Statistical Partial Constraints for 3D Model Matching and Pose Estimation Problems

M. Waite[1], M. Orr[2], R. Fisher[1], J. Hallam[1]

(1) Dept. of Artificial Intelligence, University of Edinburgh
5 Forrest Hill, Edinburgh EH1 2QL, Scotland, United Kingdom
(2) Advanced Robotics Research Ltd., University Road,
Salford M5 4PP, England, United Kingdom

Abstract

We explore the potential of variance matrices to represent not just statistical error on object pose estimates but also partially constrained degrees of freedom. Using an iterated extended Kalman filter as an estimation tool, we generate, combine and predict partially constrained pose estimates from 3D range data. We find that partial constraints on the translation component of pose which occur frequently in practice are handled well by the method. However, coupled partial constraints between rotation and translation are, in general, non-linear and cannot be represented by this method.

1 Introduction

Most model-based part recognition or location vision systems establish all model-to-data pairings during an initial matching phase, and then estimate the pose from the consistent pairings. This is less than ideal, as insufficient features may have been segmented to estimate fully the pose, or it may be desirable to improve the pose estimate by locating additional features using the current pose estimate. Some features may only provide partial or weak pose constraints.

This paper integrates three themes in computer vision to show how model matching can be improved. The themes are:

- incremental improvement of pose estimates as new evidence is found,

- a common representation for both statistical error and lack of knowledge and

- use of partial knowledge to guide model matching.

The paper demonstrates some examples of model matching or pose estimation problems where partial knowledge is integrated and used to improve the quality of scene understanding. The domain of application used for the examples here is 3D model matching using 3D image feature data, but the approach can be adapted for 2D-to-2D and 3D-to-2D problems. The examples shown in the paper are based on a surface-patch matching system where the data surface patches are extracted

from range data (by some adaptations of [3]) and the model surfaces are specialised instances of quadratic surfaces [5].

The foundation of the approach is based on representing uncertainty by the variance of an assumed normal probability distribution. This by itself is not new and a number of vision, robotics and tracking projects have followed this approach [12, 9, 13, 2]. That work has used variance to encode fully constrained but statistically uncertain poses. The advantage of this approach is that there are well-known statistical tools for estimation (e.g. the Kalman filter) and decision (e.g. the χ^2 test) problems.

However, there exists a class of problems where the uncertainty is not entirely due to statistical errors but has a component which would still be present even if the available measurements were perfectly accurate. Such problems occur when there are more parameters to estimate than measurements available – they are underconstrained problems. For example, the correspondence between a model point and a scene plane, in the sense that the transformed point must lie somewhere in the plane, is not sufficient to constrain the translational part of the transform even if the rotational part has been estimated by some other means. These problems leave degrees of freedom in the estimated transform which it would be convenient to represent in the same way as statistical uncertainty, i.e. using variance. As long as the uncertainty due to degrees of freedom is linear in parameter space, we can do this by introducing one or more large eigenvalues in the variance matrix.

In practice, the non-linear rotational part of a 3D transform is often easier to constrain than the linear translational part. This is due in part to the robustness of correspondences between plane surface normals against occlusion and segmentation errors. Correspondences between points (e.g. centre of gravity, boundary points) which are used to constrain translation are relatively fragile and subject to occlusion and must sometimes be replaced by partial constraints (such as a match between a point and a plane). Our approach is useful in these situations when the rotation is fully constrained but the translation can only be partially constrained.

The theory is given in Section 2. Illustrations matching 3D models against range images are given in Section 3. Conclusions are in Section 4.

The work reported here builds on techniques which have become standard in robotics and vision through the work of, among others, groups at INRIA [13] and Oxford University [9]. The approach to partial evidence representation is similar to that of [4] and [6] except that there intervals, which are known to be inferior to variance [10], were used to represent the bounds on the parameters. There are also links with early research into pose constraints from object relationships as specified in a robot programming language (RAPT) [11] though that work modeled noise free relationships.

2 The Statistical Approach to Uncertainty Representation

We use a Kalman filter as our basic estimation tool and begin with a brief description of its use. More details can readily be found elsewhere, e.g. [8, 1, 2].

Basically, the filter recursively processes observations to arrive at an estimate of an unobserved parameter of interest (the state). Knowledge at step k (after

processing the kth observation) about the state vector, $\mathbf{x}$, is represented by the estimated mean, $\hat{\mathbf{x}}_k$, and variance, $\mathbf{X}_k$, of an assumed Gaussian probability distribution. Observations, $\mathbf{z}_k$, pertaining to the state are themselves uncertain with means, $\hat{\mathbf{z}}_k$, and variances, $\mathbf{Z}_k$. To link the observations to the state there are measurement equations of the form

$$\mathbf{f}_k(\mathbf{x}, \mathbf{z}_k) = \mathbf{0}$$

which are usually non-linear and often under-constrained (cannot be put in the form $\mathbf{x} = \mathbf{g}_k(\mathbf{z}_k)$). The iterated extended Kalman filter (IEKF) is an adaptation of the basic filter to deal with non-linear measurement equations. In both cases incorporating the kth observation leads to an update of the state estimate to a new mean, $\hat{\mathbf{x}}_{k+1}$, and a new variance, $\mathbf{X}_{k+1}$. For the IEKF, the Jacaobians $\partial \mathbf{f}_k / \partial \mathbf{x}$ and $\partial \mathbf{f}_k / \partial \mathbf{z}_k$ (functions of $\mathbf{x}$ and $\mathbf{z}_k$), are also needed to perform the linearisation step. Appendices A and B list the measurement functions (and their Jacobians) used in this paper and Appendix C lists the Kalman Filter updating used.

The state variance matrix, $\mathbf{X}$, represents the size of an assumed Gaussian probability distribution in n-dimensional space (n is the dimension of the state vector, $\mathbf{x}$). Loosely speaking, it can be thought of as representing an n-dimensional ellipsoid centred on the mean, $\hat{\mathbf{x}}$, and containing the true state vector, $\mathbf{x}$. The ellipsoidal axes are parallel to the eigenvectors of $\mathbf{X}$ in direction and proportional to the square roots of the eigenvalues of $\mathbf{X}$ in length.

The uncertainty in a parameter estimate which has one *linear* degree of freedom can be represented by a variance matrix with a large eigenvalue in the appropriate direction. Two degrees of freedom in two different directions can be represented with two large eigenvalues, and so on. In cases where the degree of freedom is partial (over a finite range) rather than unbounded then suitably sized eigenvalues can be chosen. Of course, for very small uncertainties, non-linear constraints can be linearly approximated, and this is the basis of the usefulness of variance as a general representation of statistical measurement error.

To illustrate what the Kalman filter does, consider the following example (see also [13]). Suppose a point with unknown position $\mathbf{x}$ lies on a line with end-point $\mathbf{e}$ and direction $\mathbf{d}$ (a unit vector) at distance λ from $\mathbf{e}$ and suppose we obtain noisy observations of $\mathbf{e}$ and $\mathbf{d}$ (but not λ). The degree of freedom inherent in the lack of information about λ can be represented by assigning it a nominal mean, $\hat{\lambda}$, and a large variance, Λ. The appropriate measurement equation is

$$\mathbf{x} = \mathbf{e} + \lambda \mathbf{d} , \tag{1}$$

and the first order approximation for the variance

$$\mathbf{X} = \frac{\partial \mathbf{x}}{\partial \mathbf{e}} \mathbf{E} \frac{\partial \mathbf{x}}{\partial \mathbf{e}}^T + \frac{\partial \mathbf{x}}{\partial \lambda} \Lambda \frac{\partial \mathbf{x}}{\partial \lambda}^T + \frac{\partial \mathbf{x}}{\partial \mathbf{d}} \mathbf{D} \frac{\partial \mathbf{x}}{\partial \mathbf{D}}^T \tag{2}$$

where $\mathbf{E}$ and $\mathbf{D}$ are the variances of the estimates for, respectively, $\mathbf{e}$ and $\mathbf{d}$. From (1) the Jacobians can be derived and substituted in (2), which leads, in this case, to

$$\mathbf{X} = \mathbf{E} + \Lambda \hat{\mathbf{d}} \hat{\mathbf{d}}^T + \hat{\lambda}^2 \mathbf{D} , \tag{3}$$

where $\hat{\mathbf{d}}$ is the mean of the estimates for $\mathbf{d}$. If Λ is very large, $\mathbf{X}$ will have a large eigenvalue for an eigenvector lying in the direction of $\hat{\mathbf{d}}$.

This sort of computation is very like what the IEKF does. The main differences are (1) being recursive, the IEKF requires an initial state estimate which influences the new state estimate and (2) when the measurement equation is non-linear, the iterative linearisation produces a more accurate result. In the above example equation (1) would be used as the measurement equation. A nominal initial estimate could be set up as

$$\begin{aligned} \hat{\mathbf{x}}_0 &= \hat{\mathbf{e}} + \hat{\lambda}\hat{\mathbf{d}} , \\ \mathbf{X}_0 &= \sigma^2 \mathbf{I} \end{aligned}$$

($\hat{\mathbf{e}}$ is the mean of the estimate for $\mathbf{e}$, σ is large) and the observation is

$$\hat{\mathbf{z}}_0 = \begin{bmatrix} \hat{\mathbf{e}} \\ \hat{\mathbf{d}} \\ \hat{\lambda} \end{bmatrix} ,$$

$$\mathbf{Z}_0 = \begin{bmatrix} \mathbf{E} & 0 & 0 \\ 0 & \mathbf{D} & 0 \\ 0 & 0 & \Lambda \end{bmatrix} .$$

The IEKF computed variance would then be essentially the same as in (3).

As well as generating variance matrices containing degrees of freedom the IEKF can be used to combine together multiple observations each containing degrees of freedom into state estimates which are fully constrained. This is one of the main applications of our approach – two or more partial constraints combined to give a fuller constraint.

To combine estimates a Kalman filter is required, but for merely generating variances with degrees of freedom an alternative method can be used. The steps involved are (1) set up a diagonal matrix where one or more of the entries are large (corresponding to the degrees of freedom) and the others are small or zero and (2) rotate this matrix so that the large eigenvalues line up with the directions of the degrees of freedom. For example, the uncertainty of a point which lies somewhere along a line whose length is of the order of σ could be represented by

$$\Phi \begin{bmatrix} \epsilon^2 & 0 & 0 \\ 0 & \epsilon^2 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} \Phi^T$$

where Φ is any rotation matrix which rotates the z-axis into the line direction. The smaller eigenvalue, ϵ^2, can either be chosen to represent measurement error for a measured quantity or be set to zero for a model parameter[1]. However, this method depends on being able to sensibly choose the diagonal entries and the rotation matrix and it is not always obvious how to do this.

---

[1] Obviously, we avoid singular matrices and instead use a tiny number in place of zero.
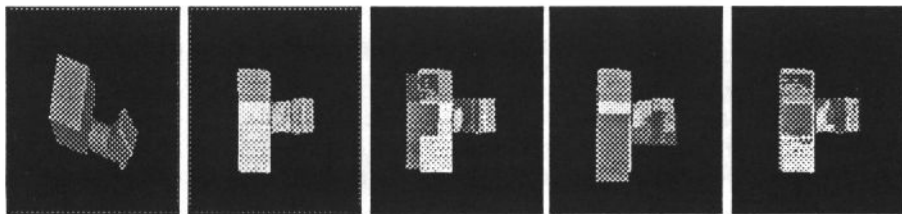
Figure 1: *A series of images showing the increased agreement between the mean position of an object model (dark) and the position of some real data (light) as two partially constrained subcomponent positions are used to refine the pose estimate. The model, which consists of one small cube and one large one, and the data are shown in the two images on the left. Neither subcomponent alone can accurately estimate the object's pose (third and fourth images) but the combination of both lead to an accurate estimate (image on the right).*

3 Some Applications of Partial Constraints

This section shows how a number of scene understanding problems can be represented and solved using this uncertainty approach. The first two problems (Subsections 3.1 and 3.2) are examples where partial translation constraints can be generated from matches between points, one of which is partially constrained. The last (Subsection 3.3) illustrates the combination of partial pose estimates.

We have also successfully applied our approach to:

- the incorporation of *a priori* partial constraints, by representing known knowledge as if it were an observed feature (e.g. knowing a given object point lies on the working surface is equivalent to measuring it as lying in a given data plane)

- the combination of partial pose estimates from disparate reference frames (i.e. integrating different subcomponent pose estimates in the common reference frame of the object - see Figure 1), and

- image feature position prediction from partial constraints (i.e. using the covariance distribution of the estimated parameters to predict a range of possible positions for a desired feature).

3.1 Planar Patch Matching

Suppose the model matching and reasoning module of a vision system has paired a number of model and data planar patch surface normals and from these estimated a rotation by using an IEKF with the measurement equation detailed in Appendix A. An estimate of the translation has yet to be made but a constraint is available from the pairing of a model patch central point and an observed (possibly partially occluded) scene patch. Three (of the six) spatial degrees of freedom are already constrained. One translational degree of freedom can be constrained by the requirement that the transformed model point must lie in the plane of the data

surface and there are loose constraints on the other two because the incomplete data patch must lie within the boundaries of the transformed model patch.

One way to account for the partial constraint is to create a pairing between the infinite plane parameters of the model and data patches. However, a better method, which accounts, at least in a crude way, for the finite size of the patch, is to create a pairing between the scene point and the model point and give the model point large variance eigenvalues in the plane of the model patch. The variance of the model point then has the characteristic elliptical shape

$$
\Phi \begin{bmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \Phi^T
$$

where σ_1 and σ_2 are the major and minor axes of the smallest ellipse fitting round the model patch. Φ is the rotation matrix which rotates the z-axis into the surface normal of the patch and the x-axis into the major axis of the surrounding ellipse.

Having created the point-to-point pairing and attached the model variance as indicated, the constraint can be processed into a new estimate for the translation using a Kalman filter and the measurement equation in Appendix B.

3.2   Cylindrical Patch Matching

As for the previous section, we suppose the rotation component of a pose estimate has already been established, but this time we suppose that the constraint on translation comes from a pairing between a cylindrical model patch and a cylindrical data patch. When rotated and translated into position, the model patch must have the same axis as the data patch (within measurement errors).

We can account for this partial constraint by pairing up the central point of the scene patch axis with the central point on the model axis and by giving the model point a degree of freedom in the direction of the cylinder axis. The variance matrix of the model point is

$$
\Phi \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma^2 \end{bmatrix} \Phi^T
$$

where σ is of the order of half the model axis length, and Φ is any rotation matrix which rotates the z-axis into the data axis. As in the previous subsection, we can then process the constraint using the measurement equation in Appendix B.

3.3   Integration of Partial Estimates

In general, if model matching has produced some number of direction matches to constrain the rotation, then there will be just as many partial constraints on translation by pairing up model and data points, since each surface patch contributes one normal and one central point. The combination of three or more
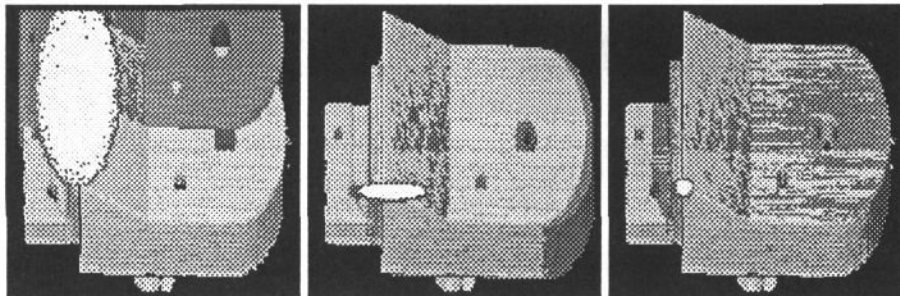
Figure 2: *A series of images showing the increased agreement between the mean position of an object model (dark) and the position of some real data (light) as three partial translation constraints are used to refine the pose estimate. The decreasing variance of the model position is also depicted by showing the decreasing size of the uncertainty ellipsoid associated with one of the model vertices.*

partial constraints from point matches will, except in degenerate cases, lead to a fully constrained translation estimate where the eigenvalues of the variance matrix are primarily determined by the measurement errors.

To achieve this combination of constraints the point-to-point pairings are processed recursively using the measurement equation in Appendix B. The output state estimate from the processing of one pair becomes the input estimate for the next pair. The initial estimate contains the previously estimated rotation and a nominal translation estimate (large variance). The final estimate, barring accidental alignment of degrees of freedom, will not have any large variance eigenvalues.

We illustrate with observations of three surfaces, two planar and one cylindrical, belonging to one object (Figure 2). The rotation mean is estimated using an SVD analysis of the paired direction means ([7], page 431) and a large variance, then an IEKF with measurement equation as in Appendix A is used to estimate the rotation. The translation is also initialised with a large variance but with an arbitrary mean (the zero vector). In Figure 2 we show the relative position of the model and data after each of three independent translation constraints have been incorporated into the pose estimate. When a model surface (dark) is close to a data surface (light) the graphics program which produced these figures tends to intermingle dark and light pixels. The intermingling effect shows clearly which surface, or surfaces, have been used to constrain the translation in each image. In each image the variance of one of the object model's vertices has been depicted by drawing an ellipse around the predicted position of the point. As each partial constraint is processed the ellipse can be seen to shrink along different directions.

4   Conclusions

Large variances are effective for encoding partial translation constraints, and the Kalman filter is an effective tool for resolving the constraints to produce fully constrained pose estimates. Moreover, the pose estimates are very good, as demonstrated by the interweaving observed between the raw range data and the projected model surfaces in Figure 2. The method is more accurate than previous methods

which used bounding intervals to represent uncertainty. However, the method is not able to cope with non-linear constraints such as the coupled constraints that are often generated between rotation and translation when there is only a partial rotation estimate.

Future work could investigate the possibility of analysing the variance matrix to deduce which large degrees of freedom remain, and thus what type of constraints would be useful for optimally reducing the uncertainty and where to search in the image for them.

Acknowledgements

References

[1] N. Ayache and O.D. Faugeras. Maintaining representations of the environment of a mobile robot. In *Robotics Research 4*, pages 337–350. MIT Press, USA, 1988.

[2] Y. Bar-Shalom and T.E. Fortmann. *Tracking and Data Association*. Academic Press, UK, 1988.

[3] P.J. Besl. *Surfaces in Range Image Understanding*. Springer-Verlag, 1987.

[4] R.A. Brooks. Symbolic reasoning among 3D models and 2-d images. *Artificial Intelligence Journal*, 17:285–348, 1981.

[5] R.B. Fisher. SMS: A suggestive modeling system for object recognition. *Image and Vision Computing*, 5(2):98–104, 1987.

[6] R.B. Fisher. *From Surfaces to Objects: Computer Vision and Three Dimensional Scene Analysis*. John Wiley, UK, 1989.

[7] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, USA, 1985.

[8] A.H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, USA, 1970.

[9] J. J. Leonard, H.F. Durrant-Whyte, and I.J. Cox. Dynamic map building for an autonomous mobile robot. *International Journal of Robotics Research*, 11(4):286–298, 1992.

[10] M.J.L. Orr, R.B. Fisher, and J. Hallam. Uncertain reasoning: Intervals versus probabilities. In *British Machine Vision Conference*, pages 351–354. Springer-Verlag, 1991.

[11] R.J. Popplestone, A.P. Ambler, and I.M. Bellos. An interpreter for a language describing assemblies. *Artificial Intelligence*, 14:79, 1980.

[12] J. Porril, S.B. Pollard, and J.E.W. Mayhew. Optimal combination of multiple sensors including stereo vision. *Image and Vision Computing*, 5(2):174–180, 1987.

[13] Z. Zhang and O.D. Faugeras. A 3D world model builder with a mobile robot. *International Journal of Robotics Research*, 11(4):269–285, 1992.

Appendix: Partial Derivatives for the IEKF

A: Estimating Rotations from Matched Directions

This is the problem of estimating a exponential representation rotation vector, $\mathbf{r}$ (the product of the rotation axis $\bar{\omega}$ and angle ϕ), from pairs of matched vectors, $\mathbf{u}_k$ and $\mathbf{v}_k$, such that $\mathbf{v}_k$ is the rotation (by $\mathbf{r}$) of $\mathbf{u}_k$. The state vector is $\mathbf{x} = \mathbf{r}$, the observation vectors are $\mathbf{z}_k = [\mathbf{v}_k^T \ \mathbf{u}_k^T]^T$ and the measurement equation for each observation is

$$\mathbf{f}(\mathbf{x}, \ \mathbf{z}_k) \ = \ \mathbf{v}_k - \mathbf{\Phi}\,\mathbf{u}_k \ = \ 0$$

where

$$\mathbf{\Phi} \ = \ \mathbf{I} + \frac{\sin\phi}{\phi}\,\mathbf{H} + \frac{1 - \cos\phi}{\phi^2}\,\mathbf{H}^2$$

$$\phi \ = \ \|\,\mathbf{r}\,\|$$

and

$$\mathbf{H} \ = \ \begin{bmatrix} 0 & -r_3 & r_2 \\ r_3 & 0 & -r_1 \\ -r_2 & r_1 & 0 \end{bmatrix}.$$

The derivatives of the measurement function are

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} \ = \ \begin{bmatrix} -\dfrac{\partial \mathbf{\Phi}}{\partial r_1}\mathbf{u}_k & -\dfrac{\partial \mathbf{\Phi}}{\partial r_2}\mathbf{u}_k & -\dfrac{\partial \mathbf{\Phi}}{\partial r_3}\mathbf{u}_k \end{bmatrix}$$

$$\frac{\partial \mathbf{f}}{\partial \mathbf{z}_k} \ = \ [\mathbf{I} \ -\mathbf{\Phi}]$$

where

$$\frac{\partial \mathbf{\Phi}}{\partial r_i} \ = \ \frac{\sin\phi}{\phi}\,\mathbf{H}_i + \frac{r_i}{\phi^3}\,(\phi\cos\phi - \sin\phi)\,\mathbf{H} +$$

$$\frac{r_i}{\phi^4}\,(\phi\sin\phi - 2\,(1 - \cos\phi))\,\mathbf{H}^2 + \frac{1 - \cos\phi}{\phi^2}\,(\mathbf{H}\mathbf{H}_i + \mathbf{H}_i\mathbf{H})$$

for $i = 1, 2, 3$ and the basis matrices $\mathbf{H}_i$ are given by

$$\mathbf{H}_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{H}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad \mathbf{H}_3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

B: Estimating Translation from Matched Points

This is the problem of estimating the translational component of a 3D transform from pairs of matched points, $\mathbf{p}_k$ and $\mathbf{q}_k$, such that $\mathbf{q}_k$ is the transform (by an already estimated rotation $\mathbf{r}$ and translation $\mathbf{t}$) of $\mathbf{p}_k$. The state vector is $\mathbf{x} = [\mathbf{r}^T \ \mathbf{t}^T]$, the observation vectors are $\mathbf{z}_k = [\mathbf{q}_k^T \ \mathbf{p}_k^T]^T$ and the measurement equation for each observation is

$$\mathbf{f}(\mathbf{x}, \mathbf{z}_k) = \mathbf{q}_k - \mathbf{\Phi}\,\mathbf{p}_k - \mathbf{t} = 0$$

where $\mathbf{\Phi}$ (a function of $\mathbf{r}$, see Appendix A) is the rotation matrix. The Jacobians are

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \left[-\frac{\partial \mathbf{\Phi}}{\partial r_1}\mathbf{p}_k \quad -\frac{\partial \mathbf{\Phi}}{\partial r_2}\mathbf{p}_k \quad -\frac{\partial \mathbf{\Phi}}{\partial r_3}\mathbf{p}_k \quad -\mathbf{I} \right]$$

$$\frac{\partial \mathbf{f}}{\partial \mathbf{z}_k} = [\mathbf{I} \ -\mathbf{\Phi}] \ .$$

The derivatives $\partial \mathbf{\Phi}/\partial r_i$, $i = 1, 2, 3$ are given above in Appendix A.

C: Kalman Filter Update Equations

This appendix gives the Iterated Extended Kalman Filter updating equations for the problems described above (using the notation given in the two previous appendices). Let:

$$\mathbf{M}_k = \frac{\partial \mathbf{f}_k}{\partial \mathbf{x}}$$

$$\mathbf{W}_k = \frac{\partial \mathbf{f}_k}{\partial \mathbf{z}_k}\mathbf{\Lambda}_k \left(\frac{\partial \mathbf{f}_k}{\partial \mathbf{z}_k}\right)^t$$

where $\mathbf{\Lambda}_k$ is the covariance matrix of the observations $\mathbf{z}_k$ and $\mathbf{f}_k$ is evaluated using the current values $(\mathbf{z}_k, \ \hat{\mathbf{x}}_{i-1})$.

Then, the update equation of *extended* Kalman filter for producing the current estimate $\mathbf{x}_k$ of the state vector is:

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{i-1} - \mathbf{K}_k\mathbf{f}_k(\mathbf{z}_k, \ \hat{\mathbf{x}}_{i-1})$$

and its associated covariance estimate is:

$$\mathbf{S}_k^{-1} = \mathbf{S}_{i-1}^{-1} + \mathbf{M}_k^t\mathbf{W}_k^{-1}\mathbf{M}_k$$

where

$$\mathbf{K}_k = \mathbf{S}_{i-1}\mathbf{M}_k^t(\mathbf{W}_k + \mathbf{M}_k\mathbf{S}_{i-1}\mathbf{M}_k^t)^{-1}$$

is the Kalman gain.

If the estimate $\hat{\mathbf{x}}_{i-1}$ around which the Taylor expansion is performed is too far from the correct parameter $\mathbf{x}$, linearisation may not be very good. A method to reduce the linearisation error is the *iterated* extended Kalman filter. This applies the update equation for the mean, $\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{i-1} - \mathbf{K}_k\mathbf{f}_k(\mathbf{z}_k, \ \hat{\mathbf{x}}_{i-1})$, as long as $\hat{\mathbf{x}}_k - \hat{\mathbf{x}}_{i-1}$ is too large, computing at each iteration a new value of $\mathbf{K}_k$ and $\mathbf{M}_k$ obtained from a re-linearisation of $\mathbf{f}_k$ about the new estimate $\hat{\mathbf{x}}_k$.