

# Nonlinear Observers via Regularized Dynamic Inversion

Anthony Yezzi and Erik I. Verriest  
School of Electrical and Computer Engineering  
Georgia Institute of Technology  
Atlanta, GA 30332, USA

**Abstract**—We propose a nonlinear observer framework in which the state estimate  $\hat{x}_k$  of a discrete time dynamical system is chosen to simultaneously minimize the final output residual  $y_k - h(x_k, u_k, t)$  while at the same time remaining close to the predicted apriori estimate  $\hat{x}_k^-$ . This latter constraint regularizes the problem of trying to instantaneously invert an overdetermined system with more states than outputs by putting a cost on the difference between the predicted and final state estimates. As the the apriori estimates used to regularize the inversion process are obtained from the modelled system dynamics, we refer to this approach as *regularized dynamic inversion*. We discuss a class of nonlinearities for which this style observer yields a computationally feasible filtering algorithm with significantly superior performance compared with its Luenberger style counterparts (EKF) in two scenarios.

## I. INTRODUCTION

The problem of inversion of a mapping from a high dimensional to a low dimensional space is ill-posed. The problem has a long history, dating back to Legendre (1805) and Gauss (1809). In the linear case the problem reduces to finding a ‘solution’ to an inconsistent overdetermined set of equations: We want to find a vector  $x$  so that for given  $H$  and  $y$ , both  $Hx$  and  $y$  are equal. However, unless  $y$  is in the range space of  $H$  equality is impossible. Hence one searches then for a ‘solution’  $x$  such that  $Hx$  and  $y$  are close in some sense. In the nonlinear overdetermined problem, we want  $h(x)$  and  $y$  to be close. If we restate the problem as  $y = h(x) + v$  then, it may be interpreted as having the result of some nonlinear measurements of  $x$  where  $v$  is the associated measurement error (this need not have a stochastic model). One simply associates a misadjustment with the measurement and, assuming that the measurement model is correct, determines a solution of the ‘doctored’ measurement  $y - v$ . Obviously, our belief in the validity of the measurement dictates that one should keep the amount of doctored minimal. We are then led to put a metric on the measurement space, which in principle may also depend on the measurement itself. A Riemannian metric,  $W(\cdot) : \mathbf{T}_y Y \times \mathbf{T}_y Y \rightarrow \mathbb{R}$  induces the norm  $\|v\|_{W(y)}^2$  on the perturbations  $v$  in the tangent space  $\mathbf{T}_y Y$  at  $y$  in the measurement space  $Y$ . A necessary condition for  $x^*$  being a minimizer of the misadjustment is

$$\nabla_x h(x^*)W(y)(y - h(x^*)) = 0.$$

In general, when the equations are made consistent by a suitable perturbation of the measurement, many (in fact infinitely many) solutions may be possible. This situation is

not very favorable in problems of tracking, i.e., when actual *dynamics* are involved. Such is for instance the case when a moving or rotating object is to be tracked against some background, or in the problem of tracking a signal that is known to evolve continuously. This is the typical filtering problem. In these cases the resolution as a *unique* solution is important, and can be accomplished by *regularization*. Here some prior information about the variable  $x$ , say  $\bar{x}$  is assumed. Then rather than minimizing just  $\|y - h(x)\|$ , one incorporates also the misadjustment in the prediction with some suitable weight in the space where  $x$  lives. The most general case involves thus a minimization of

$$J(z; y, \bar{x}) = \|y - h(z)\|_{W(y)}^2 + \|\bar{x} - z\|_{\Pi(\bar{x})}^2.$$

which yields the best trade-off between the misadjustments of the predicted value and the measurements. This idea was already explored in ([2]).

Note that the use Riemannian metrics has also been implemented in minimum sensitivity design problems, which may be seen as a dual to the present problem [7]. In principle the above step is a *corrector*-step. Many estimation algorithms indeed follow an ubiquitous predictor-corrector scheme. See for instance [1]. As shown in the next section, in the linear case both the predictor and corrector step are linear operations. The equivalence with the Kalman filter follows from an judicious choice of the weights, and are shown to be associated with the exact optimal least squares filter in the linear Gaussian case.

## II. LINEAR CASE

For the sake of completeness as well as in the interest of presenting a self-contained work, we begin with the presentation of the regularized dynamic inversion observer in the linear case. We see that in this special case, while appearing to have a different structure than the classical Luenberger style observer [3], it is actually equivalent. We note though, that this relationship between the two observers in the linear case is discussed in the book of Kailath, Sayed, and Hassibi [6] where the regularized dynamic inversion method is referred to as regularized least squares.

We will also show that when the regularizing term is weighted optimally (for minimal error variance), the observer is equivalent to the Kalman filter. While this is not a standard Kalman filter derivation, we are aware of at least one similar development [1].

### A. Observer structure

Assume we start with an unbiased estimate  $\hat{x}_k$  of the state  $x_k$  for the process whose state and output are modelled by

$$x_{k+1} = Ax_k + Bu_k + w_k \quad (1a)$$

$$y_k = Hx_k + v_k \quad (1b)$$

where  $w_k$  and  $v_k$  represent independent, zero-mean model and measurement white noise with covariance matrices  $Q_k$  and  $R_k$  respectively. Our goal is to arrive at an unbiased minimum variance estimate  $\hat{x}_{k+1}$ . We start in the standard way a-priori predicted estimate  $\hat{x}_{k+1}^-$  using the previous estimate  $\hat{x}_k$  and input  $u_k$  combined with the modeled system dynamics. We then attempt to invert the new measurement  $y_{k+1}$ , which may not be instantaneously, directly invertible, by using this a-priori estimate  $\hat{x}_{k+1}^-$  as a regularizer. This regularized inversion is carried out by minimizing a quadratic energy consisting of two terms, the first of which is simply the residual between the system modelled and actual measured outputs, the second of which is the discrepancy between the predicted and final state estimates. As such, the resulting a-posteriori estimate  $\hat{x}_{k+1}$ , unlike the a-priori estimate, will depend upon the new measurement.

$$\hat{x}_{k+1}^- = A\hat{x}_k + Bu_k \quad (2a)$$

$$\hat{x}_{k+1} = \arg \min_z \|y_{k+1} - Hz\|^2 + \|z - \hat{x}_{k+1}^-\|_W^2 \quad (2b)$$

The matrix  $W$  denotes a positive definite matrix which weights the norm in the second term according to the level of confidence we have in the predicted state  $\hat{x}_{k+1}^-$ . Note that, in this linear case, the minimizer of the quadratic energy has a unique and closed form solution.

$$\hat{x}_{k+1} = (H^T H + W)^{-1}(W\hat{x}_{k+1}^- + H^T y_{k+1}) \quad (3)$$

### B. Choosing the optimal prior weight $W$

Note that unlike the traditional gain matrix of the Luenberger style observer, which maps the error of the system output estimate into a correction to the system state estimate (an ill-posed problem when the state space has higher dimension than the output space), the prior weighting matrix  $W$  has a very clear and natural interpretation. Namely, it reflects how much we trust the various components of the a-priori estimate  $\hat{x}_{k+1}^-$ .

Given its straightforward interpretation, a number of heuristics may be derived to select reasonable choices of  $W$  from application specific considerations. However, in the event that the system and measurement noise covariances  $Q_k$  and  $R_k$  are known (allowing us to compute the covariance  $P_{k+1}$  of our state estimate  $\hat{x}_{k+1}$ ), then the natural approach would be to choose  $W$  to yield a minimum variance estimate at time  $k+1$  (in the same way that the gain matrix is chosen in the Kalman filter).

#### 1) Some notation:

$$\begin{aligned} M &\doteq H^T H + W, \\ \tilde{x}_{k+1}^- &\doteq \hat{x}_{k+1}^- - x_{k+1}, & P_{k+1}^- &\doteq E\{\tilde{x}_{k+1}^- \tilde{x}_{k+1}^{-T}\}, \\ \tilde{x}_{k+1} &\doteq \hat{x}_{k+1} - x_{k+1}, & P_{k+1} &\doteq E\{\tilde{x}_{k+1} \tilde{x}_{k+1}^T\} \end{aligned}$$

#### 2) Some expressions:

$$\begin{aligned} \tilde{x}_{k+1}^- &= A\tilde{x}_k - w_k, \\ P_{k+1}^- &= AP_k A^T + Q_k, \\ \tilde{x}_{k+1} &= (I - M^{-1}H^T H)\tilde{x}_{k+1}^- + M^{-1}H^T v_{k+1} \\ P_{k+1}^+ &= P_{k+1}^- - 2M^{-1}H^T H P_{k+1}^- \\ &\quad + M^{-1}H^T (H P_{k+1}^- H^T + R_{k+1}) H M^{-1} \end{aligned} \quad (4)$$

3) *Differentiating in  $W$* : We now choose  $W$  to minimize the trace of the a-posteriori error covariance matrix as is normally done with the gain  $K$  in the Luenberger observer to derive the Kalman filter. We compute the derivative of  $\text{Tr}\{P_{k+1}^+\}$ , using expression (4), with respect to  $W$ .

$$\begin{aligned} \frac{d}{dW} \text{Tr}\{P_{k+1}^+\} &= 2M^{-1}P_{k+1}^- H^T H M^{-1} \\ &\quad - 2M^{-2}H^T (H P_{k+1}^- H^T + R_{k+1}) H M^{-1} \end{aligned}$$

Setting this result equal to zero yields

$$\begin{aligned} H^T H P_{k+1}^- &= H^T (H P_{k+1}^- H^T + R_{k+1}) H M^{-1} \\ H P_{k+1}^- &= (H P_{k+1}^- H^T + R_{k+1}) H M^{-1} \end{aligned} \quad (5)$$

where (5) follows under the assumption that  $H$  has full rank. Now substituting  $M = H^T H + W$  we get

$$\begin{aligned} H P_{k+1}^- (H^T H + W) &= (H P_{k+1}^- H^T + R_{k+1}) H \\ H P_{k+1}^- W &= R_{k+1} H. \end{aligned} \quad (6)$$

### C. Relationship to the Kalman filter

We now show that in this linear case, regularized dynamic inversion with optimally chosen prior weighting  $W$  is equivalent to the linear Kalman filter.

First, note that expression (3) for the closed form regularized inverted output measurement may be rewritten as

$$\hat{x}_{k+1} = \hat{x}_{k+1}^- + M^{-1}H^T (y_{k+1} - H\hat{x}_{k+1}^-), \quad (7)$$

allowing us to replace the minimization step (2b) with the standard Luenberger correction step where  $M^{-1}H^T$  plays the role of the gain matrix. Next, if instead of simplifying (5) to obtain the optimality condition (6) for  $W$ , we simply rearrange its terms, we obtain the traditional Kalman gain.

$$M^{-1}H^T = P_{k+1}^- H^T (H P_{k+1}^- H^T + R_{k+1})^{-1} \quad (8)$$

Therefore, in the linear case, simply solving (8) for  $K_{Kalman} = M^{-1}H^T$  and substituting into (7) bypasses the need to solve for  $W$  directly. Further, substitution into expression (4) for  $P_{k+1}^+$  gives the standard Kalman update.

## III. THE GENERAL CASE

### A. Observer structure

We now consider a more general nonlinear process for which we have an initial estimate  $\hat{x}_k$  as whose dynamics are governed by the following equations.

$$x_{k+1} = f(x_k, u_k) + w_k \quad (9a)$$

$$y_k = h(x_k) + v_k \quad (9b)$$

Our observer model starts as before using the modelled dynamics together with the present state estimate  $\hat{x}_k$  to obtain the a-priori predicted estimate  $\hat{x}_{k+1}^-$ , independent of the new measurement. Then as before, we attempt to invert the new measurement  $y_{k+1}$ , using this a-priori estimate  $\hat{x}_{k+1}^-$  as a prior in order to regularized what may be an ill-posed inversion problem. The same type of quadratic energy is employed, but in this general case, there is not necessarily a closed form unique solution.

$$\hat{x}_{k+1}^- = f(\hat{x}_k, u_k) \quad (10a)$$

$$\hat{x}_{k+1} = \arg \min_z \|y_{k+1} - h(z)\|^2 + \|z - \hat{x}_{k+1}^-\|_W^2 \quad (10b)$$

### B. Implementation Advantages and Practicalities

Note that in the nonlinear case, the observer (10) has the exact same interpretation as in the linear case (2). Namely, the goal is to invert the measurement function  $h$  in a least squares sense using the apriori prediction based on the system dynamics as a regularizing factor to render the problem well posed. There is no need to linearize or compute Jacobians of  $h$  as in the EKF. However, the cost of this is solving a minimization problem each iteration of the filter.

Solving the minimization problem (10b) is practical in two different scenarios. One scenario is that  $h$  has a specific analytic structure which admits computable closed form solutions for the minimizers. We will explore such an example in the following section for the case of a quadratic measurement equation.

The second practical scenario occurs when the structure of  $h$  admits easily computable expressions for minimizing values of most of the state vector as a function of only a few key states. In such cases, even when no simple closed form expressions are available for these key remaining states, numerical search techniques may be used to carry out the minimization in a reasonable amount of time. We will follow the quadratic example with an example of this second scenario for a infinite dimensiona visual tracking application in computer vision.

## IV. A QUADRATIC EXAMPLE

To illustrate the above ideas we consider in this section a simple scalar linear stochastic dynamical system with a noisy energy estimator. Let

$$x_{k+1} = f x_k + g u_k + \gamma w_k \quad (11)$$

where  $f, g$  and  $\gamma$  are constant.  $u_k$  is the deterministic system input, and  $w_k$  a standard stationary gaussian white noise sequence. Standard means: zero mean and unit variance. Let the output be a noisy estimate of the signal power, i.e.

$$y_k = x_k^2 + v_k. \quad (12)$$

where  $v_k$  is a white noise sequence, independent of the driving noise, and with zero mean and variance  $r$ .

### A. Deterministic observability

In the noiseless case, and with zero input, the system response is  $x_k = f^k x_0$  giving also  $y_k = f^{2k} x_0^2$ . Obviously, the initial state cannot be uniquely determined. Its magnitude can be retrieved, but not its sign. However if an input can be applied, the ambiguity disappears. Indeed, from  $y_0 = x_0^2$  and  $y_1 = (f x_0 + g u_0)^2$ , we get

$$x_0 = \frac{y_1 - f^2 y_0 - g^2 u_0^2}{2 f g u_0}$$

at the expense of an input and an additional time step.

### B. Filtered estimate via optimization

Consider now the stochastic case. Assuming the filtered estimate (using all data up to step  $k-1$  is  $\hat{x}_{k-1}$ , the prediction of the next state, given the deterministic input is

$$\bar{x}_k = f \hat{x}_{k-1} + g u_{k-1}.$$

The new measurement update,  $\hat{x}_k$ , incorporating the measurement  $y_k$ , is obtained by minimizing

$$J_k = (y_k - z^2)^2 + W_k (\bar{x}_k - z)^2. \quad (13)$$

Here  $W - k$  is a weight to be determined later. We find a cubic equation as a necessary condition for the optimum

$$z^3 - \left(y - \frac{W}{2}\right) z - \frac{W \bar{x}}{2} = 0. \quad (14)$$

The subscript  $k$  was suppressed for notational simplicity.

It is evident that if  $y < W/2$ , the cubic can only have one real root. If  $y > W/2$ , there are three real roots if in addition

$$|\hat{x}| < \frac{1}{3\sqrt{3W}} \left(y - \frac{W}{2}\right)^{3/2}.$$

In this case ambiguity remains as to which of the three roots minimizes  $J$ . However, since  $J$  is quartic in  $z$  with positive coefficient of  $z^4$ , the middle solution may always be excluded, as it corresponds to a local maximum in  $J$ . For the special form of the cubic, the local minima are always separated by zero. The global minimum of  $J$  is found by evaluation and comparison. However note that a shortcut is possible in this computation. Indeed

$$\begin{aligned} J &= (y - z^2)^2 + W(\bar{x} - z)^2 \\ &= y^2 + W\bar{x}^2 - \left(y - \frac{W}{2}\right) z^2 - \frac{3W\bar{x}}{2} z, \end{aligned}$$

by using (14). It follows that for given  $y$  and  $\bar{x}$  and  $W$ , the minimizing root  $z$  is the one that maximizes  $\left(\left(y - \frac{W}{2}\right) z + \frac{3W\bar{x}}{2}\right) z$ , i.e. the root closest to the value

$$z_0 = \frac{3\bar{x}}{1 - \frac{2y}{W}}.$$

Clearly, this means that if  $z_0 < 0$ , we take the leftmost root of (14), and the rightmost if  $z_0 > 0$ . For  $W = 1$ , Figure 1 shows a plot of the optimal estimate,  $z_{opt}$  for various values of  $y - 1/2$  as function of  $\bar{x}$ . Note the middle (red) line, for  $y = 1/2$ , corresponding to the transition where the criterion (14) has only a single real root. The graphs correspond to

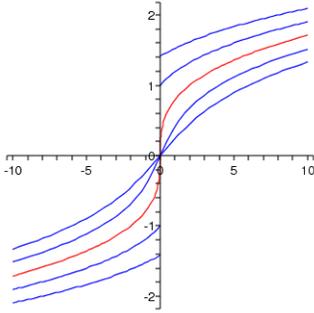


Fig. 1. Measurement update for  $m = 1$

$y - 1/2$  from  $-2$  to  $2$  in steps of  $1$ . For  $y > 1/2$  the optimal estimate is discontinuous at  $\bar{x} = 0$ .

It follows from the state equation that

$$x_k - \bar{x}_k = f(x_{k-1} - \hat{x}_{k-1}) + \gamma w_{k-1}.$$

Hence, with  $P_{k-1}^+$  is the variance of the *filtered* estimate at time  $k - 1$ , and  $P_k^-$  the variance of the *predicted* estimate at time  $k$ , we get the *time update* by ‘squaring up’

$$P_k^- = f^2 P_{k-1}^+ + \gamma^2.$$

The measurement updated variance equation is a lot more complicated considering the nonlinearities.

### C. Extended Kalman filter

This follows the exposition in [1, p.374]. The filter is ‘designed’ as a generalization of the linear case, with  $Hx$  replaced by  $\frac{\partial h}{\partial x}(\hat{x})$ . Since, the resulting update formulas for  $P^+$  and  $P^-$  are now *not* exact covariances, we use  $M$  for  $P^-$  and  $P$  for  $P^+$  in the recursions instead.

$$\begin{aligned} \hat{x}_k &= \bar{x}_k + K_k(y_k - \bar{x}_k^2), & \bar{x}_{k+1} &= f\hat{x}_k + gu_k, \\ K_k &= \frac{2}{r} P_k \bar{x}_k, & P_k &= M_k - \frac{4\bar{x}_k^2 M_k^2}{4\bar{x}_k^2 M_k + r}, \\ M_{k+1} &= f^2 P_k + \gamma^2. \end{aligned}$$

### D. Choice of weight

The weight  $W$  for the regularized dynamic inversion follows from the optimality condition as explained in section II.B, for which the recursions for  $P$  and  $M$  in the extended Kalman are instrumental. Summarizing, the proposed nonlinear filter for the power measurement system is given by

$$\begin{aligned} \bar{x}_k &= f\hat{x}_{k-1} + gu_{k-1}, & \hat{x}_k &= \operatorname{argmin}_z J_k(z; y_k, \bar{x}_k, W_k), \\ P_k &= M_k - \frac{4\bar{x}_k^2 M_k^2}{4\bar{x}_k^2 M_k + r}, & M_{k+1} &= f^2 P_k + \gamma^2, \\ W_k &= M_k^{-1} \end{aligned}$$

where  $\operatorname{argmin} J_k(z; \cdot)$  requires the solution of a cubic (explicit formulas exist) as described above.

### E. Simulation results

Preliminary results have indicated that for certain values of the parameters, substantial improvement is obtained over the extended Kalman filter solutions. In Figure 2a, the performance of the two estimators for the system with

$x_{k+1} = 0.9x_k + 1 + \gamma w_k$  and  $y_k = x_k^2 + \sqrt{r}v_k$ , and initial condition equal to the (noiseless) steady state value  $x_0 = 10$  is compared. The sequences  $w_k$  and  $v_k$  are standard white noise (zero mean and unit variance). The initial uncertainty  $M_0$  was taken to be  $100$  in both cases. The graphs show the (empirically estimated) filtered error covariance  $\mathbf{E}(x - \hat{x})^2$  for the extended Kalman filter and the regularized dynamic inversion in function of  $\gamma$ , ranging from  $0$  (no driving noise) to  $2$  and for fixed  $r$ . The different curves in Figure 2 correspond to  $r$  from  $0.2$  to  $2$  in steps of  $0.1$ . We observe that for small values of  $\gamma$  (up to about  $0.5$ ), the EKF and the regularized dynamic inversion have comparable performance. The sharply rising curves for  $\gamma > 0.5$  are the ones for the EKF. The two curves with the larger initial covariance correspond to  $r = 10$ , while the lower ones have  $r = 0.2$ . For much larger values of  $\gamma$  the nonlinear filter substantially outperforms the extended Kalman filter. Besides the predicted state error covariance, other measures of performance may be looked at. We computed the residuals,  $y - h(\hat{x})$ , and found that both filters have a bias. The empirical average for this bias, denoted  $\langle y - h(\hat{x}) \rangle$  under the same steady state conditions as discussed above is displayed in Figure 2b. For the nonlinear filter, the bias is remarkably insensitive to process and measurement noise covariance, and remains in fact close to zero. The EKF shows a bias that is quite sensitive to the driving noise covariance, but not the observation noise covariance. In addition, Figure 2c gives the empirical RMS value of the fluctuation around the bias, of these residuals. We discover that these fluctuations increase with increasing measurement noise, but for the EKF, they increase also sharply with increasing driving noise once  $\gamma$  exceeds  $1$ . On the other hand, for the nonlinear filter, the RMS values are quite insensitive to the driving noise.

It was also found by computing the power spectrum of the residuals  $y_k - \hat{x}_k^2$  that these residuals are ‘closer’ to a white noise sequence for the nonlinear filter than for the EKF. This is indicative of near optimality as would be obtained with the exact least squares filter [5]. We experienced one problem shared by both filters: If the filters are started near the *zero state*, and the (deterministic) input is zero, then neither filter may be able to discern between the state  $x$  or its mirror image  $-x$ , as both are consistent with the quadratic observation and time update. In fact such a dichotomy arises each time the state passes close to zero. This is due to the non-observability in the absence of an input. Note that in this case the residuals  $y - h(\hat{x})$  are close to zero, whereas the estimation errors  $\|x - \hat{x}\|$  are not. There is one exception: if the nonlinear filter for the autonomous system starts exactly in the zero state, it can never leave this state, whereas the EKF is able to track ‘‘half of the time’’ (the remaining half producing the mirror state). It is debatable which is the better.

## V. AN APPLICATION TO VISUAL TRACKING

We now apply the proposed observer to an infinite dimensional problem for which no closed form solution to the minimization problem exists but where minimizing values of almost all of the state vector can be written as a function of

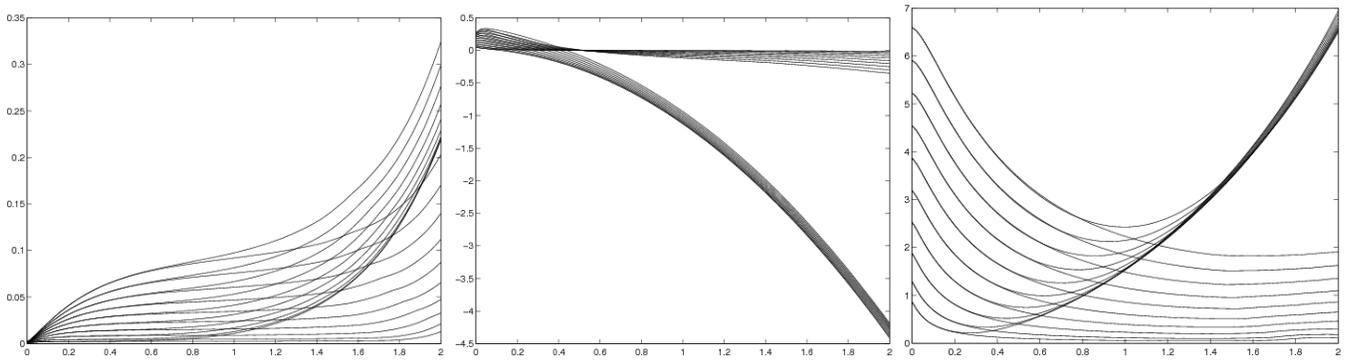


Fig. 2. (a) Filtered state - error covariance, (b) Bias in residuals, (c) RMS value of fluctuations in residuals about the bias

just a few key states. We proceed to locate the minimum numerically by searching over these key states. In particular, we consider a computer vision application of tracking a moving object within a video sequence.

#### A. Plant Description

1) *Foreground-Background model*: We consider a simplified model of a video sequence  $I$  which we treat as a time indexed sequence of images (with infinite spatial resolution for the moment), modelled as real-valued functions over the plane  $I_k : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ . Note that the image  $I_k$  at time  $k$  represents the plant output, which is infinite dimensional as it associates to each  $x \in \Omega$  an intensity  $I_k(x)$ . As such,  $x$  may be interpreted as an index into the infinite dimensional vector  $I_k$ , and no longer carries the usual interpretation of being the state vector. To further discourage the misinterpretation of  $x$  as the state vector, we have chosen to use the symbol  $I$  for the plant output (the image sequence) rather than the usual symbol  $y$  which will remain unused in this section.

We will utilize an infinite dimensional state vector as well in our simplified video plant model. However, we will divide the state into two infinite dimensional pieces and one finite dimensional piece. We assume that the video sequence describes a scene with an unchanging background and a single rigidly moving foreground object which occludes portions of the static background as it moves. We also assume that the motion of the foreground object can be described by a simple time varying 2D displacement (translation)  $d_k \in \mathbb{R}^2$  and that modulo this displacement and noise effects, the appearance of the foreground object remains constant over time (as do unoccluded portions of the background). The displacement  $d_k$  will constitute the finite dimensional part of our state vector. The background and foreground appearances (what we actually see in each image), however, will be infinite dimensional, both modelled as real-valued functions which we will call the background radiance  $b_k : \Omega \rightarrow \mathbb{R}$  and the foreground radiance  $f_k : \Sigma \rightarrow \mathbb{R}$  respectively. Notice that the background radiance  $b_k$  is defined over the same domain  $\Omega$  as the output image  $I_k$  based on our assumption of no

motion of the scene background<sup>1</sup>. The foreground domain  $\Sigma$  is separate, and we will assume that  $\Sigma + d_k \subset \Omega$  for all  $k$ .

2) *State equations*: The assumption of constant appearance of both the foreground and background yields trivial dynamics for their respective radiance functions. The displacement signal  $d_k$ , instead, would exhibit scenario dependent dynamics. For now, we just assume constant velocity. We therefore obtain the following linear dynamical equations (note that there is no input  $u$  nor modelling noise  $w_k$ ).

$$\begin{aligned} b_{k+1} &= b_k, & \dot{d}_{k+1} &= \dot{d}_k \\ f_{k+1} &= f_k, & d_{k+1} &= d_k + \dot{d}_k \end{aligned} \quad (15)$$

3) *Output equation*: While the state equations are rather simple, the measurement equation

$$y_k = h(b_k, f_k, d_k) + v_k,$$

where

$$\begin{aligned} h(b, f, d)(x) &= \\ & f(x-d)\chi_\Sigma(x-d) + b(x)(1 - \chi_\Sigma(x-d)), \end{aligned} \quad (16)$$

exhibits two difficult nonlinearities. Both nonlinearities are in the displacement  $d$ . One is through the discontinuous characteristic function  $\chi_\Sigma : \mathbb{R}^2 \rightarrow \{0, 1\}$  which is defined as follows.

$$\chi_\Sigma(x) = \begin{cases} 1, & x \in \Sigma \\ 0 & x \notin \Sigma \end{cases}$$

The other nonlinearity in  $d$  is through composition with  $f$ , which could contain an even larger set of discontinuities than  $\chi_\Sigma$  since  $f$  is intended to represent the appearance of a foreground object which may be highly textured (consider a running zebra for example). As a result, it is rather useless to linearize  $h$  (which precludes the extended Kalman filter), and extremely difficult to approximate these nonlinearities by any sort of polynomial or other truncated expansion over common types of bases.

The interpretation of (16) is straight forward. The role of the characteristic function  $\chi_\Sigma$  is to model the occlusion of the

<sup>1</sup>To keep the presentation simple we have not delineated all of the additional assumptions necessary to render this model valid (e.g. no camera motion, constant illumination, Lambertian reflection, no shadow effects, etc.) as this is not intended to be an article on computer vision but rather a practical application to illustrate the performance of the proposed observer.

background by the moving foreground object. Thus, a point  $x \in \Omega$  in the image  $I_k$  will exhibit an intensity taken from the foreground radiance function  $f_k$  if, after adjusting for the foreground displacement  $d_k$ , the point is determined to fall within the foreground domain  $\Sigma$ . Otherwise, it will exhibit an intensity taken from the background radiance function  $b_k$ .

**B. Observer equations**

Running the proposed observer for this video plant model, despite its dimensionality and nonlinearity, is computationally efficient and simple. The prediction step (10a) is carried out in the usual way by using equations (15) to obtain apriori estimates  $\hat{b}_{k+1}^-$ ,  $\hat{f}_{k+1}^-$ , and  $\hat{d}_{k+1}^-$  for the background/foreground radiances and for the displacement respectively.

The correction step (10b) involves minimizing the following energy for all choices of  $z_b$ ,  $z_f$ , and  $z_d$  (representing  $b_{k+1}$ ,  $f_{k+1}$ , and  $d_{k+1}$  respectively) given the output image  $I_{k+1}$  and the apriori estimates.

$$J(z_b, z_f, z_d, \hat{b}_{k+1}^-, \hat{f}_{k+1}^-, \hat{d}_{k+1}^-, I_{k+1}) = \left\| h(z_b, z_f, z_d) - I_{k+1} \right\|_{\mathcal{L}^2\{\Omega\}}^2 + W_b \left\| z_b - \hat{b}_{k+1}^- \right\|_{\mathcal{L}^2\{\Omega\}}^2 + W_f \left\| z_f - \hat{f}_{k+1}^- \right\|_{\mathcal{L}^2\{\Sigma\}}^2 + W_d \left\| z_d - \hat{d}_{k+1}^- \right\|^2 \tag{17}$$

The minimization of (17) is rendered computationally simpler by the fact that globally optimal radiances  $z_b^*$  and  $z_f^*$  (which comprise the two infinite dimensional parts of the state vector) may be computed directly as a function of the displacement  $z_d$ . This is due to the fact that  $h$ , as given in (16), is linear in the radiances  $b$  and in  $f$ . As such, we need only minimize the following cost  $J^*$  which is already optimized in  $z_b$  and  $z_f$  and depends only upon  $z_d$ :

$$J^*(z_d, \hat{b}_{k+1}^-, \hat{f}_{k+1}^-, \hat{d}_{k+1}^-, I_{k+1}) = J(z_b^*, z_f^*, z_d, \hat{b}_{k+1}^-, \hat{f}_{k+1}^-, \hat{d}_{k+1}^-, I_{k+1}) \tag{18}$$

where the optimal radiances  $z_b^*$  and  $z_f^*$  for a given  $z_d$  are

$$z_b^*(x) = \begin{cases} \frac{W_b \hat{b}_{k+1}^-(x) + I_{k+1}(x)}{W_b + 1}, & x - z_d \notin \Sigma \\ \hat{b}_{k+1}^-(x), & x - z_d \in \Sigma \end{cases} \tag{19a}$$

$$z_f^*(x) = \frac{W_f \hat{f}_{k+1}^-(x) + I_{k+1}(x + z_d)}{W_f + 1}, \tag{19b}$$

Now that the minimization step has been reduced to two dimensions, a search is computationally feasible. In our experiments, we perform this search by quantizing the displacement state space for  $d_{k+1}$  to correspond with the resolution of the video image  $I_{k+1}$  such that only translations in exact pixel units are considered (thereby avoiding the need to interpolate when evaluating  $I_{k+1}(x + d)$ ). The search is limited to a rectangular window centered around the apriori estimate. Since the global minimizer is expected to lie somewhere in the vicinity of the predicted estimate, this

procedure typically locates the global minimum  $z_d^*$  of (18) so long as the window size is not too small. The final estimates are then obtained by setting  $\hat{d}_{k+1} = z_d^*$ ,  $\hat{b}_{k+1} = z_b^*$ , and  $\hat{f}_{k+1} = z_f^*$ .

**C. Experiments**

In figure 3 we show the result of our nonlinear observer in tracking a person walking in a parking lot with very high additive image noise plus an additional black bar of missing image information (simulating a partial sensor failure). Note that the extremely high noise levels and the highly discontinuous nature of the vertical bar disturbance makes the Extended Kalman Filter unsuitable for this application. However, the nonlinear regularized dynamic inversion observer tracks very successfully as demonstrated. The top row shows the image sequence without the noise, the middle row shows the measured noisy output along with the displacement estimate  $\hat{d}_k$  (the moving white box), and the third row shows the estimated output based on the foreground and background estimates  $\hat{f}_k$  and  $\hat{b}_k$  as well. In this experiment we just chose fixed weights  $W_f=10$ ,  $W_b=15$ , and  $W_d=1$ .

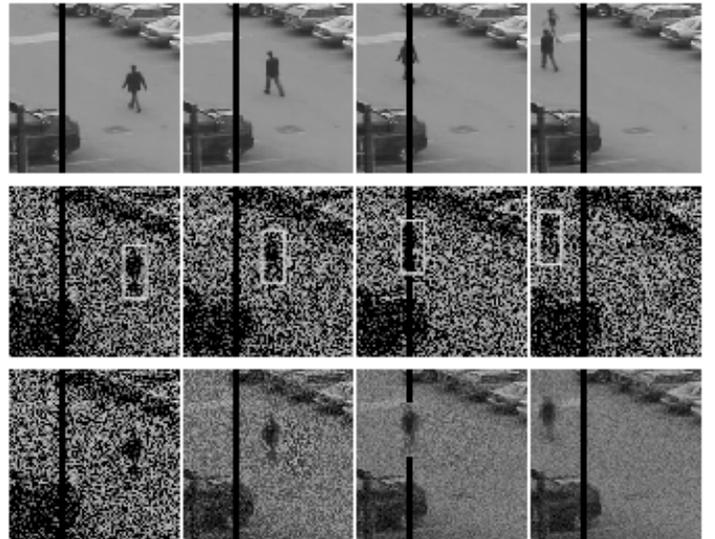


Fig. 3. Application of the nonlinear observer for visual tracking.

**REFERENCES**

- [1] A. Bryson and Y.C. Ho, *Applied Optimal Control*, Ginn and Co., Waltham, MA; 1969.
- [2] E. Verriest, "Linear Filters for Linear Systems with Multiplicative Noise and Nonlinear Filters for Linear Systems with Non-Gaussian Additive Noise," *Proc. Am. Control Conf.*, June 1985, pp. 182–184.
- [3] D. Luenberger, "An introduction to observers," *IEEE Transactions on Automatic Control*, Vol. 16, No. 6, Dec 1971, pp. 596–602
- [4] R. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, Mar. 1960, pp. 33–45.
- [5] P.A. Frost and T. Kailath, "An Innovations Approach to Least-Squares Estimation-Part I11 : Nonlinear Estimation in White Gaussian Noise," *IEEE Transactions on Automatic Control*, Vol. AC-16, No. 3, June 1971, pp. 217-226.
- [6] T. Kailath, A. Sayed, and B. Hassibi, *Linear Estimation*, Prentice Hall, New Jersey, 2000.
- [7] E.I. Verriest and W.S. Gray, "Geometric Approach to the Minimum Sensitivity Design Problem," *SIAM Journal on Control and Optimization*, vol. 33, no. 3, pp. 863-881, May 1995.