

# Filter Based Estimates of Depth

SJ Maybank

Hirst Research Centre

East Lane, Wembley, Middlesex HA9 7PP, UK

---

We compare three filters for estimating the depth and velocity of a translating point, namely, (i) the optimal filter; (ii) a second order approximation to the optimal filter; and (iii) the extended Kalman filter. The filter input is a sequence of measurements of the image of the point under polar projection. The measurements are subject to white Gaussian noise.

The optimal filter requires an excessive amount of computation. The second order approximation is close to the optimal filter, especially if the noise level is low; and it requires far less computation. The extended Kalman filter also requires less computation than the optimal filter, but it usually performs very badly. In particular, it seriously underestimates covariance.

---

We compare the performance of three filters applied to a problem in computer vision, the estimation of the position and velocity of a uniformly translating point, using image measurements corrupted by Gaussian noise. We choose this very simple application in order to make the comparison between the three filters as clear as possible. The filters are (i) the optimal filter [1]; (ii) a second order approximation to the optimal filter; and (iii) the extended Kalman filter (EKF) [1,2,3]. The significance of the optimal filter is that it provides an absolute standard against which the performance of any approximating filter can be judged.

Our main result is that the second order approximation is closer to the optimal filter than the EKF. The result is significant because the EKF is currently a standard tool in computer vision. The second order approximation gives a simple way of assessing how far the performance of the EKF is degraded from the optimal.

The use of filters for estimating the depth and velocity of moving objects is described in [4,5]. A method for improving on the EKF in a different context is described in [6], where it is used to fit an ellipse to a set of points. It is shown that an adjustment to the usual formulation of the EKF leads to better results, at only a small increase in computational cost.

In our application of filtering to the estimation of depth and velocity we assume that we have a sequence of images of a point moving in the  $x, y$  plane with constant translational velocity  $(v, 1)^T$ , as illustrated in Figure 1. We scale the velocity such that the  $y$  component is equal to one, in order to reduce the number of parameters. This scaling does not cause difficulty provided the true  $y$  component of the velocity is not excessively small.

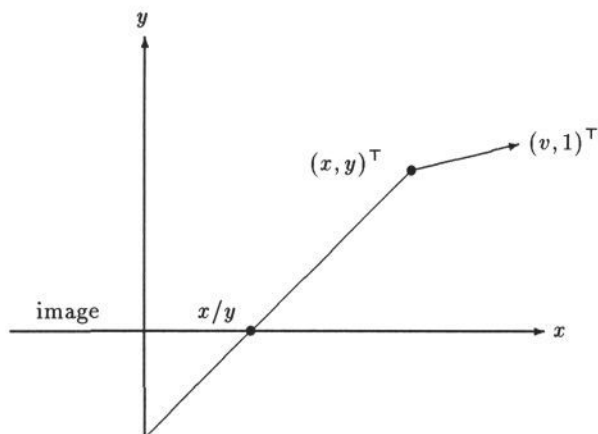


Figure 1: A translating point

We describe the state of the moving point at time  $i$  by the vector,  $\mathbf{x}_i = (x_i, y_i, v_i)^T$ . The state  $\mathbf{x}_{i+1}$  is given in terms of  $\mathbf{x}_i$  by

$$\mathbf{x}_{i+1} = (x_i + v_i, y_i + 1, v_i)^T \quad (1)$$

We write (1) in the form  $\mathbf{x}_{i+1} = G\mathbf{x}_i$ . The transformation  $G$  is affine, deterministic, and independent of time.

We take measurements  $\mu_i$  of the image of the point at discrete times  $i = 1, 2, 3, \dots$ . The  $\mu_i$  are given by

$$\mu_i = x_i/y_i + w_i$$

where  $w_i$  is zero-mean white Gaussian noise with a known standard deviation  $\rho$ . Any information that we may have about the motion of the point additional to the information obtained from the measurements is summarised in a prior probability density function (pdf)  $p(\mathbf{x})$ . The pdf  $p(\mathbf{x})$  describes our knowledge of the state at time 1, prior to the first measurement. We denote by  $I^k$  the information about the state available at time  $k$ . Thus  $I^k$  consists of the prior density  $p(\mathbf{x})$  and the measurements  $\mu_i, 1 \leq i \leq k$ .

Our entire knowledge about the state  $\mathbf{x}_k$  is contained in the conditional pdf  $p(\mathbf{x}_k|I^k)$ . The optimal filter is based on an equation which expresses  $p(\mathbf{x}_k|I^k)$  in terms of  $p(\mathbf{x}_{k-1}|I^{k-1})$ . Let  $p(\mu_k|\mathbf{x}_k)$  be the pdf of the measurement  $\mu_k$  conditional on  $\mathbf{x}_k$ , and let  $p(\mathbf{x}_k|\mathbf{x}_{k-1})$  be the pdf of  $\mathbf{x}_k$  conditional on  $\mathbf{x}_{k-1}$ . Thus  $p(\mu_k|\mathbf{x}_k)$  describes the measurement process, and  $p(\mathbf{x}_k|\mathbf{x}_{k-1})$  describes the time evolution of the state. The equation

for the optimal filter is [1]

$$p(\mathbf{x}_k|I^k) = c p(\mu_k|\mathbf{x}_k) \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|I^{k-1}) d\mathbf{x}_{k-1} \quad (2)$$

where  $c$  is a normalising constant. The integral on the right-hand side of (2) serves to propagate from time  $k - 1$  to time  $k$  the conditional pdf that summarises our knowledge of the state. The coefficient  $p(\mu_k|\mathbf{x}_k)$  of the integral describes the effect of incorporating the measurement  $\mu_k$ . In our application to tracking  $\mathbf{x}_k$  is a deterministic function (1) of  $\mathbf{x}_{k-1}$ . It follows that

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}) = \delta(\mathbf{x}_k - G\mathbf{x}_{k-1}) \quad (3)$$

where  $\delta$  is the Dirac delta function. The pdf  $p(\mu_k|\mathbf{x}_k)$  is Gaussian, of the form

$$p(\mu_k|\mathbf{x}_k) = \frac{1}{\sqrt{2\pi\rho}} \exp\left[-\frac{1}{2\rho^2}\left(\frac{x_k}{y_k} - \mu_k\right)\right] \quad (4)$$

The optimal filter is rarely used because it requires an excessive amount of computation. The major exception to this rule is the Kalman filter [2], which is the optimal filter for linear systems subject to white Gaussian noise. In the non-linear case filters have been developed which approximate to the optimal filter, but which require less computation [2,3]. The best known of these approximating filters is the EKF.

## THE FILTERS

We regard filtering as a way of obtaining approximations to the optimal pdfs  $p(\mathbf{x}_k|I^k)$  as the time  $k$  increases. For example, the EKF calculates the mean and covariance of a Gaussian approximation to  $p(\mathbf{x}_k|I^k)$ . Let  $p_o(\mathbf{x})$ ,  $p_s(\mathbf{x})$ ,  $p_e(\mathbf{x})$  be the pdfs produced respectively by the optimal filter, the second order approximation, and the EKF, after *three* measurements have been taken. We choose three measurements because this is the least number required to specify the system state  $(x, y, v)^\top$  exactly in the noise free case.

We obtain expressions for  $p_o(\mathbf{x})$ ,  $p_s(\mathbf{x})$  and  $p_e(\mathbf{x})$ . We define the function  $f(\mathbf{x})$  by

$$f(\mathbf{x}) = \left(\frac{x}{y} - \mu_3\right)^2 + \left(\frac{x-v}{y-1} - \mu_2\right)^2 + \left(\frac{x-2v}{y-2} - \mu_1\right)^2 \quad (5)$$

Equations (3), (4) and (2) yield the following expression for  $p_o(\mathbf{x})$ .

$$p_o(\mathbf{x}) = c \exp\left[-\frac{1}{2\rho^2}f(\mathbf{x})\right] p(G^{-2}\mathbf{x}) \quad (6)$$

In (6),  $c$  is a normalising constant. We assume that  $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}, \mathbf{m}, P)$ , where  $\mathcal{N}(\mathbf{x}, \mathbf{m}, P)$  is the Gaussian pdf with mean  $\mathbf{m}$  and covariance  $P$ .

We define  $p_s(\mathbf{x})$ . Let  $\mathbf{c}$  be the vector such that  $f(\mathbf{c}) = 0$ . The coordinates of  $\mathbf{c}$  are

$$\begin{aligned} c_1 &= (2\mu_3(\mu_1 - \mu_2)/(\mu_3 + \mu_1 - 2\mu_2)) \\ c_2 &= 2(\mu_1 - \mu_2)/(\mu_3 + \mu_1 - 2\mu_2) \\ c_3 &= (2\mu_1\mu_3 - \mu_2\mu_3 - \mu_1\mu_2)/(\mu_3 + \mu_1 - 2\mu_2) \end{aligned} \quad (7)$$

If the measurements are noise free then  $\mathbf{c}$  is the state at time  $k = 3$ . In other words, the point being tracked is at the position  $(c_1, c_2)^\top$  and moving with velocity  $(c_3, 1)^\top$ . In the more realistic case of non-zero noise we expect  $p_o(\mathbf{x})$  to be concentrated about  $\mathbf{c}$ .

The point  $\mathbf{c}$  is a global minimum for  $f(\mathbf{x})$ , thus  $\nabla f(\mathbf{c}) = 0$ . Let  $J$  be the Hessian matrix of second order partial derivatives of  $f(\mathbf{x})$  evaluated at  $\mathbf{c}$ . On taking the Taylor expansion of  $f(\mathbf{x})$  about  $\mathbf{c}$  we obtain

$$f(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{m})^\top J(\mathbf{x} - \mathbf{m}) + O(\|\mathbf{x} - \mathbf{c}\|^3) \quad (8)$$

We obtain  $p_s(\mathbf{x})$  by using the quadratic terms on the right-hand side of (8) to approximate to  $f(\mathbf{x})$  in (5). This yields

$$p_s(\mathbf{x}) = \mathcal{N}(\mathbf{x}, \mathbf{c}, \rho^2 J^{-1}) p(G^{-2}\mathbf{x})$$

The pdf  $p_s(\mathbf{x})$  is a product of Gaussians, therefore it is also Gaussian.

The pdf  $p_e(\mathbf{x})$  obtained by the EKF is a Gaussian approximation to  $p_o(\mathbf{x})$ . The mean and covariance of  $p_e(\mathbf{x})$  are obtained recursively as follows. We define the matrix  $F$  and the vector  $\mathbf{e}$  by

$$F = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{e} = (0, 1, 0)^\top$$

Thus equation (1), describing the time evolution of the state, is equivalent to

$$\mathbf{x}_{i+1} = F\mathbf{x}_i + \mathbf{e} \quad (9)$$

Let  $\mathbf{m}_i$ ,  $P_i$  be the mean and covariance obtained by the EKF at time  $i$  after incorporating the measurement  $\mu_i$ . The predicted mean and covariance  $\mathbf{n}$ ,  $C$  at time  $i + 1$ , but prior to the measurement  $\mu_{i+1}$ , are given by

$$\mathbf{n} = F\mathbf{m}_i + \mathbf{e} \quad C = FP_iF^\top \quad (10)$$

We define the observation vector  $\mathbf{h}_{i+1}$  by

$$\mathbf{h}_{i+1} = (1/n_2, -n_1/n_2^2, 0)^\top \quad (11)$$

The updated mean and covariance  $\mathbf{m}_{i+1}$ ,  $P_{i+1}$  are obtained from  $\mathbf{n}$ ,  $C$ ,  $\mathbf{h}_{i+1}$  by the equations [3]

$$\begin{aligned} P_{i+1} &= C - (\rho^2 + \mathbf{h}_{i+1}^\top C \mathbf{h}_{i+1})^{-1} C \mathbf{h}_{i+1} \otimes \mathbf{h}_{i+1} C \\ \mathbf{m}_{i+1} &= \mathbf{n} + \left[ \frac{\mu_{i+1} - n_1/n_2}{\rho^2 + \mathbf{h}_{i+1}^\top C \mathbf{h}_{i+1}} \right] C \mathbf{h}_{i+1} \end{aligned} \quad (12)$$

We define the pdf  $p_e(\mathbf{x})$  by

$$p_e(\mathbf{x}) = \mathcal{N}(\mathbf{x}, \mathbf{m}_3, P_3)$$

We initialise the EKF prior to the first measurement by setting  $\mathbf{n} = \mathbf{m}$ ,  $C = P$ , where  $\mathbf{m}$ ,  $P$  are the mean and covariance of the prior pdf  $p(\mathbf{x})$ . To obtain  $\mathbf{m}_3$  and  $P_3$  we incorporate the measurement  $\mu_1$  using (12), and we then carry out the prediction step (10) and the update step (12) two times.

## MARGINAL DENSITIES

We use marginal densities to compare the effectiveness of the three pdfs  $p_o(\mathbf{x})$ ,  $p_s(\mathbf{x})$ ,  $p_e(\mathbf{x})$ . Let the three marginal densities on the depth  $y$  be  $p_{oy}(y)$ ,  $p_{sy}(y)$ ,  $p_{ey}(y)$  respectively, and let the three marginal densities on the velocity component  $v$  be  $p_{ov}(v)$ ,  $p_{sv}(v)$ ,  $p_{ev}(v)$ . The marginal densities are obtained by integration. For example, we have

$$p_{oy}(y) = \int \int p_o(\mathbf{x}) dx dv \quad (13)$$

In the case of the Gaussian density,  $\mathcal{N}(\mathbf{x}, \mathbf{m}, P)$ , the marginal densities are easy to obtain. For example, the marginal density of the component  $y$  of  $\mathbf{x}$  is Gaussian with mean  $m_2$  and variance  $P_{22}$ . It follows that  $p_{sy}(y)$ ,  $p_{ey}(y)$  and  $p_{sv}(v)$ ,  $p_{ev}(v)$  are all Gaussian.

The optimal marginal density  $p_{oy}(y)$  is more difficult to obtain. We write  $p_o(\mathbf{x})$  in the form

$$p_o(\mathbf{x}) = c \exp\left[-\frac{1}{2}(Ax^2 + Bxv + Cv^2 + Dx + Ev)\right] \quad (14)$$

where  $A$ ,  $B$ ,  $C$ ,  $D$ ,  $E$  are functions of  $y$  and of the measurements  $\mu_1, \mu_2, \mu_3$ . We define the matrix  $M$  and the vector  $\mathbf{l}$  by

$$M = \begin{pmatrix} A & B/2 \\ B/2 & C \end{pmatrix} \quad \mathbf{l} = \begin{pmatrix} D \\ E \end{pmatrix} \quad (15)$$

We define the vector  $\mathbf{u}$  by  $\mathbf{u} = (x, v)^\top$ . It follows from (14) and (15) that

$$\begin{aligned} p_{oy}(y) &= c \int \int p_o(\mathbf{x}) dx dv \\ &= c \int \int \exp\left[-\frac{1}{2}(\mathbf{u}^\top M \mathbf{u} + \mathbf{l} \cdot \mathbf{u})\right] dx dv \\ &= \frac{c' \exp(\mathbf{l}^\top M^{-1} \mathbf{l}/8)}{\sqrt{\det(M)}} \end{aligned} \quad (16)$$

where  $c'$  is a normalising constant independent of  $y$ .

## EXPERIMENTAL RESULTS

We use the Mathematica computer algebra program [7] to draw graphs of the marginal densities for 'typical' choices of the noise level  $\rho$  and of the measurements  $\mu_1,$

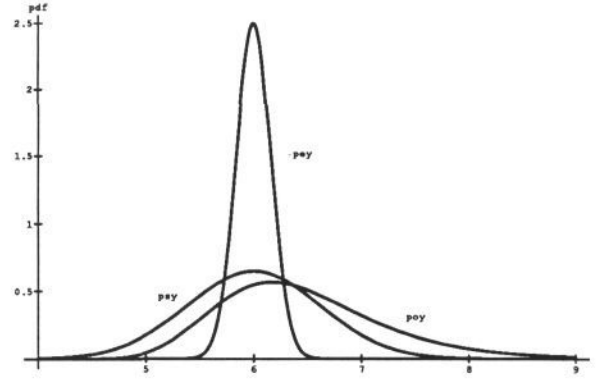


Figure 2: Marginal densities for  $y$  ( $\rho = 1/200$ )

$\mu_2, \mu_3$ . For most of the experiments we use  $\rho = 1/200$  rad. In an image of size  $512 \times 512$  pixels with a  $30^\circ$  field of view, this value of  $\rho$  corresponds to an error in locating a feature point of

$$\frac{1}{200} \times \frac{180}{\pi} \times \frac{512}{30} \sim 5 \text{ pixels}$$

The prior pdf,  $p(G^{-2}\mathbf{x})$ , is chosen to be Gaussian with mean  $\mathbf{m}$  given in terms of the vector  $\mathbf{c}$  of (7) by

$$\mathbf{m} = (c_1 - 2c_3, c_2 - 2, 0)^\top$$

The first two components of  $\mathbf{m}$  coincide with the first two components of the state vector  $G^{-2}\mathbf{c}$ . The third component of  $\mathbf{m}$  is set equal to zero. We choose the prior covariance to be  $1000 I$ , where  $I$  is the  $3 \times 3$  identity matrix. The covariance of the prior pdf is large in order to reduce the effect of the prior pdf on  $p_o(\mathbf{x})$ ,  $p_s(\mathbf{x})$  and  $p_e(\mathbf{x})$ . This makes it easier to assess the extent to which  $p_s(\mathbf{x})$  and  $p_e(\mathbf{x})$  use the information available in the measurements.

We plot the pdfs  $p_{oy}(y)$ ,  $p_{sy}(y)$ ,  $p_{ey}(y)$  in Figure 2 for the triple of measurements

$$(\mu_1, \mu_2, \mu_3)^\top = (-3/10, 0, 1/5)^\top \quad (17)$$

The corresponding value of  $\mathbf{c}$  is  $\mathbf{c} = (6/5, 6, 6/5)^\top$ . We see from Figure 2 that  $p_{sy}(y)$  is fairly close to  $p_{oy}(y)$ , except that the graph of  $p_{sy}(y)$  is shifted slightly towards the lower values of  $y$ . The pdf  $p_{oy}(y)$ , is skewed with a bias towards the higher values of  $y$ . The graph of  $p_{ey}(y)$  is sharply peaked, but because the optimal filter is available for comparison, we know that  $p_{ey}(y)$  severely underestimates the covariance. The mean values and standard deviations of the three pdfs plotted in Figure 2 are

$$\begin{aligned} E(p_{oy}) &= 6.47 & \sigma_{oy} &= 0.80 \\ E(p_{sy}) &= 6.00 & \sigma_{sy} &= 0.62 \\ E(p_{ey}) &= 6.00 & \sigma_{ey} &= 0.16 \end{aligned}$$

We plot the pdfs  $p_{oy}(y)$ ,  $p_{sy}(y)$ ,  $p_{ey}(y)$  in Figure 3 for the triple of measurements

$$(\mu_1, \mu_2, \mu_3)^\top = (-1/4, 0, 1/5)^\top \quad (18)$$

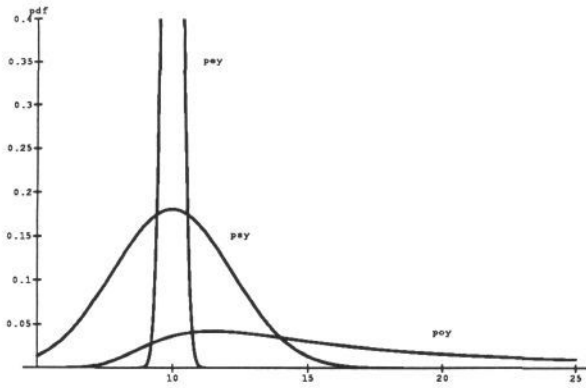


Figure 3: Marginal densities for  $y$ , ( $\rho = 1/200$ )

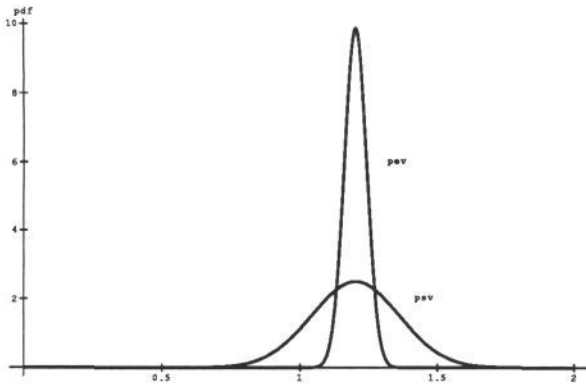


Figure 4: Marginal densities for  $v$  ( $\rho = 1/200$ )

In this case  $\mathbf{c} = (2, 10, 2)^\top$ . In this second example the depth  $c_2$  is increased from 6 to 10. The graphs of  $p_{oy}(y)$  and  $p_{ey}(y)$  are greatly flattened compared with Figure 2. The pdf  $p_{ey}(y)$  retains a sharp peak (truncated in the figure), but as we have already noted in Figure 2, this does not indicate that the EKF is accurate. Instead it shows that the EKF seriously underestimates the covariance. As in the first example,  $p_{sy}$  is shifted towards the lower values of  $y$ , and  $p_{oy}(y)$  is skewed, with a bias in favour of the higher values of  $y$ . The mean values and standard deviations of the three pdfs plotted in Figure 3 are

$$\begin{aligned} E(p_{oy}) &= 67.45 & \sigma_{oy} &= 67.45 \\ E(p_{sy}) &= 10.00 & \sigma_{sy} &= 2.20 \\ E(p_{ey}) &= 10.00 & \sigma_{ey} &= 0.28 \end{aligned}$$

It is clear from the graph of  $p_{oy}(y)$  in Figure 3 that a good estimate of  $y$  cannot be obtained. It is only possible to give a sharp lower bound on  $y$ .

In Figure 4 we plot the marginal densities  $p_{sv}(v)$ ,  $p_{ev}(v)$  for the triple of measurements (17). The optimal pdf,  $p_{ov}(v)$ , is not shown because of difficulties in calculating its values.

In Figure 5 we plot the marginal densities  $p_{oy}(y)$ ,

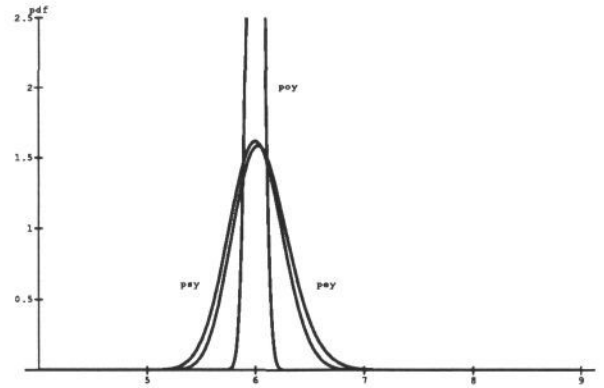


Figure 5: Marginal densities for  $y$  ( $\rho = 1/500$ )

$p_{sy}(y)$ ,  $p_{ey}(y)$  for the measurements (17), with the noise level  $\rho$  decreased to  $\rho = 1/500$ , which corresponds to an error of about 1 pixel in locating the image of the translating point (in a  $512 \times 512$  image with a  $30^\circ$  field of view). The densities  $p_{oy}(y)$  and  $p_{sy}(y)$  are much closer to each other than in Figure 2, and more sharply peaked about their mean values.

Experiments show that the performance of the EKF depends critically on the prior mean  $\mathbf{m}$ . We illustrate this overdependence of the EKF on  $\mathbf{m}$  in Figure 6. We set  $\rho = 1/200$ , and we employ the same measurements (17) used to produce Figure 2. To produce Figure 6 we set

$$\mathbf{m} = (c_1 - 2c_3, c_2, 0)^\top$$

The  $y$  coordinate of  $\mathbf{m}$  is increased two units over the  $y$  coordinate of the prior mean used in Figure 2. This shift of two units is transmitted by the EKF straight through to the estimate of depth. In contrast, the optimal filter and the second order filter are almost unaffected because the contributions to the pdfs from the measurements swamp the effect of the prior pdf.

There is one case in which the EKF does perform well. We show in the appendix that if the prior mean  $\mathbf{m}$  coincides with the state  $G^{-2}\mathbf{c}$  then  $p_e(\mathbf{x})$  exactly coincides with  $p_s(\mathbf{x})$ . If  $\mathbf{m}$  does not coincide with  $G^{-2}\mathbf{c}$  then the covariance estimates produced by the EKF are far too optimistic, as illustrated in Figures 2-6.

## CONCLUSION

Our results show that in the simple tracking problem we can find an approximation to the optimal filter that is computationally tractable, and that in general far outperforms the extended Kalman filter. The approximation is obtained by accumulating enough measurements to specify the system state exactly in the noise free case, and then taking a Gaussian approximation to the optimal pdf with the mean of the Gaussian located at the zero noise estimate of the system state.

The weaknesses of the EKF are that it is too strongly

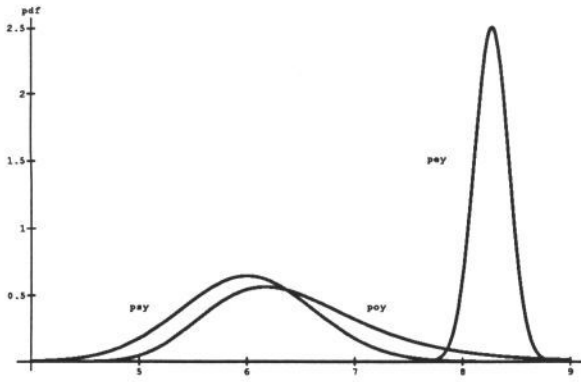


Figure 6: Effect of prior mean on the EKF ( $\rho = 1/200$ )

affected by the prior mean, and that it seriously underestimates covariances. We conjecture that alternative approximations to the optimal filter of the type described above are preferable to the extended Kalman filter in a wide range of applications, particularly when the noise level is low.

**Acknowledgements.** This work is funded by ESPRIT Project 2502 (Voilà). The author thanks Bernard Buxton and David Castelow for useful discussions.

## REFERENCES

1. Bar-Shalom, Y. & Fortmann, T.E. *Tracking and Data Association*. Mathematics in Science and Engineering Series, Vol. 179, Academic Press Inc., Orlando, Florida (1988).
2. Maybeck, P.S. *Stochastic Models, Estimation, and Control, Volume 1*. Mathematics in Science and Engineering Series, Vol. 141-1, Academic Press Inc., San Diego, CA, USA (1979).
3. Maybeck, P.S. *Stochastic Models, Estimation, and Control, Volume 2*. Mathematics in Science and Engineering Series, Vol. 141-2, Academic Press Inc., New York, NY, USA (1982).
4. Broida, T.J. & Chellappa, R. "Estimation of object motion parameters from noisy images". *IEEE Trans. Pattern Analysis and Machine Intelligence* Vol. 8 (1986) pp 90-99.
5. Matthies, L. & Kanade, T. "Kalman filter based algorithms for estimating depth from image sequences". *Int. J. Computer Vision* Vol. 3 (1989) pp 209-236.
6. Porrill, J. "Fitting ellipses and predicting confidence envelopes using a bias corrected Kalman filter." *Proc. Fifth Alvey Vision Conf., Reading, UK.* (1989) pp 175-180.
7. Wolfram, S. *Mathematica, A System for Doing Mathematics by Computer*. Addison-Wesley (1988).

## APPENDIX

We show that in one special case the EKF is identical to the second order approximation to the optimal filter.

**Theorem 1.** Let the prior mean  $\mathbf{m}$  be chosen such that  $\mathbf{m} = G^{-2}\mathbf{c}$ , where  $\mathbf{c}$  is defined by (7). Then  $p_s(\mathbf{x}) = p_e(\mathbf{x})$ .

**Proof.** It follows from the definition of  $\mathbf{c}$  and the hypothesis  $\mathbf{m} = G^{-2}\mathbf{c}$  that in each update by the EKF calculated from (12) the factor  $\mu_{i+1} - n_1/n_2$  on the right-hand side of the second equation of (12) is equal to zero. The three mean values  $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3$  produced by the EKF at times 1, 2, 3 respectively are thus given by

$$\mathbf{m}_1 = G^{-2}\mathbf{c} \quad \mathbf{m}_2 = G^{-1}\mathbf{c} \quad \mathbf{m}_3 = \mathbf{c} \quad (19)$$

It follows that the mean of  $p_s(\mathbf{x})$  is equal to the mean of  $p_e(\mathbf{x})$ . To prove the theorem it remains only to show that  $p_s(\mathbf{x})$  and  $p_e(\mathbf{x})$  have the same covariance.

In our construction of  $p_s(\mathbf{x})$  we take a quadratic approximation to the function  $f(\mathbf{x})$  of (5) about the point  $\mathbf{c}$ . We note that  $\mathbf{c} = \mathbf{m}_3 = \mathbf{n}$  where  $\mathbf{n}$  is the mean predicted by the EKF at time  $k = 3$  prior to the third measurement. The measurements  $\mu_1, \mu_2, \mu_3$  are given in terms of  $\mathbf{n}$  and  $G$  by the equations

$$\frac{n_1}{n_2} = \mu_3 \quad \frac{(G^{-1}\mathbf{n})_1}{(G^{-1}\mathbf{n})_2} = \mu_2 \quad \frac{(G^{-2}\mathbf{n})_1}{(G^{-2}\mathbf{n})_2} = \mu_1$$

We expand  $x/y - \mu_3$  about the predicted mean  $\mathbf{n}$  to obtain

$$\begin{aligned} \frac{x}{y} - \mu_3 &= \frac{x}{n_2 + y - n_2} - \mu_3 \\ &= \frac{x}{n_2} \left( 1 - \frac{y - n_2}{n_2} \right) - \mu_3 + O(\|\mathbf{x} - \mathbf{n}\|^2) \\ &= \frac{x}{n_2} - \frac{n_1}{n_2^2}(y - n_2) - \mu_3 + O(\|\mathbf{x} - \mathbf{n}\|^2) \\ &= \frac{x}{n_2} - \frac{n_1 y}{n_2^2} + O(\|\mathbf{x} - \mathbf{n}\|^2) \\ &= \mathbf{h}_3 \cdot \mathbf{x} + O(\|\mathbf{x} - \mathbf{n}\|^2) \end{aligned} \quad (20)$$

where  $\mathbf{h}_3$  is the observation vector defined by (11). Similarly, on expanding  $(x-v)/(y-1) - \mu_2$  and  $(x-2v)/(y-2) - \mu_1$  about  $\mathbf{n}$  we obtain

$$\frac{x-v}{y-1} - \mu_2 = (G^{-1}\mathbf{x}) \cdot \mathbf{h}_2 + O(\|\mathbf{x} - \mathbf{n}\|^2) \quad (21)$$

$$\frac{x-2v}{y-2} - \mu_1 = (G^{-2}\mathbf{x}) \cdot \mathbf{h}_1 + O(\|\mathbf{x} - \mathbf{n}\|^2) \quad (22)$$

It follows from (20), (21) and (22) that

$$\begin{aligned} p_s(\mathbf{x}) &= \\ c \exp \left[ -\frac{1}{2\rho^2} ((\mathbf{h}_3 \cdot \mathbf{x})^2 + (\mathbf{h}_2 \cdot G^{-1}\mathbf{x})^2 + (\mathbf{h}_1 \cdot G^{-2}\mathbf{x})^2) \right] \times \\ &\quad \mathcal{N}(G^{-2}\mathbf{x}, \mathbf{m}, P) \end{aligned} \quad (23)$$

Let  $D$  be the covariance of  $p_s(\mathbf{x})$ . Equation (23) yields

$$D^{-1} =$$

$$\frac{1}{\rho^2} (\mathbf{h}_3 \otimes \mathbf{h}_3 + G^{-\top} \mathbf{h}_2 \otimes \mathbf{h}_2 G^{-1} + (G^{-2})^\top \mathbf{h}_1 \otimes \mathbf{h}_1 G^{-2}) + (F^2 P (F^2)^\top)^{-1} \quad (24)$$

It follows from (24) that  $D$  is equal to the covariance  $P_3$  obtained by the EKF.