e.g.*, [2], [3] and [4]. A photo of such a setup is shown in Fig. 1. When a ball is thrown toward the robot, the box is moved to make the ball hit the hole. High-speed cameras provide information about the position of the ball, and a Kalman filter is used to estimate the position of the ball and predict its future trajectory. Only a limited number of measurements are available during the flight of the ball and due to the limited acceleration of the robot, it has to start moving as soon as it is possible to estimate where the ball will hit. Thus, it is essential to have a good estimate from the very start of the measurement series so the robot can get to its target in time.

Magnus Linderoth, Kristian Soltesz, Anders Robertsson, and Rolf Johansson are with the Department of Automatic Control, LTH, Lund University, Sweden, magnus.linderoth@control.lth.se

The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 Challenge 2 Cognitive Systems, Interaction, Robotics under grant agreement No 230902 - ROSETTA. This document reflects only the authors views and the European Community is not liable for any use that may be made of the information contained herein.

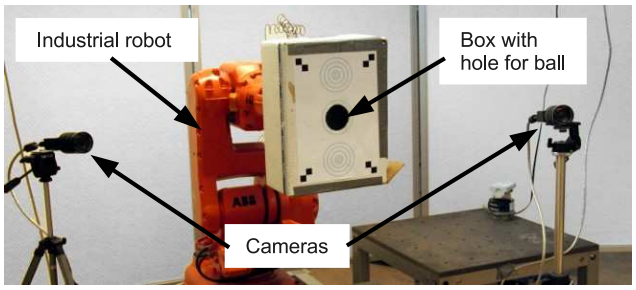

Fig. 1. Picture of a ball-catching robot.

Widely used algorithms are based on the extended Kalman filter (EKF), whose application to SLAM problems was developed in a series of seminal papers [5], [6], [7]. The EKF calculates a Gaussian posterior over the locations of environmental features and the robot itself. Information-oriented Kalman-type filters were proposed for feature-based SLAM [8], [9] and for maneuvering target tracking [10] with attention to approximation and computational speed.

Vision is commonly used for tracking in robotic applications. An introduction to available algorithms is given in [11].

Many specialized approaches for making an informed initialization of the Kalman filter have been proposed for specific problems and can be found in, *e.g.*, [12], [13] and [14]. A more general way of initializing a system on the linear state-space form (1) can be done by using information filter theory, but then with additional requirements, *e.g.*, that Φ is non-singular or that (1) is controllable [15].

This paper presents a method to initialize the Kalman filter without making any assumptions on the initial value, only assuming that the system is on the linear state-space form (1). This is done by transforming the state estimate $\hat{x}$ into a space with a basis separating directions of infinite variance from those with finite variance. A different approach for solving the same problem is presented in [16].

II. Preliminaries

A. State Space Description

The problem considered relates to discrete-time time-varying linear systems on the form

$$\begin{aligned} x(k+1) &= \Phi(k)x(k) + \Gamma(k)u(k) + v(k) \\ y(k) &= C(k)x(k) + e(k) \end{aligned} \tag{1}$$

where x is the state, u is the input and y is the measurement. The disturbances v and e are assumed to be white noise processes with zero mean values, $\mathrm{E}[v(k)v^T(k)] = R_v(k) \succeq 0$, $\mathrm{E}[e(k)e^T(k)] = R_e(k) \succ 0$, and $\mathrm{E}[v(k)e^T(k)] = 0$.

B. The Kalman Filter

The Kalman filter can be used to estimate the state of (1) recursively as described by (2), $\hat{x}(l|k)$ denoting the estimate of $x(l)$ based on measurements up to sample k and $P(l|k)$ being the covariance matrix of $\hat{x}(l|k)$.

$$\begin{aligned}
\hat{x}(k|k) &= \hat{x}(k|k-1) + \\
&\quad + K(k)\left(y(k) - C(k)\hat{x}(k|k-1)\right) \\
K(k) &= P(k|k-1)C^T(k)\cdot \\
&\quad \cdot \left(C(k)P(k|k-1)C^T(k) + R_e(k)\right)^{-1} \\
P(k|k) &= P(k|k-1) - K(k)C(k)P(k|k-1) \\
\hat{x}(k+1|k) &= \Phi(k)\hat{x}(k|k) + \Gamma(k)u(k) \\
P(k+1|k) &= \Phi(k)P(k|k)\Phi^T(k) + R_v(k)
\end{aligned} \tag{2}$$

C. Singular Value Decomposition

Consider a matrix $A \in \mathbb{R}^{m\times n}$ with $\mathrm{rank}(A) = r$. Using singular value decomposition (SVD) it can be factorized as

$$A = U\Sigma V^T \tag{3}$$

where $U \in \mathbb{R}^{m\times r}$ satisfies $U^TU = I$, $V \in \mathbb{R}^{n\times r}$ satisfies $V^TV = I$ and $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$ with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$

III. Optimal Solution of a Linear System of Equations with Noise

This section describes methods for solving linear systems of equations, which will be used later.

A. Over-determined System

Consider a system of linear equations with disturbances:

$$z = Gx + w \tag{4}$$

where $z \in \mathbb{R}^m$ and $G \in \mathbb{R}^{m\times n}$ are known and $w \in \mathbb{R}^m$ is a disturbance with $\mathrm{E}[w] = 0$ and $\mathrm{E}[ww^T] = R_w \succ 0$.

Assume that $\mathrm{rank}(G) = n < m$, *i.e.*, the system is over-determined. Let $\hat{x}$ denote the minimum variance unbiased estimate of x. The problem of finding $\hat{x}$ was solved in [17] with the solution

$$\hat{x} = (G^TR_w^{-1}G)^{-1}G^TR_w^{-1}z \tag{5}$$

$$R_x = \mathrm{E}\left[(\hat{x} - x)(\hat{x} - x)^T\right] = (G^TR_w^{-1}G)^{-1} \tag{6}$$

B. Under-determined System

Again consider the system (4), but now assume that $\mathrm{rank}(G) = r < n$, *i.e.*, the system of equations is under-determined. Still x can be partly determined. By singular value decomposition G can be factorized as

$$G = U\Sigma V^T \tag{7}$$

It is possible, as a part of the SVD algorithm, to construct

$$S = \left[\begin{array}{cc} S_f & S_i \end{array}\right] \in \mathbb{R}^{n\times n} \tag{8}$$

such that $S_f = V$ and $S^TS = I$. Define $x_f \in \mathbb{R}^r$ and $x_i \in \mathbb{R}^{n-r}$ as the unique solution to

$$x = S\left[\begin{array}{c} x_f \\ x_i \end{array}\right] = S_fx_f + S_ix_i \tag{9}$$

Note that x_f is a parametrization of the part of x that can be estimated by (4), and x_i is a parametrization of the null space of G. Inserting (9) into (4) and noting that $V^T[S_f\ S_i] = [I\ 0]$ one obtains

$$\begin{aligned}
z &= Gx + w \\
&= U\Sigma V^T\left(S_fx_f + S_ix_i\right) + w \\
&= U\Sigma x_f + w
\end{aligned} \tag{10}$$

Since $U\Sigma$ has full rank, one can solve (10) for $\hat{x}_f$, using the method described in Section III-A.

IV. Filter Initialization

A. State Partitioning

During the initialization of the Kalman filter there may be times when the variance of the state estimate is finite in some directions of the state space and infinite in other directions. To handle this case the state can by a linear transformation be reoriented to a space where the directions with infinite variance are orthogonal to as many base vectors as possible. Let the state $\bar{x}$ in this alternative space be defined by

$$T\bar{x} = x \tag{11}$$

where $T \in \mathbb{R}^{n\times n}$ and $T^TT = I$. Let $\hat{\bar{x}}$ denote the estimate of $\bar{x}$, and denote the estimation error by

$$\tilde{\bar{x}} = \bar{x} - \hat{\bar{x}} \tag{12}$$

Throughout the paper it is assumed that the estimates are designed to be unbiased, *i.e.*, so that $\mathrm{E}\left[\tilde{\bar{x}}\right] = 0$. The state can be partitioned as

$$\bar{x} = \left[\begin{array}{c} \bar{x}_f \\ \bar{x}_i \end{array}\right] \tag{13}$$

such that the covariance-matrix of $\hat{\bar{x}}_f$ is finite and the variance of $\hat{\bar{x}}_i$ is infinite in all directions. Define n_f and n_i such that $n_f + n_i = n$, and $\bar{x}_f \in \mathbb{R}^{n_f}$, $\bar{x}_i \in \mathbb{R}^{n_i}$. The transformation matrix T can similarly be partitioned as

$$T = \left[\begin{array}{cc} T_f & T_i \end{array}\right],\quad T_f \in \mathbb{R}^{n\times n_f},\quad T_i \in \mathbb{R}^{n\times n_i} \tag{14}$$

Since T is orthonormal we have

$$T^{-1} = \left[\begin{array}{cc} T_f & T_i \end{array}\right]^{-1} = \left[\begin{array}{cc} T_f & T_i \end{array}\right]^T = \left[\begin{array}{c} T_f^T \\ T_i^T \end{array}\right] \tag{15}$$

Note for future reference that

$$\left[\begin{array}{c} \bar{x}_f \\ \bar{x}_i \end{array}\right] = \left[\begin{array}{c} T_f^T \\ T_i^T \end{array}\right]x \tag{16}$$

and

$$x = T_f\bar{x}_f + T_i\bar{x}_i \tag{17}$$

which shows that T_f spans the directions in which $\hat{x}$ has finite variance and T_i spans the directions in which $\hat{x}$ has infinite variance.

Let $\bar{P}_f$ denote the covariance matrix of $\hat{\bar{x}}_f$:

$$\bar{P}_f = \mathrm{E}\left[(\hat{\bar{x}}_f - \bar{x}_f)(\hat{\bar{x}}_f - \bar{x}_f)^T\right] \tag{18}$$

In the remainder of the paper all quantities defined in this section may be appended with time indices so that, *e.g.*, $\hat{\bar{x}}_f(l|k)$ is the estimate of $\bar{x}_f(l)$ based on measurements up to sample k, and $\hat{x}(l|k) = T(l|k)\hat{\bar{x}}(l|k)$. Note, however, the slightly different case $x(l) = T(l|k)\bar{x}(l|k)$. The actual state x has only a single time index, since the second time index is meaningful only for estimates. Still $\bar{x}$ has two time indices to indicate which T was used for the transformation.

To conclude, all the knowledge about $\hat{x}(l|k)$ can be fully specified by $T(l|k)$, $\hat{\bar{x}}_f(l|k)$ and $\bar{P}_f(l|k)$.

B. Time Step

Assume that $T(k|k)$, $\hat{\bar{x}}_f(k|k)$ and $\bar{P}_f(k|k)$ are known. The state model (1) gives the time update

$$x(k+1) = \Phi(k)x(k) + \Gamma(k)u(k) + v(k). \tag{19}$$

The purpose of the time step is to calculate $T(k+1|k)$, $\hat{\bar{x}}_f(k+1|k)$ and $\bar{P}_f(k+1|k)$.

Choose $T(k+1|k)$ such that

$$T_f^T(k+1|k)\Phi(k)T_i(k|k) = 0 \tag{20}$$

$$n_f(k+1|k) = n - \operatorname{rank}\left(\Phi(k)T_i(k|k)\right) \tag{21}$$

$$T^T(k+1|k)T(k+1|k) = I \tag{22}$$

This can be interpreted as finding a $T(k+1|k)$ such that it is orthonormal and its $n_f(k+1|k)$ leftmost columns span the left null space of $\Phi(k)T_i(k|k)$, which can be done, *e.g.*, by means of SVD. Note that

$$n_f(k+1|k) \geq n_f(k|k) \tag{23}$$

where strict inequality holds if and only if $\Phi(k)$ is singular and $\mathcal{N}(\Phi(k)) \cap \mathcal{R}(T_i(k|k)) \neq \emptyset$.

Premultiplying (19) with $T_f^T(k+1|k)$ gives

$$\begin{aligned}
&\bar{x}_f(k+1|k) \\
&= T_f^T(k+1|k)\Phi(k)x(k) \\
&\quad + T_f^T(k+1|k)\Gamma(k)u(k) \\
&\quad + T_f^T(k+1|k)v(k) \\
&= T_f^T(k+1|k)\Phi(k)\left(T_f(k|k)\bar{x}_f(k|k) + T_i(k|k)\bar{x}_i(k|k)\right) \\
&\quad + T_f^T(k+1|k)\Gamma(k)u(k) \\
&\quad + T_f^T(k+1|k)v(k) \\
&= T_f^T(k+1|k)\Phi(k)T_f(k|k)\bar{x}_f(k|k) \\
&\quad + T_f^T(k+1|k)\Gamma(k)u(k) \\
&\quad + T_f^T(k+1|k)v(k)
\end{aligned} \tag{24}$$

where the second and third equalities result from (17) and (20) respectively. Here the advantage of choosing $T(k+1|k)$ according to (20) becomes clear. Because of this choice $\bar{x}_f(k+1)$ is independent of $\bar{x}_i(k)$ and only depends on quantities with finite variance. Condition (21) guarantees that $T_f(k+1)$ has the highest possible rank.

Motivated by (24), let the update of the state estimate be defined by

$$\begin{aligned}
\hat{\bar{x}}_f(k+1|k) &= T_f^T(k+1|k)\Phi(k)T_f(k|k)\hat{\bar{x}}_f(k|k) \\
&\quad + T_f^T(k+1|k)\Gamma(k)u(k)
\end{aligned} \tag{25}$$

The estimation error is then given by

$$\begin{aligned}
\tilde{\bar{x}}_f(k+1|k) &= \bar{x}_f(k+1|k) - \hat{\bar{x}}_f(k+1|k) \\
&= T_f^T(k+1|k)\Phi(k)T_f(k|k)\tilde{\bar{x}}_f(k|k) \\
&\quad + T_f^T(k+1|k)v(k)
\end{aligned} \tag{26}$$

It is easily verified that $\mathrm{E}\left[\tilde{\bar{x}}_f(k+1|k)\right] = 0$ as required. The variance of the estimate becomes

$$\begin{aligned}
\bar{P}_f(k+1|k) &= \mathrm{E}\left[\tilde{\bar{x}}_f(k+1|k)\tilde{\bar{x}}_f^T(k+1|k)\right] \\
&= Q\bar{P}_f(k|k)Q^T \\
&\quad + T_f^T(k+1|k)R_v(k)T_f(k+1|k),
\end{aligned} \tag{27}$$

$$Q = T_f^T(k+1|k)\Phi(k)T_f(k|k)$$

C. Correction Step

Assume that $T(k|k-1)$, $\hat{\bar{x}}_f(k|k-1)$ and $\bar{P}_f(k|k-1)$ are known. The state model (1) gives the measurement

$$y(k) = C(k)x(k) + e(k) \tag{28}$$

The purpose of the correction step is to calculate $T(k|k)$, $\hat{\bar{x}}_f(k|k)$ and $\bar{P}_f(k|k)$.

Combining (12) and (16) gives

$$\begin{aligned}
\hat{\bar{x}}_f(k|k-1) &= \bar{x}_f(k) - \tilde{\bar{x}}_f(k|k-1) \\
&= T_f^T(k|k-1)x(k) - \tilde{\bar{x}}_f(k|k-1)
\end{aligned} \tag{29}$$

Equations (28) and (29) can be formulated as a single linear system of equations:

$$\begin{aligned}
&\underbrace{\begin{bmatrix} y(k) \\ \hat{\bar{x}}_f(k|k-1) \end{bmatrix}}_{z} \\
&= \underbrace{\begin{bmatrix} C(k) \\ T_f^T(k|k-1) \end{bmatrix}}_{G} x(k) + \underbrace{\begin{bmatrix} e(k) \\ -\tilde{\bar{x}}_f(k|k-1) \end{bmatrix}}_{w}
\end{aligned} \tag{30}$$

which can be solved by the method described in Section III-B with

$$R_w = \begin{bmatrix} R_e(k) & 0 \\ 0 & \bar{P}_f(k|k-1) \end{bmatrix} \tag{31}$$

The solution is given by

$$T(k|k) = S \tag{32}$$

$$\hat{\bar{x}}_f(k|k) = (\Sigma U^T R_w^{-1} U\Sigma)^{-1}\Sigma U^T R_w^{-1} z \tag{33}$$

$$\bar{P}_f(k|k) = (\Sigma U^T R_w^{-1} U\Sigma)^{-1} \tag{34}$$

$$n_f(k|k) = \operatorname{rank}(\Sigma) = \operatorname{rank}(G) \tag{35}$$

where U, Σ and S are defined in (7) and (8).

From the definition of G in (30) it can be seen that

$$\operatorname{rank}(G) \geq \operatorname{rank}\left(T_f(k|k-1)\right) \tag{36}$$

Further, the orthonormality of $T(k|k-1)$ in combination with (14) gives

$$\operatorname{rank}\left(T_f(k|k-1)\right) = n_f(k|k-1) \tag{37}$$

Combining (35) - (37) results in

$$n_f(k|k) \geq n_f(k|k-1) \tag{38}$$

where equality holds if and only if $\mathcal{R}(C^T(k)) \subseteq \mathcal{R}\left(T_f(k|k-1)\right)$. Equations (23) and (38) together show that n_f never decreases and give conditions for when n_f increases.

If G has full rank the variance of $\hat{x}(k|k)$ will be finite in all directions and $n_f(k|k) = n$.

Remark: For $n_f(k|k-1) = n$ and $T_f(k|k-1) = I$ it can be shown that the solution of (30) is equivalent to the correction step of the ordinary Kalman filter (2).

D. How to start and when to stop

Assuming that nothing is known about x when the estimation starts out ($n_f = 0$), the first thing to do is to apply the correction step to the first measurement. The lower blocks of the matrices z and G, and all blocks except R_e in R_w, will then be empty.

If the initial variance of x is infinite only in some directions ($0 < n_f < n$), the available information can be represented by a triple of matrices, T, $\hat{\bar{x}}_f$ and $\bar{P}_f$, and then plugged into the algorithm without any modification.

If the measurements provide enough information, the variance of the estimate will be finite in all directions ($n_i = 0$) after a number of iterations of the filter. Then it is no longer necessary to use the algorithm described in this section and one can just as well use the standard Kalman filter (2), since the methods are equivalent for $n_i = 0$.

V. Simulation

To illustrate the use of the filter, consider a ball flying in a gravity field and with negligible air drag. The ball is tracked by a vision system, where each camera can provide an estimate of the line that intersects both the ball and the focal point of the camera, but no depth information is available. The process model is given on state space form (1) with

$$\Phi(k) = \begin{bmatrix} 1 & 0 & 0 & h & 0 & 0 \\ 0 & 1 & 0 & 0 & h & 0 \\ 0 & 0 & 1 & 0 & 0 & h \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \forall k \tag{39}$$

$$\Gamma(k) = \begin{bmatrix} 0 \\ 0 \\ h^2/2 \\ 0 \\ 0 \\ h \end{bmatrix}, \forall k \tag{40}$$

$$u(k) = -g, \ \forall k \tag{41}$$

$$R_v(k) = 10^{-6} I_{6\times 6}, \ \forall k \tag{42}$$

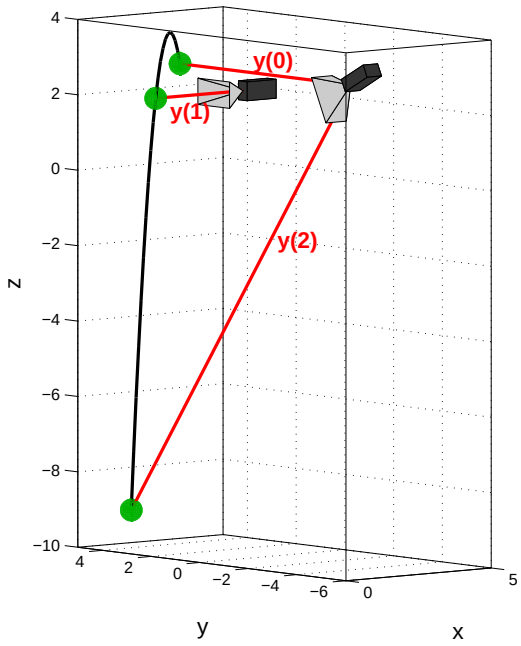

Fig. 2. Simulated ball trajectory and measurements marked as green balls.

where the state vector $x = [x_b \ \ y_b \ \ z_b \ \ \dot{x}_b \ \ \dot{y}_b \ \ \dot{z}_b]^T$ consists of three positions followed by three velocities. To make the example easy to follow, use the time step $h = 1$ and the earth gravitation $g = 10$. Let the initial state of the system be $x(0) = [1 \ \ 2 \ \ 3 \ \ 0 \ \ 1 \ \ 4]^T$. The trajectory of the ball is shown as a black curve in Fig. 2. The positions of the ball at the measuring instants are marked with green circles and the corresponding lines that are extracted from the images are marked in red. One camera observes the ball at time steps 0 and 2, and a second camera observes the ball at time step 1. The simulated measurements are given by

$$y(0) = \begin{bmatrix} 1 \\ 3 \end{bmatrix}, C(0) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{43}$$

$$y(1) = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, C(1) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{44}$$

$$y(2) = \begin{bmatrix} 1 \\ -1.9 \end{bmatrix}, C(2) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.4 & 0.3 & 0 & 0 & 0 \end{bmatrix} \tag{45}$$

$$R_e(0) = R_e(1) = R_e(2) = 10^{-4} I_{2\times 2} \tag{46}$$

Performing the state estimation on the given data gives the following results:

$$T_f(0|0) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{47}$$

$$T_i(0|0) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{48}$$

$$\hat{\bar{x}}_f(0|0) = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \tag{49}$$

$$\bar{P}_f(0|0) = 10^{-4} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{50}$$

$$T_f(1|0) = \begin{bmatrix} 0.707 & 0 \\ 0 & 0 \\ 0 & 0.707 \\ -0.707 & 0 \\ 0 & 0 \\ 0 & -0.707 \end{bmatrix} \tag{51}$$

$$T_i(1|0) = \begin{bmatrix} 0 & 0.707 & 0 & 0 \\ -0.707 & 0 & 0 & 0.707 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0.707 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0.5 & 0 & 0 & 0.5 \end{bmatrix} \tag{52}$$

$$\hat{\bar{x}}_f(1|0) = \begin{bmatrix} 0.707 \\ 5.657 \end{bmatrix} \tag{53}$$

$$\bar{P}_f(1|0) = 10^{-4} \begin{bmatrix} 0.51 & 0 \\ 0 & 0.51 \end{bmatrix} \tag{54}$$

$$T_f(1|1) = \begin{bmatrix} 0.707 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -0.707 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{55}$$

$$T_i(1|1) = \begin{bmatrix} 0 & -0.707 \\ 0 & 0 \\ 0 & 0 \\ 0 & -0.707 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \tag{56}$$

$$\hat{\bar{x}}_f(1|1) = \begin{bmatrix} 0.707 \\ 3 \\ 2 \\ -6 \end{bmatrix} \tag{57}$$

$$\bar{P}_f(1|1) = 10^{-4} \begin{bmatrix} 0.51 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 2.02 \end{bmatrix} \tag{58}$$

$$T_f(2|1) = \begin{bmatrix} 0.447 & 0 & 0 & 0 \\ 0 & 0.707 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -0.894 & 0 & 0 & 0 \\ 0 & -0.707 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{59}$$

$$T_i(2|1) = \begin{bmatrix} 0 & -0.894 \\ 0.707 & 0 \\ 0 & 0 \\ 0 & -0.447 \\ 0.707 & 0 \\ 0 & 0 \end{bmatrix} \tag{60}$$

$$\hat{\bar{x}}_f(2|1) = \begin{bmatrix} 0.447 \\ 2.121 \\ -9 \\ -16 \end{bmatrix} \tag{61}$$

$$\bar{P}_f(2|1) = 10^{-4} \begin{bmatrix} 0.214 & 0 & 0 & 0 \\ 0 & 0.51 & 0 & 0 \\ 0 & 0 & 5.03 & 3.02 \\ 0 & 0 & 3.02 & 2.03 \end{bmatrix} \tag{62}$$

$$T_f(2|2) = I_{6\times 6} \tag{63}$$

$$T_i(2|2) \in \mathbb{R}^{6\times 0} \tag{64}$$

$$\hat{\bar{x}}_f(2|2) = \begin{bmatrix} 1 \\ 2 \\ -9 \\ 0 \\ -1 \\ -16 \end{bmatrix} \tag{65}$$

$$\bar{P}_f(2|2)$$

$$= 10^{-4} \begin{bmatrix} 1 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 9.08 & -3.77 & 0 & 9.08 & -2.27 \\ 0 & -3.77 & 5.03 & 0 & -3.77 & 3.02 \\ 0.5 & 0 & 0 & 0.518 & 0 & 0 \\ 0 & 9.08 & -3.77 & 0 & 10.1 & -2.27 \\ 0 & -2.27 & 3.02 & 0 & -2.27 & 2.03 \end{bmatrix} \tag{66}$$

Most insight on the estimation progress is given by studying T_f. The first measurement locates the ball on a line in the y_b-direction, which gives information about the position in the x_b- and z_b-directions. This is reflected in the columns of $T_f(0|0)$. After the time step the position is no longer known. Only linear combinations of the positions and velocities can be determined, as seen in $T_f(1|0)$. With the second measurement y_b and z_b are given. Since it is the second measurement in the z-direction, $\dot{z}_b$ can be determined. Still, no information about $\dot{y}_b$ is available and hence y_b is no longer known after the time step, as indicated by $T_f(2|1)$. The last measurement gives information in the remaining directions with infinite variance, and thus $T_f(2|2)$ spans the entire $\mathbb{R}^n$ and an estimate $\hat{x}(2|2) = T_f(2|2)\hat{\bar{x}}(2|2)$ can finally be calculated.

VI. Discussion

Alternative frameworks to the one used in this paper would be Bayesian networks or conditional expectations [18].

The reason for doing the partitioning suggested in this paper, is the difficulty of representing matrices with infinite singular values. An alternative approach to this is used in information filters [15]. Instead of the covariance matrix P and the state estimate $\hat{x}$, the information matrix $Y = P^{-1}$ and the information vector $\hat{y} = P^{-1}\hat{x}$ are used to represent the information about the system. If no information about the state is available this is conveniently represented by $Y = 0$. Measurement updates get a very simple form with information filters. However, the time update is complicated and does not work for a general system on the form (1), as stated in the introduction.

The equations (20) - (22) and (30) do not in general have a unique solution for T. In this paper SVD based methods for solving the equations are suggested, but other methods can be used. The transformation T can be replaced by any T' fulfilling (11) - (18) such that $\mathcal{R}(T_f) = \mathcal{R}(T'_f)$. Of course $\hat{\bar{x}}_f$ and $\bar{P}_f$ have to be modified accordingly. In the example in Section V the T matrices were chosen to align the base vectors of $\bar{x}_f$ with the base vectors of the original state space as far as possible to improve human readability.

The presented initialization procedure is useful when very little is known about the initial state. If a priori knowledge is available, this should of course be used to improve the estimate.

The state x, and hence also $\bar{x}_f$ and $\bar{x}_i$, are assumed to have exact and finite values, however not known exactly. More specifically it is assumed that no information at all is available about $\bar{x}_i$, which is modeled as $\hat{\bar{x}}_i$ having infinite variance.

It is in general not meaningful to give any numerical values of $\hat{x}$ if $n_f < n$. To see this recall (17). If a row in T_i has any non-zero element, the corresponding element of $\hat{x}$ is completely unknown. The knowledge about x can, however be described by T, $\hat{\bar{x}}_f$ and $\bar{P}_f$.

Even though it may not be possible to calculate any value of $\hat{x}$ in the original state space, the information in $\hat{\bar{x}}_f$ can still be useful. For instance it may be of interest to know the altitude of an aerial vehicle before its longitude and latitude can be estimated.

As an example of state partitioning consider the scenario where an object o is known to be near a given line l in 3D-space, but nothing is known about its position along the line. Choose a coordinate system such that its first two base vectors are orthogonal to l and the third base vector is parallel to l. The position of o can then be partly described by the first two components with a finite covariance matrix, even though the variance in the direction of the third component (parallel to l) is infinite.

VII. Conclusions

A new way of initializing the Kalman filter has been presented, making it possible to calculate a state estimate that is not influenced by any guess of the initial value of the state. Instead the estimate can be determined completely based on the first measurements.

References

[1] R. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME–Journal of Basic Engineering*, vol. 82, no. Series D, pp. 35–45, 1960.

[2] U. Frese, B. Baeuml, G. Schreiber, I. Schaefer, M. Haehnle, G. Hirzinger, and S. Haidacher, "Off-the-shelf vision for a robotic ball catcher," in *In Proceedings of the IEEE/RSJ Intl. Conference on Intelligent Robots and Systems Oktober 2001, Maui*, 2001, pp. 1623–1629.

[3] M. Linderoth, A. Robertsson, K. Åström, and R. Johansson, "Vision based tracker for dart-catching robot," in *Preprints 9th IFAC International Symposium on Robot Control (SYROCO'09)*, Gifu, Japan, September 9-12, 2009, pp. 883-888, Sept. 2009.

[4] ——, "Object tracking with measurements from single or multiple cameras," in *Proc. International Conference on Robotics and Automation (ICRA2010)*, Anchorage, AK, May 2010, pp. 4525–4530.

[5] R. C. Smith and P. Cheeseman, "On the representation and estimation of spatial uncertainly," *Int. J. Rob. Res.*, vol. 5, pp. 56–68, December 1986.

[6] P. Moutarlier and R. Chatila, "An experimental system for incremental environment modelling by an autonomous mobile robot," in *The First International Symposium on Experimental Robotics I*. London, UK: Springer-Verlag, 1990, pp. 327–346.

[7] R. Smith, M. Self, and P. Cheeseman, *Estimating uncertain spatial relationships in robotics*. New York, NY, USA: Springer-Verlag New York, Inc., 1990, pp. 167–193.

[8] M. R. Walter, R. M. Eustice, and J. J. Leonard, "Exactly sparse extended information filters for feature-based SLAM," *IJRR*, vol. 26, no. 4, pp. 335–359, 2007.

[9] S. Thrun, Y. Liu, D. Koller, A. Y. Ng, Z. Ghahramani, and H. Durrant-Whyte, "Simulataneous localization and mapping with sparse extended information filters," *IJRR*, vol. 23, no. 7–8, pp. 693–716, 2004.

[10] M. Farooq and S. Bruder, "Information type filters for tracking a maneuvering target," *Trans. on Aerospace and Electronic Systems*, vol. 26, no. 3, pp. 441–454, 1990.

[11] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An Invitation to 3-D Vision: From Images to Geometric Models*. SpringerVerlag, 2003.

[12] E. Einhorn, C. Schröter, H.-J. Böhme, and H.-M. Gross, "A hybrid Kalman filter based algorithm for real-time visual obstacle detection," *Proc. 52nd Int. Scientific Colloquium (IWK)*, vol. II, pp. 353–358, 2007.

[13] J. Hyland, "An iterated-extended Kalman filter algorithm for tracking surface and sub-surface targets," *OCEANS '02 MTS/IEEE*, vol. 3, pp. 1283–1290, 2002.

[14] L. B. Weiner, "Kalman filter initialization with large initial uncertainty and strong measurement nonlinearity," *Proceedings of the Region 3 Conference and Exhibit, Huntsville, AL; United States*, pp. 150–151, 1981.

[15] T. Kailath, A. Sayed, and B. Hassibi, "Linear estimation," *Prentice Hall, Upper Saddle River, NJ*, pp. 310–361, 2000.

[16] P. Hagander, "Operator factorization and other aspects of the analysis of linear systems," Ph.D. dissertation, 7336, Department of Automatic Control, Lund University, Sweden, Dec. 1973.

[17] R. Plackett, "Some theorems in least squares," *Biometrika 37 (1-2)*, pp. 149–157, 1950.

[18] R. Neapolitan, "Learning Bayesian networks," *Prentice Hall, Upper Saddle River, NJ*, 2004.