Chapter 19

Temporal models

This chapter concerns temporal models and tracking. The goal is to infer a sequence of world states $\{\mathbf{w}_t\}_{t=1}^T$ from a noisy sequence of measurements $\{\mathbf{x}_t\}_{t=1}^T$. The world states are not independent; each is modeled as being contingent on the previous one. We exploit this statistical dependency to help estimate the state $\mathbf{w}_t$, even when the associated observation $\mathbf{x}_t$ is partially or completely uninformative.

Since the states form a chain, the resulting models are similar to those in chapter 11. However, there are two major differences. First, in this chapter, we consider models where the world state is continuous rather than discrete. Second, the models here are designed for real-time applications; a judgment is made based on only the past and present measurements and without knowledge of those in the future.

A prototypical use of these temporal models is for *contour tracking*. Consider a parameterized model of an object contour (figure 19.1). The goal is to track this contour through a sequence of images so that it remains firmly attached to the object. A good model should be able to cope with non-rigid deformations of the object, background clutter, blurring, and occasional occlusion.

19.1 Temporal estimation framework

Each model in this chapter consists of two components.

- The *measurement model* describes the relationship between the measurements $\mathbf{x}_t$ and the state $\mathbf{w}_t$ at time t. We treat this as generative, and model the likelihood $Pr(\mathbf{x}_t|\mathbf{w}_t)$. We assume that the data at time t depend only on the state at time t and not those at any other time. In mathematical terms, we assume that $\mathbf{x}_t$ is conditionally independent of $\mathbf{w}_{1\ldots t-1}$, given $\mathbf{w}_t$.

- The *temporal model* describes the relationship between states. Typically, we make the *Markov assumption*: we assume that each state depends only upon its predecessor. More formally, we assume that $\mathbf{w}_t$ is conditionally independent of the states $\mathbf{w}_{1\ldots t-2}$ given its immediate predecessor $\mathbf{w}_{t-1}$, and just model the relationship $Pr(\mathbf{w}_t|\mathbf{w}_{t-1})$.

Figure 19.1 Contour tracking. The goal is to track the contour (solid blue line) through the sequence of images so that it remains firmly attached to the object (see chapter 17 for more information about contour models). The estimation of the contour parameters is based on a temporal model relating nearby frames, and local measurements from the image (e.g., between the dashed blue lines). Adapted from Blake & Isard (1998). ©1998 Springer.

Together, these assumptions lead to the graphical model shown in figure 19.2.

19.1.1   Inference

In inference, the general problem is to compute the marginal posterior distribution $Pr(\mathbf{w}_t|\mathbf{x}_{1...t})$ over the world state $\mathbf{w}_t$ at time t given all of the measurements $\mathbf{x}_{1...t}$ up until this time. At time $t = 1$, we have only observed a single measurement $\mathbf{x}_1$, so our prediction is based entirely on this datum. To compute the posterior distribution $Pr(\mathbf{w}_1|\mathbf{x}_1)$, Bayes' rule is applied:

$$Pr(\mathbf{w}_1|\mathbf{x}_1) = \frac{Pr(\mathbf{x}_1|\mathbf{w}_1)Pr(\mathbf{w}_1)}{\int Pr(\mathbf{x}_1|\mathbf{w}_1)Pr(\mathbf{w}_1)\,d\mathbf{w}_1}, \tag{19.1}$$

where the distribution $Pr(\mathbf{w}_1)$ contains our prior knowledge about the initial state.

At time $t = 2$ we observe a second measurement $\mathbf{x}_2$. We now aim to compute the posterior distribution of the state at time $t = 2$ based on both $\mathbf{x}_1$ and $\mathbf{x}_2$. Again, we apply Bayes' rule,

$$Pr(\mathbf{w}_2|\mathbf{x}_1, \mathbf{x}_2) = \frac{Pr(\mathbf{x}_2|\mathbf{w}_2)Pr(\mathbf{w}_2|\mathbf{x}_1)}{\int Pr(\mathbf{x}_2|\mathbf{w}_2)Pr(\mathbf{w}_2|\mathbf{x}_1)\,d\mathbf{w}_2}. \tag{19.2}$$

Notice that the likelihood term $Pr(\mathbf{x}_2|\mathbf{w}_2)$ is only dependent on the current measurement (according to assumption stated earlier). The prior $Pr(\mathbf{w}_2|\mathbf{x}_1)$ is now based on what we have learned from the previous measurement; the possible values of the state at this time depend on our knowledge of what happened at the previous time and how these are affected by the temporal model.

Generalizing this procedure to time t, we have

$$Pr(\mathbf{w}_t|\mathbf{x}_{1...t}) = \frac{Pr(\mathbf{x}_t|\mathbf{w}_t)Pr(\mathbf{w}_t|\mathbf{x}_{1...t-1})}{\int Pr(\mathbf{x}_t|\mathbf{w}_t)Pr(\mathbf{w}_t|\mathbf{x}_{1...t-1})\,d\mathbf{w}_t}. \tag{19.3}$$
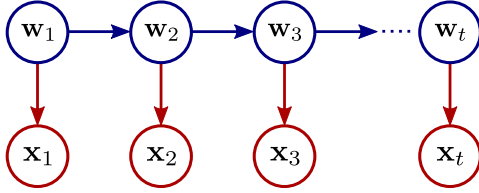
Figure 19.2 Graphical model for Kalman filter and other temporal models in this chapter. This implies the following conditional independence relation: the state $\mathbf{w}_t$ is conditionally independent of the states $\mathbf{w}_{1...t-2}$ and the measurements $\mathbf{x}_{1...t-1}$, given the previous state $\mathbf{w}_{t-1}$.

To evaluate this, we must compute $Pr(\mathbf{w}_t|\mathbf{x}_{1...t-1})$, which represents our prior knowledge about $\mathbf{w}_t$ before we look at the associated measurement $\mathbf{x}_t$. This prior depends on our knowledge $Pr(\mathbf{w}_{t-1}|\mathbf{x}_{1...t-1})$ of the state at the previous time and the temporal model $Pr(\mathbf{w}_t|\mathbf{w}_{t-1})$, and is computed recursively as

$$Pr(\mathbf{w}_t|\mathbf{x}_{1...t-1}) = \int Pr(\mathbf{w}_t|\mathbf{w}_{t-1})Pr(\mathbf{w}_{t-1}|\mathbf{x}_{1...t-1})\, d\mathbf{w}_{t-1}, \tag{19.4}$$

which is known as the *Chapman-Kolmogorov* relation. The first term in the integral represents the prediction for the state at time t given a known state $\mathbf{w}_{t-1}$ at time $t-1$. The second term represents the uncertainty about what the state actually was at time $t-1$. The Chapman-Kolmogorov equation amalgamates these two pieces of information to predict what will happen at time t.

Hence, inference consists of two alternating steps. In the *prediction step*, we compute the prior $Pr(\mathbf{w}_t|\mathbf{x}_{1...t-1})$ using the Chapman-Kolmogorov relation (equation 19.4). In the *measurement incorporation* step, we combine this prior with the new information from the measurement $\mathbf{x}_t$ using Bayes's rule (equation 19.3).

19.1.2   Learning

The goal of learning is to estimate the parameters $\boldsymbol{\theta}$ that determine the relationship $Pr(\mathbf{w}_t|\mathbf{w}_{t-1})$ between adjacent states and the relationship $Pr(\mathbf{x}_t|\mathbf{w}_t)$ between the state and the data, based on several observed time sequences.

If we know the states for these sequences, then this can be achieved using the maximum likelihood method. If the states are unknown, then they can be treated as *hidden variables*, and the model can be learned using the EM algorithm (section 7.8). In the E-step, we compute the posterior distribution over the states for each time sequence. This is a process similar to the inference method described earlier, except that it also uses data from later in the sequence (section 19.2.6). In the M-step, we update the EM bound with respect to $\boldsymbol{\theta}$.

The rest of this chapter focusses on inference in this type of temporal model. We will first consider the *Kalman filter*. Here the uncertainty over the world state is described by a normal distribution,[1] the relationship between the measurements

---

[1]In the original formulation of the Kalman filter, it was only assumed that the noise was white; however, if the distribution is normal, we can compute the exact marginal posterior distributions, so we favor this assumption.

and the world is linear with additive normal noise, and the relationship between the state at adjacent times is also linear with additive normal noise.

19.2   Kalman filter

To define the Kalman filter, we must specify the temporal and measurement models. First, the temporal model relates the states at times $t-1$ and t and is given by

$$\mathbf{w}_t = \boldsymbol{\mu}_p + \boldsymbol{\Psi}\mathbf{w}_{t-1} + \boldsymbol{\epsilon}_p, \tag{19.5}$$

where $\boldsymbol{\mu}_p$ is a $D_{\mathbf{w}} \times 1$ vector, which represents the mean change in the state and $\boldsymbol{\Psi}$ is a $D_{\mathbf{w}} \times D_{\mathbf{w}}$ matrix, which relates the mean of the state at time t to the state at time $t-1$. This is known as the *transition* matrix. The term $\boldsymbol{\epsilon}_p$ is a realization of the transition noise, which is normally distributed with covariance $\boldsymbol{\Sigma}_p$ and determines how closely related the states are at times t and $t-1$. Alternately, we can write this in probabilistic form:

$$Pr(\mathbf{w}_t|\mathbf{w}_{t-1}) = \text{Norm}_{\mathbf{w}_t}[\boldsymbol{\mu}_p + \boldsymbol{\Psi}\mathbf{w}_{t-1}, \boldsymbol{\Sigma}_p]. \tag{19.6}$$

Second, the measurement model relates the data $\mathbf{x}_t$ at time t to the state $\mathbf{w}_t$,

$$\mathbf{x}_t = \boldsymbol{\mu}_m + \boldsymbol{\Phi}\mathbf{w}_t + \boldsymbol{\epsilon}_m, \tag{19.7}$$

where $\boldsymbol{\mu}_m$ is a $D_{\mathbf{x}} \times 1$ mean vector and $\boldsymbol{\Phi}$ is a $D_{\mathbf{x}} \times D_{\mathbf{w}}$ matrix relating the $D_{\mathbf{x}} \times 1$ measurement vector to the $D_{\mathbf{w}} \times 1$ state. The term $\boldsymbol{\epsilon}_m$ is a realization of the measurement noise, which is normally distributed with covariance $\boldsymbol{\Sigma}_m$. In probabilistic notation, we have

$$Pr(\mathbf{x}_t|\mathbf{w}_t) = \text{Norm}_{\mathbf{x}_t}[\boldsymbol{\mu}_m + \boldsymbol{\Phi}\mathbf{w}_t, \boldsymbol{\Sigma}_m]. \tag{19.8}$$

Notice that the measurement equation is identical to the relation between the data and the hidden variable in the factor analysis model (section 7.6); here the state $\mathbf{w}$ replaces the hidden variable $\mathbf{h}$. In the context of the Kalman filter, the dimension $D_{\mathbf{w}}$ of the state $\mathbf{w}$ is often larger than the dimension $D_{\mathbf{x}}$ of the measurements $\mathbf{x}$, so $\boldsymbol{\Phi}$ is a *landscape* matrix, and the measurement noise $\boldsymbol{\Sigma}_m$ is not necessarily diagonal.

The form of both the temporal and measurement equations is the same: each is a normal probability distribution where the mean is a linear function of another variable and the variance is constant. This form has been carefully chosen because it ensures that if the marginal posterior $Pr(\mathbf{w}_{t-1}|\mathbf{x}_{1...t-1})$ at time $t-1$ was normal, then so is the marginal posterior $Pr(\mathbf{w}_t|\mathbf{x}_{1...t})$ at time t. Hence, the inference procedure consists of a recursive updating of the means and variances of these normal distributions. We now elaborate on this procedure.

19.2.1 Inference

In inference, the goal is to compute the posterior probability $Pr(\mathbf{w}_t|\mathbf{x}_{1...t})$ over the state $\mathbf{w}_t$ given all of the measurements $\mathbf{x}_{1...t}$ so far. As before, we apply the prediction and measurement-incorporation steps to recursively estimate $Pr(\mathbf{w}_t|\mathbf{x}_{1...t})$ from $Pr(\mathbf{w}_{t-1}|\mathbf{x}_{1...t-1})$. The latter distribution is assumed to be normal with mean $\boldsymbol{\mu}_{t-1}$ and variance $\boldsymbol{\Sigma}_{t-1}$.

In the prediction step, we compute the prior at time t using the Chapman-Kolmogorov equation

Problem 19.1

$$
\begin{aligned}
Pr(\mathbf{w}_t|\mathbf{x}_{1...t-1}) &= \int Pr(\mathbf{w}_t|\mathbf{w}_{t-1})Pr(\mathbf{w}_{t-1}|\mathbf{x}_{1...t-1})\, d\mathbf{w}_{t-1}\\
&= \int \mathrm{Norm}_{\mathbf{w}_t}[\boldsymbol{\mu}_p + \boldsymbol{\Psi}\mathbf{w}_{t-1}, \boldsymbol{\Sigma}_p]\mathrm{Norm}_{\mathbf{w}_{t-1}}[\boldsymbol{\mu}_{t-1}, \boldsymbol{\Sigma}_{t-1}]\, d\mathbf{w}_{t-1}\\
&= \mathrm{Norm}_{\mathbf{w}_t}[\boldsymbol{\mu}_p + \boldsymbol{\Psi}\boldsymbol{\mu}_{t-1}, \boldsymbol{\Sigma}_p + \boldsymbol{\Psi}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Psi}^T]\\
&= \mathrm{Norm}_{\mathbf{w}_t}[\boldsymbol{\mu}_+, \boldsymbol{\Sigma}_+],
\end{aligned}
\tag{19.9}
$$

where we have denoted the predicted mean and variance of the state by $\boldsymbol{\mu}_+$ and $\boldsymbol{\Sigma}_+$. The integral between lines 2 and 3 was solved by using equations 5.17 and 5.14 to rewrite the integrand as proportional to a normal distribution in $\mathbf{w}_{t-1}$. Since the integral of any pdf is one, the result is the constant of proportionality, which is itself a normal distribution in $\mathbf{w}_t$.

In the measurement incorporation step, we apply Bayes' rule,

Problem 19.2

$$
\begin{aligned}
Pr(\mathbf{w}_t|\mathbf{x}_{1...t}) &= \frac{Pr(\mathbf{x}_t|\mathbf{w}_t)Pr(\mathbf{w}_t|\mathbf{x}_{1...t-1})}{Pr(\mathbf{x}_{1...t})}\\
&= \frac{\mathrm{Norm}_{\mathbf{x}_t}[\boldsymbol{\mu}_m + \boldsymbol{\Phi}\mathbf{w}_t, \boldsymbol{\Sigma}_m]\mathrm{Norm}_{\mathbf{w}_t}[\boldsymbol{\mu}_+, \boldsymbol{\Sigma}_+]}{Pr(\mathbf{x}_{1...t})}\\
&= \mathrm{Norm}_{\mathbf{w}_t}\left[\left(\boldsymbol{\Phi}^T\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\Phi} + \boldsymbol{\Sigma}_+^{-1}\right)^{-1}\left(\boldsymbol{\Phi}^T\boldsymbol{\Sigma}_m^{-1}(\mathbf{x}_t - \boldsymbol{\mu}_m) + \boldsymbol{\Sigma}_+^{-1}\boldsymbol{\mu}_+\right),\right.\\
&\qquad\qquad\qquad\qquad\qquad\qquad\left. \left(\boldsymbol{\Phi}^T\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\Phi} + \boldsymbol{\Sigma}_+^{-1}\right)^{-1}\right]\\
&= \mathrm{Norm}_{\mathbf{w}_t}[\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t],
\end{aligned}
\tag{19.10}
$$

where we have used equation 5.17 on the likelihood term and then combined the likelihood and prior using equation 5.14. The right-hand side is now proportional to a normal distribution in $\mathbf{w}_t$, and the constant of proportionality must be one to ensure that the posterior on the left hand side is a valid distribution.

Notice that the posterior $Pr(\mathbf{w}_t|\mathbf{x}_{1...t})$ is normal with mean $\boldsymbol{\mu}_t$ and covariance $\boldsymbol{\Sigma}_t$. We are in the same situation as at the start, so the procedure can be repeated for the next time step.

It can be shown that the mean of the posterior is a weighted sum of the values predicted by the measurements and the prior knowledge, and the covariance is smaller than either. However, this is not particularly obvious from the equations. In the following section we re-write the equation for the posterior in a form that makes these properties more obvious.
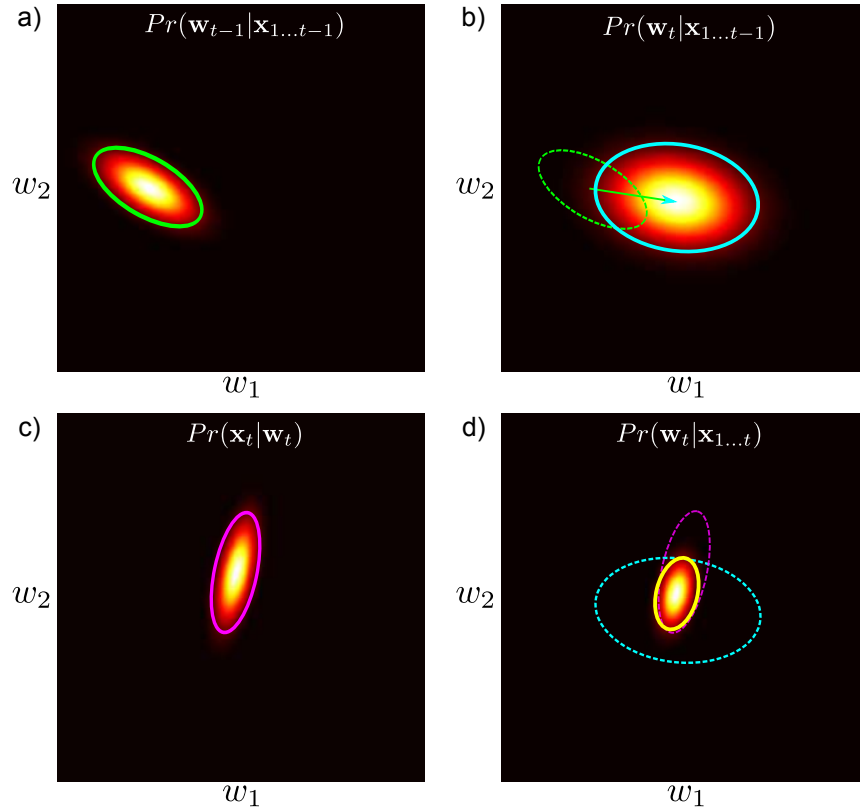
Figure 19.3 Recursive inference in the Kalman filter. a) The posterior probability $Pr(\mathbf{w}_{t-1}|\mathbf{x}_{1...t-1})$ of the state $\mathbf{w}_{t-1}$, given all of the measurements $\mathbf{x}_{1...t-1}$ up to that time, takes the form of a normal distribution (green ellipse). b) In the prediction step, we apply the Chapman-Kolmogorov relation to estimate the prior $Pr(\mathbf{w}_t|\mathbf{x}_{1...t-1})$ which is also a normal distribution (cyan ellipse). c) The measurement likelihood $Pr(\mathbf{x}_t|\mathbf{w}_t)$ is proportional to a normal distribution (magenta ellipse). d) To compute the posterior probability $Pr(\mathbf{w}_t|\mathbf{x}_{1...t})$, we apply Bayes' rule by taking the product of the prior from panel (b) and the likelihood from panel (c) and normalizing. This yields a new normal distribution (yellow ellipse) and the procedure can begin anew.

19.2.2   Rewriting measurement incorporation step

The measurement incorporation step is rarely presented in the above form in practice. One reason for this is that the equations for $\boldsymbol{\mu}_t$ and $\boldsymbol{\Sigma}_t$ contain an inversion that is of size $D_{\mathbf{w}} \times D_{\mathbf{w}}$. If the world state is much higher dimensional than the observed data, then it would be more efficient to reformulate this as an inverse of size $D_{\mathbf{x}} \times D_{\mathbf{x}}$. To this end, we define the *Kalman gain* as

$$\mathbf{K} = \mathbf{\Sigma}_+ \mathbf{\Phi}^T (\mathbf{\Sigma}_m + \mathbf{\Phi}\mathbf{\Sigma}_+\mathbf{\Phi}^T)^{-1}. \tag{19.11}$$

We will use this to modify the expressions for $\boldsymbol{\mu}_t$ and $\mathbf{\Sigma}_t$ from equation 19.10. Starting with $\boldsymbol{\mu}_t$, we apply the matrix inversion lemma (appendix C.8.4):

$$
\begin{aligned}
(\mathbf{\Phi}^T\mathbf{\Sigma}_m^{-1}\mathbf{\Phi} + \mathbf{\Sigma}_+^{-1})^{-1}&(\mathbf{\Phi}^T\mathbf{\Sigma}_m^{-1}(\mathbf{x}_t - \boldsymbol{\mu}_m) + \mathbf{\Sigma}_+^{-1}\boldsymbol{\mu}_+) \\
&= \mathbf{K}(\mathbf{x}_t - \boldsymbol{\mu}_m) + (\mathbf{\Phi}^T\mathbf{\Sigma}_m^{-1}\mathbf{\Phi} + \mathbf{\Sigma}_+^{-1})^{-1}\mathbf{\Sigma}_+^{-1}\boldsymbol{\mu}_+ \\
&= \mathbf{K}(\mathbf{x}_t - \boldsymbol{\mu}_m) + (\mathbf{\Sigma}_+ - \mathbf{\Sigma}_+\mathbf{\Phi}^T(\mathbf{\Phi}\mathbf{\Sigma}_+\mathbf{\Phi}^T + \mathbf{\Sigma}_m)^{-1}\mathbf{\Phi}\mathbf{\Sigma}_+)\mathbf{\Sigma}_+^{-1}\boldsymbol{\mu}_+ \\
&= \mathbf{K}(\mathbf{x}_t - \boldsymbol{\mu}_m) + \boldsymbol{\mu}_+ - \mathbf{\Sigma}_+\mathbf{\Phi}^T(\mathbf{\Phi}\mathbf{\Sigma}_+\mathbf{\Phi}^T + \mathbf{\Sigma}_m)^{-1}\mathbf{\Phi}\boldsymbol{\mu}_+ \\
&= \mathbf{K}(\mathbf{x}_t - \boldsymbol{\mu}_m) + \boldsymbol{\mu}_+ - \mathbf{K}\mathbf{\Phi}\boldsymbol{\mu}_+ \\
&= \boldsymbol{\mu}_+ + \mathbf{K}\left(\mathbf{x}_t - \boldsymbol{\mu}_m - \mathbf{\Phi}\boldsymbol{\mu}_+\right), \tag{19.12}
\end{aligned}
$$

The expression in brackets in the final line is known as the *innovation*, and is the difference between the actual measurements, $\mathbf{x}_t$ and the predicted measurements $\boldsymbol{\mu}_m + \mathbf{\Phi}\boldsymbol{\mu}_+$ based on the prior estimate of the state. It is easy to see why $\mathbf{K}$ is termed the Kalman gain: it determines the amount that the measurements contribute to the new estimate in each direction in state space. If the Kalman gain is small in a given direction, then this implies that the measurements are unreliable relative to the prior and should not influence the mean of the state too much. If the Kalman gain is large in a given direction, then this suggests that the measurements are more reliable than the prior and should be weighted more highly.

We now return to the covariance term of equation 19.10. Using the matrix inversion lemma, we get

$$
\begin{aligned}
(\mathbf{\Phi}^T\mathbf{\Sigma}_m\mathbf{\Phi} + \mathbf{\Sigma}_+^{-1})^{-1} &= \mathbf{\Sigma}_+ - \mathbf{\Sigma}_+\mathbf{\Phi}^T(\mathbf{\Phi}\mathbf{\Sigma}_+\mathbf{\Phi}^T + \mathbf{\Sigma}_m)^{-1}\mathbf{\Phi}\mathbf{\Sigma}_+ \\
&= \mathbf{\Sigma}_+ - \mathbf{K}\mathbf{\Phi}\mathbf{\Sigma}_+ \\
&= (\mathbf{I} - \mathbf{K}\mathbf{\Phi})\mathbf{\Sigma}_+, \tag{19.13}
\end{aligned}
$$

which also has a clear interpretation: the posterior covariance is equal to the prior covariance less a term that depends on the Kalman gain: we are always more certain about the state after incorporating information due to the measurement, and the Kalman gain modifies how much more certain we are. When the measurements are more reliable, the Kalman gain is high, and the covariance decreases more.

After these manipulations, we can rewrite equation 19.10 as

$$Pr(\mathbf{w}_t|\mathbf{x}_{1\ldots t}) = \text{Norm}_{\mathbf{w}_t}\left[\boldsymbol{\mu}_+ + \mathbf{K}(\mathbf{x}_t - \boldsymbol{\mu}_m - \mathbf{\Phi}\boldsymbol{\mu}_+), (\mathbf{I} - \mathbf{K}\mathbf{\Phi})\mathbf{\Sigma}_+\right]. \tag{19.14}$$

19.2.3 Inference summary

Developing the inference equations was rather long-winded, so here we summarize the inference process in the Kalman filter (see also figure 19.3). We aim to compute the marginal posterior probability $Pr(\mathbf{w}_t|\mathbf{x}_{1\ldots t})$ based on a normally distributed

estimate of the marginal posterior probability $Pr(\mathbf{w}_{t-1}|\mathbf{x}_{1...t-1})$ at the previous time and a new measurement $\mathbf{x}_t$. If the posterior probability at time $t-1$ has mean $\boldsymbol{\mu}_{t-1}$ and variance $\boldsymbol{\Sigma}_{t-1}$, then the Kalman filter updates take the form

$$
\begin{aligned}
\text{State Prediction:} \quad & \boldsymbol{\mu}_+ = \boldsymbol{\mu}_p + \boldsymbol{\Psi}\boldsymbol{\mu}_{t-1} && (19.15)\\
\text{Covariance Prediction:} \quad & \boldsymbol{\Sigma}_+ = \boldsymbol{\Sigma}_p + \boldsymbol{\Psi}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Psi}^T \\
\text{State Update:} \quad & \boldsymbol{\mu}_t = \boldsymbol{\mu}_+ + \mathbf{K}(\mathbf{x}_t - \boldsymbol{\mu}_m - \boldsymbol{\Phi}\boldsymbol{\mu}_+) \\
\text{Covariance Update:} \quad & \boldsymbol{\Sigma}_t = (\mathbf{I} - \mathbf{K}\boldsymbol{\Phi})\boldsymbol{\Sigma}_+,
\end{aligned}
$$

where

$$
\mathbf{K} = \boldsymbol{\Sigma}_+\boldsymbol{\Phi}^T(\boldsymbol{\Sigma}_m + \boldsymbol{\Phi}\boldsymbol{\Sigma}_+\boldsymbol{\Phi}^T)^{-1}. \qquad (19.16)
$$

In the absence of prior information at time $t = 1$, it is usual to initialize the prior mean $\boldsymbol{\mu}_0$ to any reasonable value and the covariance $\boldsymbol{\Sigma}_0$ to a large multiple of the identity matrix representing our lack of knowledge.

19.2.4 Example 1

The Kalman filter update equations are not very intuitive. To help understand the properties of this model, we present two toy examples. In the first case, we consider an object that is approximately circling a central point in two dimensions. The state consists of the two-dimensional position of the object. The actual set of states can be seen in figure 19.4a.

Let us assume that we do not have a good temporal model that describes this motion. Instead, we will assume the simplest possible model. The Brownian motion model assumes that the state at time $t+1$ is similar to the state at time t:

$$
Pr(\mathbf{w}_t|\mathbf{w}_{t-1}) \quad = \quad \text{Norm}_{\mathbf{w}_t}[\mathbf{w}_{t-1}, \sigma_p^2\mathbf{I}]. \qquad (19.17)
$$

This is a special case of the more general formulation in equation 19.6. Notice that this model has no insight into the fact that the object is rotating.

For ease of visualization, we will assume that the observations are just noisy realizations of the true 2D state so that

$$
Pr(\mathbf{x}_t|\mathbf{w}_t) \quad = \quad \text{Norm}_{\mathbf{x}_t}[\mathbf{w}_t, \boldsymbol{\Sigma}_m], \qquad (19.18)
$$

where $\boldsymbol{\Sigma}_m$ is diagonal. This is a special case of the general formulation in 19.8.

In inference, the goal is to estimate the posterior probability over the state at each time step given the observed sequence so far. Figure 19.4b shows the sequence of posterior probabilities $Pr(\mathbf{w}_t|\mathbf{x}_{1...t})$ over the state after running the Kalman recursions. It is now easy to understand why this is referred to as the Kalman *filter*: the MAP states (the peaks of the marginal posteriors) are smoothed relative to estimates from the measurements alone. Notice that the Kalman filter estimate has a lower covariance than the estimate due to the measurements alone
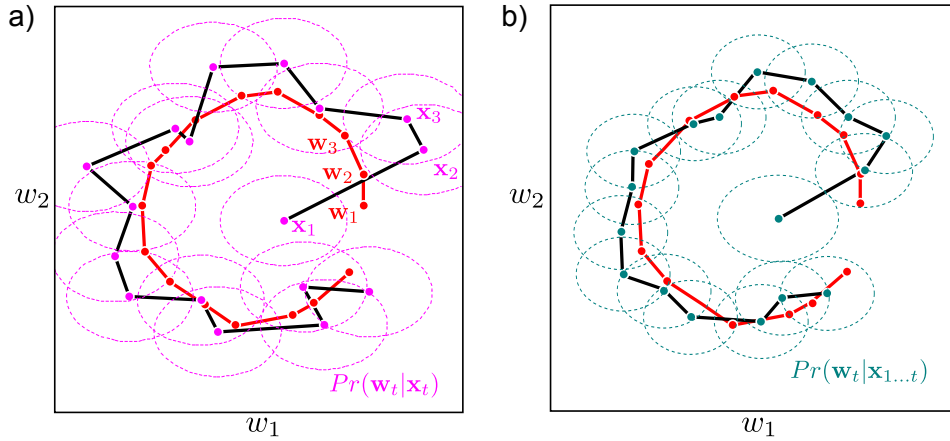
Figure 19.4 Kalman filter example 1. The true state (red dots) evolves in a 2D circle. The measurements (magenta dots) consist of direct noisy observations of the state. a) Posterior probability $Pr(\mathbf{w}_t|\mathbf{x}_t)$ of state using measurements alone (uniform prior assumed). The mean of each distribution is centered on the associated data point, and the covariance depends on the measurement noise. b) Posterior probabilities based on Kalman filter. Here the temporal equation describes the state as a random perturbation of the previous state. Although this is not a good model (the circular movement is not accounted for), the posterior covariances (blue ellipses) are smaller than without the Kalman filter (magenta ellipses in (a)) and the posterior means (blue dots) are closer to the true states (red dots) than the estimates due to the measurements alone (magenta dots in (a)).

(figure 19.4a). In this example, the inclusion of the temporal model makes the estimates both more accurate and more certain, even though the particular choice of temporal model is wrong.

19.2.5 Example 2

Figure 19.5 shows a second example, where the setup is the same except that the observation equation differs at alternate time steps. At even time steps, we observe a noisy estimate of just the first dimension of the state $\mathbf{w}$. At odd time steps, we observe only a noisy estimate of the second dimension. The measurement equations are

$$
\begin{aligned}
Pr(x_t|\mathbf{w}_t) &= \text{Norm}_{x_t}\left[\begin{bmatrix} 1 & 0 \end{bmatrix}\mathbf{w}_t, \sigma_m^2\right], && \text{for } t = 1, 3, 5 \ldots \\
Pr(x_t|\mathbf{w}_t) &= \text{Norm}_{x_t}\left[\begin{bmatrix} 0 & 1 \end{bmatrix}\mathbf{w}_t, \sigma_m^2\right], && \text{for } t = 2, 4, 6 \ldots . \quad (19.19)
\end{aligned}
$$

This is an example of a *non-stationary model*: the model changes over time.

Figure 19.5 Kalman filter example 2. As for example 1, the true state (red points) proceeds in an approximate circle. We don't have access to this temporal model and just assume that the new state is a perturbation of the previous state. At even time steps, the measurement is a noisy realization of the first dimension (horizontal coordinate) of the state. At odd time steps, it is a noisy realization of the second dimension (vertical coordinate). The Kalman filter produces plausible estimates (blue and cyan points and ellipses) even though a single measurement is being used to determine a 2D state at any time step.
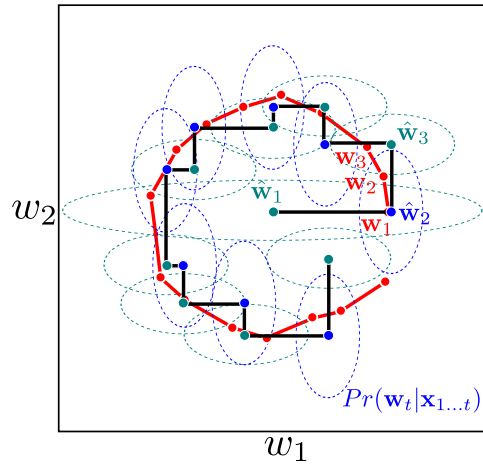
We use the same set of ground truth 2D states as for example 1, and simulate the associated measurements using equation 19.19. The posterior probability over the state was computed using the Kalman filter recursions with the relevant measurement matrix $\boldsymbol{\Phi}$ at each time step.

The results (figure 19.5) show that the Kalman filter maintains a good estimate of the 2D state, despite only having a 1D measurement at each time-step. At odd time steps, the variance in the first dimension decreases (due to the information from the measurement) but the variance in the second dimension increases (due to the uncertainty in the temporal model). At even time steps, the opposite occurs.

This model may seem esoteric, but it is common in many imaging modalities for different aspects of the measurements to arrive at different times. One option is to wait until a full set of measurements is present before estimating the state, but this means that many of them are out of date. Incorporation of each measurement when it arrives using a Kalman filter is a superior approach.

19.2.6 Smoothing

The preceding inference procedure is designed for real-time applications where the estimation of the state is based only on measurements from the past and present. However, there may be occasions where we wish to infer the state using measurements that lie in the future. In the parlance of Kalman filters, this is known as *smoothing*.

We consider two cases. The *fixed lag smoother* is still intended for on-line estimation, but it delays the decision about the state by a fixed number of time steps. The *fixed interval smoother* assumes that we have observed all of the measurements in the entire sequence before we make a judgment about the world state.

Fixed lag smoother

The fixed lag smoother relies on a simple trick. To estimate the state delayed by τ time steps, we *augment* the state vector to contain the delayed estimates of the previous τ times. The time update equation now takes the form

$$
\begin{bmatrix} \mathbf{w}_t \\ \mathbf{w}_t^{[1]} \\ \mathbf{w}_t^{[2]} \\ \vdots \\ \mathbf{w}_t^{[\tau]} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Psi} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w}_{t-1} \\ \mathbf{w}_{t-1}^{[1]} \\ \mathbf{w}_{t-1}^{[2]} \\ \vdots \\ \mathbf{w}_{t-1}^{[\tau]} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_p \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0,} \end{bmatrix}, \tag{19.20}
$$

where the notation $\mathbf{w}_t^{[m]}$ refers to the state at time $t-m$ based on measurements up to time t, so $\mathbf{w}_t^{[\tau]}$ is the quantity we wish to estimate. This state evolution equation clearly fits the general form of the Kalman filter (equation 19.5). It can be parsed as follows: in the first equation of this system, the temporal model is applied to the current estimate of the state. In the remaining equations, the delayed states are created by simply copying the previous state.

The associated measurement equation is

$$
\mathbf{x}_t = \begin{bmatrix} \boldsymbol{\Phi} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w}_t \\ \mathbf{w}_t^{[1]} \\ \mathbf{w}_t^{[2]} \\ \vdots \\ \mathbf{w}_t^{[\tau]}, \end{bmatrix} + \boldsymbol{\epsilon}_m, \tag{19.21}
$$

which fits the form of the Kalman filter measurement model (equation 19.7). The current measurements are based on the current state, and the time-delayed versions play no part. Applying the Kalman recursions with these equations computes not just the current estimate of the state, but also the time-delayed estimates.

Fixed interval smoother

The fixed interval smoother consists of a backward set of recursions that estimate the marginal posterior distributions $Pr(\mathbf{w}_t|\mathbf{x}_{1\dots T})$ of the state at each time step, taking into account all of the measurements $\mathbf{x}_{1\dots T}$. In these recursions, the marginal posterior distribution $Pr(\mathbf{w}_t|\mathbf{x}_{1\dots T})$ of the state at time t is updated, and, based on this result, the marginal posterior $Pr(\mathbf{w}_{t-1}|\mathbf{x}_{1\dots T})$ at time $t-1$ is updated and so on. We denote the mean and variance of the marginal posterior $Pr(\mathbf{w}_t|\mathbf{x}_{1\dots T})$ at time t by $\boldsymbol{\mu}_{t|T}$ and $\boldsymbol{\Sigma}_{t|T}$, respectively, and use

$$
\begin{aligned}
\boldsymbol{\mu}_{t|T} &= \boldsymbol{\mu}_t + \mathbf{C}_t(\boldsymbol{\mu}_{t+1|T} - \boldsymbol{\mu}_{+|t}) \\
\boldsymbol{\Sigma}_{t|T} &= \boldsymbol{\Sigma}_t + \mathbf{C}_t(\boldsymbol{\Sigma}_{t+1|T} - \boldsymbol{\Sigma}_{+|t})\mathbf{C}_t^T,
\end{aligned} \tag{19.22}
$$

where the notation $\boldsymbol{\mu}_{+|t}$ and $\boldsymbol{\Sigma}_{+|t}$ denotes the mean and variance of the posterior distribution $Pr(\mathbf{w}_t|\mathbf{x}_{1\dots t-1})$ of the state at time t based on the measurements up

Figure 19.6 Fixed lag Kalman smoothing. The estimated state (light blue dots) and the associated covariance (light blue ellipses) were estimated using a lag of 1 time step: each estimate is based on all of the data in the past, the present observation, and the observation one step into the future. The resulting estimated states have a smaller variance than for the standard Kalman filter (compare to figure 19.4b) and are closer to the true state (red dots). In effect, the estimate averages out the noise in the measurements (dark blue dots). The cost of this improvement is that there is a delay of one time step.
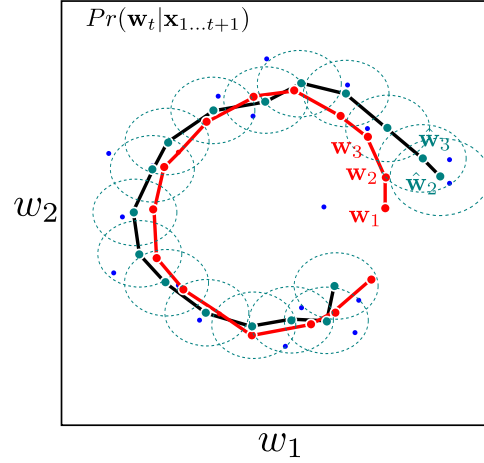
to time $t-1$ (i.e., what we denoted as $\boldsymbol{\mu}_+$ and $\boldsymbol{\Sigma}_+$ during the forward Kalman filter recursions) and

$$\mathbf{C}_t = \boldsymbol{\Sigma}_{t|t}\boldsymbol{\Psi}^T\boldsymbol{\Sigma}_{+|t}^{-1}. \tag{19.23}$$

It can be shown that the backward recursions correspond to the backward pass of sum-product belief propagation in the Kalman filter graphical model.

19.2.7   Temporal and measurement models

The choice of temporal model in the Kalman filter is restricted to be linear and is dictated by the matrix $\boldsymbol{\Psi}$. Despite this limitation, it is possible to build a surprisingly versatile family of models within this framework. In this section we review a few of the best known examples.

1. *Brownian motion:* The simplest model is Brownian motion (figure 19.7a) in which the state is operated upon by the identity matrix so that

$$\mathbf{w}_t = \mathbf{w}_{t-1} + \boldsymbol{\epsilon}_p. \tag{19.24}$$

2. *Geometric transformations:* The family of linear filters includes geometric transformations such as rotations, stretches and shears. For example, choosing $\boldsymbol{\Psi}$ so that $\boldsymbol{\Psi}^T\boldsymbol{\Psi} = \mathbf{I}$ and $|\boldsymbol{\Psi}| = 1$ creates a rotation around the origin (figure 19.7b): this is the true temporal model in figures 19.4-19.6.

3. *Velocities and accelerations:* The Brownian motion model can be extended by adding a constant velocity $\mathbf{v}$ to the motion model (figure 19.7c)so that

$$\mathbf{w}_t = \mathbf{v} + \mathbf{w}_{t-1} + \boldsymbol{\epsilon}_p. \tag{19.25}$$
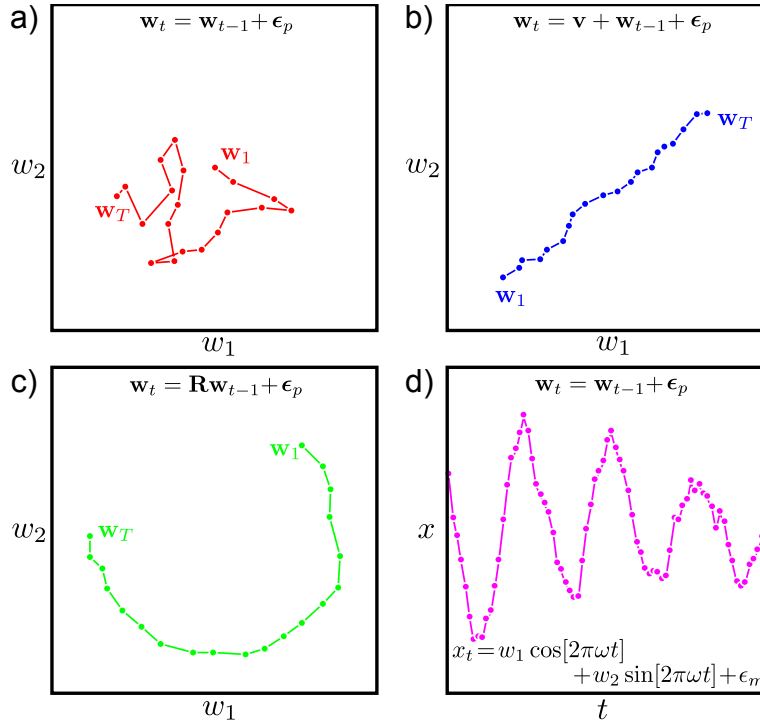
Figure 19.7 Temporal and measurement models. a) Brownian motion model. At each time step the state is randomly perturbed from the previous position, so a drunken walk through state space occurs. b) Constant velocity model. At each time step, a constant velocity is applied in addition to the noise. c) Transformation model. At each time step the state is rotated about the origin and noise is added. d) Oscillatory measurements. These quasi-sinusoidal measurements are created from a two-dimensional state that itself undergoes Brownian motion. The two elements of the state control the sinusoidal and co-sinusoidal parts of the measurements.

To incorporate a changing velocity, we can enhance the state vector to include the velocity term so that

$$\begin{bmatrix} \mathbf{w}_t \\ \dot{\mathbf{w}}_t \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{w}_{t-1} \\ \dot{\mathbf{w}}_{t-1} \end{bmatrix} + \boldsymbol{\epsilon}_p. \tag{19.26}$$

This has the natural interpretation that the velocity term $\dot{\mathbf{w}}$ contributes an offset to the state at each time. However, the velocity is itself uncertain and is related by a Brownian motion model to the velocity at the previous time step. For the measurement equation, we have

$$\mathbf{x}_t = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{w}_t \\ \dot{\mathbf{w}}_t \end{bmatrix} + \boldsymbol{\epsilon}_m. \tag{19.27}$$

In other words, the measurements at time t do not directly depend on the velocity term, only the state itself. This idea can be easily extended to include an acceleration term as well.

4. *Oscillatory data:* Some data are naturally oscillatory. To describe oscillatory 1D data, we could use a state containing a 2×1 state vector $\mathbf{w}$ using Brownian motion as the temporal model. We then implement the non-stationary measurement equation

$$x_t = \begin{bmatrix} \cos[2\pi\omega t] & \sin[2\pi\omega t] \end{bmatrix} \mathbf{w}_t + \epsilon_m. \tag{19.28}$$

For a fixed state $\mathbf{w}$, this model produces noisy sinusoidal data. As the state varies due to the Brownian motion of the temporal model, the phase and amplitude of the quasi-sinusoidal output will change (19.7d).

19.2.8 Problems with the Kalman filter

Although the Kalman filter is a flexible tool, it has a number of shortcomings (figure 19.8). Most notably,

- it requires the temporal and measurement equations to be linear, and
- it assumes that the marginal posterior is unimodal and can be well captured by a mean and covariance; hence, it can only ever have one hypothesis about the position of the object.

In the following sections, we discuss models that address these problems. The *extended Kalman filter* and the *unscented Kalman filter* both allow nonlinear state update and measurement equations. *Particle filtering* abandons the use of the normal distribution and describes the state as a complex multi-modal distribution.

19.3 Extended Kalman filter

The extended Kalman filter (EKF) is designed to cope with more general temporal models, where the relationship between the states at time t is an arbitrary nonlinear function $\mathbf{f}[\bullet, \bullet]$ of the state at the previous time step and a stochastic contribution $\boldsymbol{\epsilon}_p$

$$\mathbf{w}_t = \mathbf{f}[\mathbf{w}_{t-1}, \boldsymbol{\epsilon}_p], \tag{19.29}$$

where the covariance of the noise term $\boldsymbol{\epsilon}_p$ is $\boldsymbol{\Sigma}_p$ as before. Similarly, it can cope with a nonlinear relationship $\mathbf{g}[\bullet, \bullet]$ between the state and the measurements

$$\mathbf{x}_t = \mathbf{g}[\mathbf{w}_t, \boldsymbol{\epsilon}_m], \tag{19.30}$$

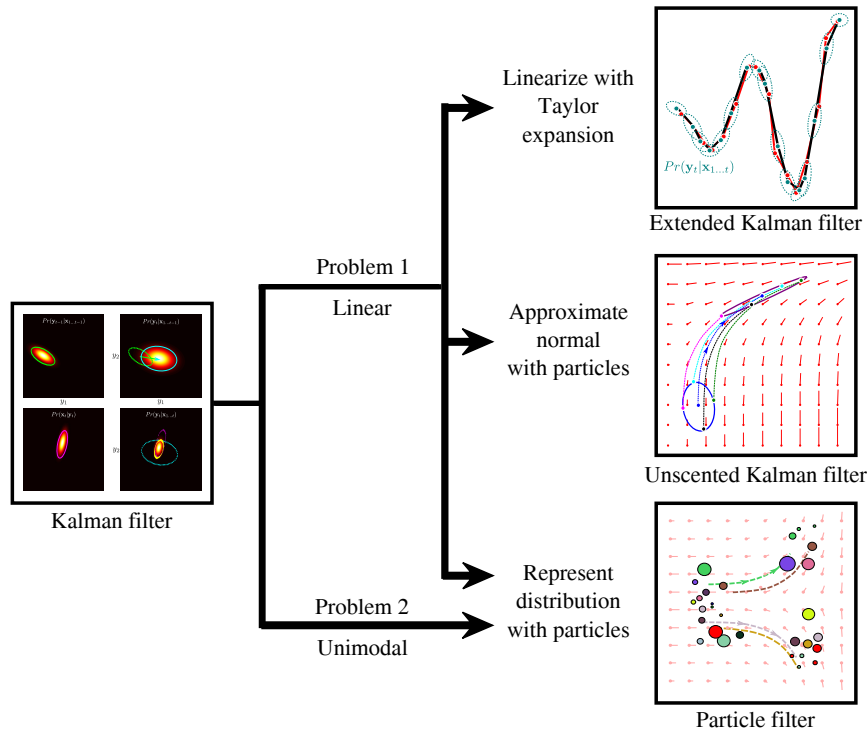where the covariance of $\boldsymbol{\epsilon}_m$ is $\boldsymbol{\Sigma}_m$.

Figure 19.8 Temporal models for tracking. The Kalman filter has two main problems: first, it requires the temporal and measurement models to be linear. This problem is directly addressed by the extended and unscented Kalman filters. Second, it represents the uncertainty with a normal distribution that is unimodal, and so cannot maintain multiple hypotheses about the state. This problem is tackled by particle filters.

The extended Kalman filter works by taking linear approximations to the nonlinear functions at the peak $\boldsymbol{\mu}_t$ of the current estimate using the Taylor expansion. If the function is not too nonlinear, then this approximation will adequately represent the function in the region of the current estimate and we can proceed as usual. We define the Jacobian matrices,

$$\Psi = \left.\frac{\partial \mathbf{f}[\mathbf{w}_{t-1}, \boldsymbol{\epsilon}_p]}{\partial \mathbf{w}_{t-1}}\right|_{\boldsymbol{\mu}_{t-1}, \mathbf{0}}$$

$$\boldsymbol{\Upsilon}_p = \left.\frac{\partial \mathbf{f}[\mathbf{w}_{t-1}, \boldsymbol{\epsilon}_p]}{\partial \boldsymbol{\epsilon}_p}\right|_{\boldsymbol{\mu}_{t-1}, \mathbf{0}}$$

$$\boldsymbol{\Phi} = \left.\frac{\partial \mathbf{g}[\mathbf{w}_t, \boldsymbol{\epsilon}_m]}{\partial \mathbf{w}_t}\right|_{\boldsymbol{\mu}_+, \mathbf{0}}$$

$$\boldsymbol{\Upsilon}_m = \left.\frac{\partial \mathbf{g}[\mathbf{w}_t, \boldsymbol{\epsilon}_m]}{\partial \boldsymbol{\epsilon}_m}\right|_{\boldsymbol{\mu}_+, \mathbf{0}}, \tag{19.31}$$

where the notation $|_{\boldsymbol{\mu}_+, \mathbf{0}}$ denotes that the derivative is computed at position $\mathbf{w} = \boldsymbol{\mu}_+$ and $\boldsymbol{\epsilon} = \mathbf{0}$. Note that we have overloaded the meaning of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ here. Previously they represented the linear transforms between states at adjacent times, and between the state and the measurements, respectively. Now they represent the local linear approximation of the nonlinear functions relating these quantities.

The update equations for the extended Kalman filter are

$$
\begin{aligned}
\text{State Prediction:} \qquad & \boldsymbol{\mu}_+ = \mathbf{f}[\boldsymbol{\mu}_{t-1}, \mathbf{0}] \\
\text{Covariance Prediction:} \qquad & \boldsymbol{\Sigma}_+ = \boldsymbol{\Psi}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Psi}^T + \boldsymbol{\Upsilon}_p\boldsymbol{\Sigma}_p\boldsymbol{\Upsilon}_p^T \\
\text{State Update:} \qquad & \boldsymbol{\mu}_t = \boldsymbol{\mu}_+ + \mathbf{K}(\mathbf{x}_t - \mathbf{g}[\boldsymbol{\mu}_+, \mathbf{0}]) \\
\text{Covariance Update:} \qquad & \boldsymbol{\Sigma}_t = (\mathbf{I} - \mathbf{K}\boldsymbol{\Phi})\boldsymbol{\Sigma}_+,
\end{aligned} \tag{19.32}
$$

where

$$\mathbf{K} = \boldsymbol{\Sigma}_+\boldsymbol{\Phi}^T(\boldsymbol{\Upsilon}_m\boldsymbol{\Sigma}_m\boldsymbol{\Upsilon}_m^T + \boldsymbol{\Phi}\boldsymbol{\Sigma}_+\boldsymbol{\Phi}^T)^{-1}. \tag{19.33}$$

In the context of fixed interval smoothing, the results can be improved by performing several passes back and forth through the data, re-linearizing around the previous estimates of the state in each sweep. This is called the *iterated extended Kalman filter*.

Algorithm 19.4

In conclusion, the extended Kalman filter is conceptually simple, but only copes with relatively benign nonlinearities. It is a heuristic solution to the nonlinear tracking problem and may diverge from the true solution.

19.3.1 Example

Figure 19.9 shows an worked example of the extended Kalman filter. In this case, the time update model is nonlinear, but the observation model is still linear:

$$
\begin{aligned}
\mathbf{w}_t &= \mathbf{f}[\mathbf{w}_{t-1}, \boldsymbol{\epsilon}_p] \\
\mathbf{x}_t &= \mathbf{w} + \boldsymbol{\epsilon}_m,
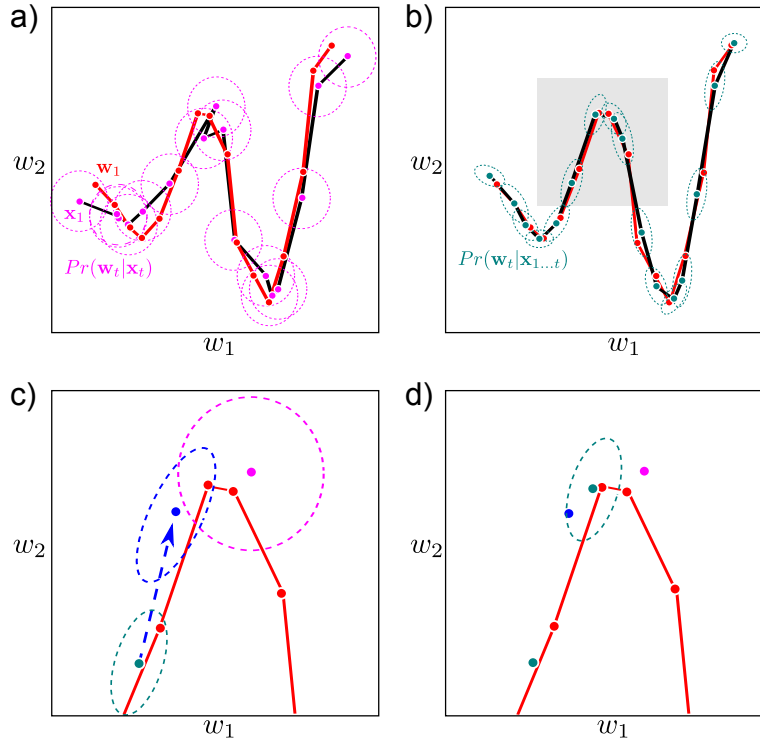\end{aligned} \tag{19.34}
$$

Figure 19.9 The extended Kalman filter (EKF). a) The temporal model that transforms the state (red dots) here is nonlinear. Each observed data point (magnenta dots) is a noisy copy of the state at that time. Based on the data alone, the posterior probability over the state is given by the magenta ellipses. b) Estimated marginal posterior distributions using the extended Kalman filter: the estimates are more certain and more accurate than using the measurements alone. c) Close up of shaded region from (b). The EKF makes a normal prediction (dark blue ellipse) based on the previous state (light blue ellipse) and the linearized motion model. The measurement makes a different prediction (magenta ellipse). d) The EKF combines these two predictions to get an improved estimate (light blue ellipse).

where $\boldsymbol{\epsilon}_p$ is a normal noise term with covariance $\boldsymbol{\Sigma}_p$, $\boldsymbol{\epsilon}_m$ is a normal noise term with covariance $\boldsymbol{\Sigma}_m$, and the nonlinear function $\mathbf{f}[\bullet, \bullet]$ is given by

$$\mathbf{f}[\mathbf{w}, \boldsymbol{\epsilon}_p] = \begin{bmatrix} w_1 \\ w_1 \sin[w_1] + \boldsymbol{\epsilon}_p \end{bmatrix}. \tag{19.35}$$

The Jacobian matrices for this model are easily computed. The matrices $\boldsymbol{\Upsilon}_p, \boldsymbol{\Phi}$, and $\boldsymbol{\Upsilon}_m$ are all equal to the identity. The Jacobian matrix $\boldsymbol{\Psi}$ is

$$\boldsymbol{\Psi} = \begin{bmatrix} 1 & 0 \\ \sin[w_1] + w_1 \cos[w_1] & 0 \end{bmatrix}. \tag{19.36}$$

Figure 19.10 Problems with the EKF.
A 2D temporal function $\mathbf{f}[\bullet, \bullet]$ oper-
ates on the state (red lines indicate
gradient direction). In the prediction
step, the previous estimate (blue el-
lipse) should be passed through this
function (to create distorted green el-
lipse) and noise added (not shown).
The EKF approximates this by pass-
ing the mean through the function
and updating the covariance based on
a linear approximation of the func-
tion at the previous state estimate.
Here, the function changes quite non-
linearly, and so this approximation
is bad, and the predicted covariance
(magenta ellipse) is inaccurate.
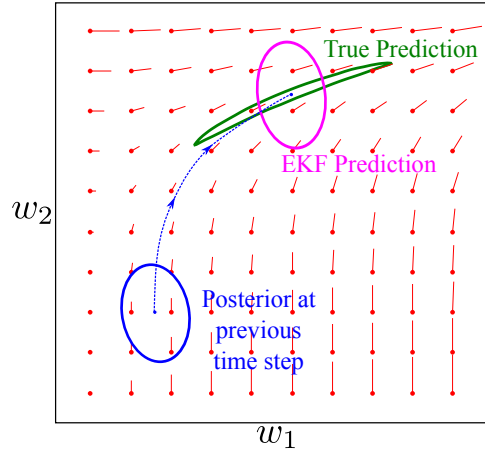
The results of using this system for tracking are illustrated in figure 19.9. The
extended Kalman filter successfully tracks this nonlinear model giving results that
are both closer to the true state and more confident than estimates based on the
individual measurements alone. The EKF works well here because the nonlinear
function is smooth and departs from linearity slowly relative to the time steps:
consequently, the linear approximation is reasonable.

19.4   Unscented Kalman filter

The extended Kalman filter is only reliable if the linear approximations to the
functions $\mathbf{f}[\bullet, \bullet]$ or $\mathbf{g}[\bullet, \bullet]$ describe them well in the region of the current position.
Figure 19.10 shows a situation where the local properties of the temporal function
are not representative and so the EKF estimate of the covariance is inaccurate.

Algorithm 19.5

The unscented Kalman filter (UKF) is a derivative-free approach that partially
circumvents this problem. It is suited to nonlinear models with additive, normally
distributed noise, so that the temporal and measurement equations are

$$
\begin{aligned}
\mathbf{w}_t &= \mathbf{f}[\mathbf{w}_{t-1}] + \boldsymbol{\epsilon}_p \\
\mathbf{x}_t &= \mathbf{g}[\mathbf{w}_t] + \boldsymbol{\epsilon}_m.
\end{aligned}
\tag{19.37}
$$

As such it is slightly more restricted in its applicability compared to the extended
Kalman filter. However, it can cope with more extreme nonlinearities in the state
evolution and measurement equations than the EKF.

To understand how the UKF works, consider a nonlinear temporal model.
Given the mean and covariance of the normally distributed posterior distribution
$Pr(\mathbf{w}_{t-1}|\mathbf{x}_{1...t-1})$ over the state at time $t-1$, we wish to predict a normally dis-
tributed prior $Pr(\mathbf{w}_t|\mathbf{x}_{1...t-1})$ over the state at time t. We proceed by

- approximating the normally distributed posterior $Pr(\mathbf{w}_{t-1}|\mathbf{x}_{1...t-1})$ with a
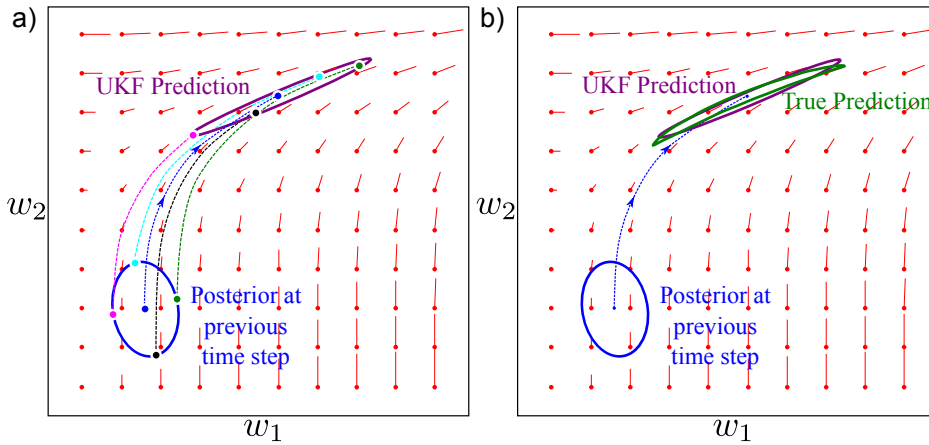
Figure 19.11 Temporal update for unscented Kalman filter. a) In the temporal update step, the posterior distribution from the previous time step (which is a normal) is approximated by a weighted set of point masses that collectively have the same mean and covariance as this normal. Each of these is passed through the temporal model (dashed lines). A normal prediction is formed by estimating the mean and the covariance of the transformed point masses. In a real system, the variance would subsequently be inflated to account for additive uncertainty in the position. b) The resulting prediction is quite close to the true prediction and for this case is much better than that of the extended Kalman filter (see figure 19.10).

set of point masses that are deterministically chosen so that they have the same mean $\boldsymbol{\mu}_{t-1}$ and covariance $\boldsymbol{\Sigma}_{t-1}$ as the original distribution,

- passing each of the point masses through the nonlinear function, and then

- setting the predicted distribution $Pr(\mathbf{w}_t|\mathbf{x}_{1...t-1})$ at time t to be a normal distribution with the mean and covariance of the transformed point masses.

This process is illustrated in figure 19.11.

As for the extended Kalman filter, the predicted state $Pr(\mathbf{w}_t|\mathbf{x}_{1...t-1})$ in the UKF is a normal distribution. However, this normal distribution is a provably better approximation to the true distribution than that provided by the EKF. A similar approach is applied to cope with a nonlinearity in the measurement equations. We will now consider each of these steps in more detail.

19.4.1   State evolution

As for the standard Kalman filter, the goal of the state evolution step is to form a prediction $Pr(\mathbf{w}_t|\mathbf{x}_{1...t-1})$ about the state at time t by applying the temporal model to the posterior distribution $Pr(\mathbf{w}_{t-1}|\mathbf{x}_{1...t-1})$ at the previous time step.

This marginal posterior distribution is normal with mean $\boldsymbol{\mu}_{t-1}$ and covariance $\boldsymbol{\Sigma}_{t-1}$, and the prediction will also be normal with mean $\boldsymbol{\mu}_+$ and covariance $\boldsymbol{\Sigma}_+$.

We proceed as follows. We approximate the posterior distribution at the previous time step by a weighted sum of $2D_{\mathbf{w}}+1$ delta functions, where $D_{\mathbf{w}}$ is the dimensionality of the state, so that

$$
\begin{aligned}
Pr(\mathbf{w}_{t-1}|\mathbf{x}_{1\ldots t-1}) &= \text{Norm}_{\mathbf{w}_{t-1}}[\boldsymbol{\mu}_{t-1}, \boldsymbol{\Sigma}_{t-1}] \\
&\approx \sum_{j=0}^{2D_{\mathbf{w}}} a_j \delta[\mathbf{w}_{t-1} - \hat{\mathbf{w}}^{[j]}],
\end{aligned}
\tag{19.38}
$$

where the weights $\{a_j\}_{j=0}^{2D_{\mathbf{w}}}$ are positive and sum to one. In this context, the delta functions are referred to as *sigma points*. The positions $\{\hat{\mathbf{w}}^{[j]}\}_{j=0}^{2D_{\mathbf{w}}}$ of the sigma points are carefully chosen so that

$$
\begin{aligned}
\boldsymbol{\mu}_{t-1} &= \sum_{j=0}^{2D_{\mathbf{w}}} a_j \hat{\mathbf{w}}^{[j]} \\
\boldsymbol{\Sigma}_{t-1} &= \sum_{j=0}^{2D_{\mathbf{w}}} a_j (\hat{\mathbf{w}}^{[j]} - \boldsymbol{\mu}_{t-1})(\hat{\mathbf{w}}^{[j]} - \boldsymbol{\mu}_{t-1})^T.
\end{aligned}
\tag{19.39}
$$

Problem 19.9

One possible scheme is to choose sigma points

$$
\begin{aligned}
\hat{\mathbf{w}}^{[0]} &= \boldsymbol{\mu}_{t-1} \\
\hat{\mathbf{w}}^{[j]} &= \boldsymbol{\mu}_{t-1} + \sqrt{\frac{D_{\mathbf{w}}}{1-a_0}} \boldsymbol{\Sigma}_{t-1}^{1/2} \mathbf{e}_j && \text{for all } j \in \{1\ldots D_{\mathbf{w}}\} \\
\hat{\mathbf{w}}^{[D_{\mathbf{w}}+j]} &= \boldsymbol{\mu}_{t-1} - \sqrt{\frac{D_{\mathbf{w}}}{1-a_0}} \boldsymbol{\Sigma}_{t-1}^{1/2} \mathbf{e}_j && \text{for all } j \in \{1\ldots D_{\mathbf{w}}\}, \quad (19.40)
\end{aligned}
$$

where $\mathbf{e}_j$ is the unit vector in the jth direction. The associated weights are chosen so that $a_0 \in [0, 1]$ and

$$
a_j = \frac{1-a_0}{2D_{\mathbf{w}}},
\tag{19.41}
$$

where the choice of the weight a_0 of the first sigma point determines how far the remaining sigma points are from the mean.

We pass the sigma points through the nonlinearity, to create a new set of samples $\hat{\mathbf{w}}_+^{[j]} = \mathbf{f}[\hat{\mathbf{w}}^{[j]}]$ that collectively form a prediction for the state. We then compute the mean and the variance of the predicted distribution $Pr(\mathbf{w}_t|\mathbf{x}_{1\ldots t-1})$ from the mean and variance of the transformed points so that

$$\boldsymbol{\mu}_+ \quad = \quad \sum_{j=0}^{2D_{\mathbf{w}}} a_j \hat{\mathbf{w}}_+^{[j]}$$

$$\boldsymbol{\Sigma}_+ \quad = \quad \sum_{j=0}^{2D_{\mathbf{w}}} a_j (\hat{\mathbf{w}}_+^{[j]} - \boldsymbol{\mu}_+)(\hat{\mathbf{w}}_+^{[j]} - \boldsymbol{\mu}_+)^T + \boldsymbol{\Sigma}_p, \qquad (19.42)$$

where we have added an extra term $\boldsymbol{\Sigma}_p$ to the predicted covariance to account for the additive noise in the temporal model.

19.4.2   Measurement incorporation

The measurement incorporation process in the UKF uses a similar idea: we approximate the predicted distribution $Pr(\mathbf{w}_t|\mathbf{x}_{1...t-1})$ as a set of delta functions or sigma points

$$Pr(\mathbf{w}_t|\mathbf{x}_{1...t-1}) \quad = \quad \text{Norm}_{\mathbf{w}_{t-1}}[\boldsymbol{\mu}_+, \boldsymbol{\Sigma}_+]$$

$$\approx \quad \sum_{j=0}^{2D_{\mathbf{w}}} a_j \delta[\mathbf{w}_t - \hat{\mathbf{w}}^{[j]}], \qquad (19.43)$$

where we choose the centers of the sigma points and the weights so that

$$\boldsymbol{\mu}_+ \quad = \quad \sum_{j=0}^{2D_{\mathbf{w}}} a_j \hat{\mathbf{w}}^{[j]}$$

$$\boldsymbol{\Sigma}_+ \quad = \quad \sum_{j=0}^{2D_{\mathbf{w}}} a_j (\hat{\mathbf{w}}^{[j]} - \boldsymbol{\mu}_+)(\hat{\mathbf{w}}^{[j]} - \boldsymbol{\mu}_+)^T. \qquad (19.44)$$

For example, we could use the scheme outlined in equations 19.40 and 19.41.

Then the sigma points are passed through the measurement model $\hat{\mathbf{x}}^{[j]} = \mathbf{g}[\hat{\mathbf{w}}^{[j]}]$ to create a new set of points $\{\hat{\mathbf{x}}^{[j]}\}_{j=0}^{2D_{\mathbf{w}}}$ in the measurement space. We compute the mean and covariance of these predicted measurements using the relations

$$\boldsymbol{\mu}_x \quad = \quad \sum_{j=0}^{2D_{\mathbf{w}}} a_j \hat{\mathbf{x}}^{[j]}$$

$$\boldsymbol{\Sigma}_x \quad = \quad \sum_{j=0}^{2D_{\mathbf{w}}} a_j (\hat{\mathbf{x}}^{[j]} - \boldsymbol{\mu}_x)(\hat{\mathbf{x}}^{[j]} - \boldsymbol{\mu}_x)^T + \boldsymbol{\Sigma}_m \qquad (19.45)$$

.

The measurement incorporation equations are now

$$\begin{aligned} \boldsymbol{\mu}_t &= \boldsymbol{\mu}_+ + \mathbf{K}\left(\mathbf{x}_t - \boldsymbol{\mu}_x\right) \\ \boldsymbol{\Sigma}_t &= \boldsymbol{\Sigma}_+ - \mathbf{K}\boldsymbol{\Sigma}_x\mathbf{K}^T, \end{aligned} \tag{19.46}$$

where the Kalman gain $\mathbf{K}$ is now re-defined as

$$\mathbf{K} = \left(\sum_{j=0}^{2D_\mathbf{w}} a_j(\hat{\mathbf{w}}^{[j]} - \boldsymbol{\mu}_+)^T(\hat{\mathbf{x}}^{[j]} - \boldsymbol{\mu}_x)^T\right)\boldsymbol{\Sigma}_x^{-1}. \tag{19.47}$$

As for the prediction step, it can be shown that the UKF approximation is better than that of the EKF.

19.5   Particle filtering

The extended and unscented Kalman filters can partially cope with nonlinear temporal and measurement models. However, they both represent the uncertainty over the state as a normal distribution. They are hence ill-equipped to deal with situations where the true probability distribution over the state is multi-modal. Figure 19.12 illustrates a temporal model that maps nearby states into two distinct regions. In this situation, neither the EKF nor the UKF suffice: the EKF models only one of the resulting clusters and the UKF tries to model both with a single normal model, assigning a large probability to the empty region between the clusters.

Particle filtering resolves this problem by representing the probability density as a set of particles in the state space. Each particle can be thought of as representing a hypothesis about the possible state. When the state is tightly constrained by the data, all of these particles will lie close to one another. In more ambiguous cases, they will be widely distributed or clustered into groups of competing hypotheses.

The particles can be evolved through time or projected down to simulate measurements regardless of how nonlinear the functions are. The latter exercise leads us to another nice property of the particle filter: since the state is multi-modal, so are the predicted measurements. In turn, the measurement density may also be multi-modal. In vision systems, this means that the system copes much better with clutter in the scene (figure 19.14). As long as some of the predicted measurements agree with the measurement density, the tracker should remain stable.

One of the simplest particle filter methods is the *conditional density propagation* or *condensation* algorithm. The probability distribution $Pr(\mathbf{w}_{t-1}|\mathbf{x}_{1\ldots t-1})$ is represented by a weighted sum of J weighted particles:

Algorithm 19.6

$$Pr(\mathbf{w}_{t-1}|\mathbf{x}_{1\ldots t-1}) = \sum_{j=1}^{J} a_j\delta[\mathbf{w}_{t-1} - \hat{\mathbf{w}}_{t-1}^{[j]}], \tag{19.48}$$

where the weights are positive and sum to one. Each particle represents a hypothesis about the state, and the weight of the particle indicates our confidence in that
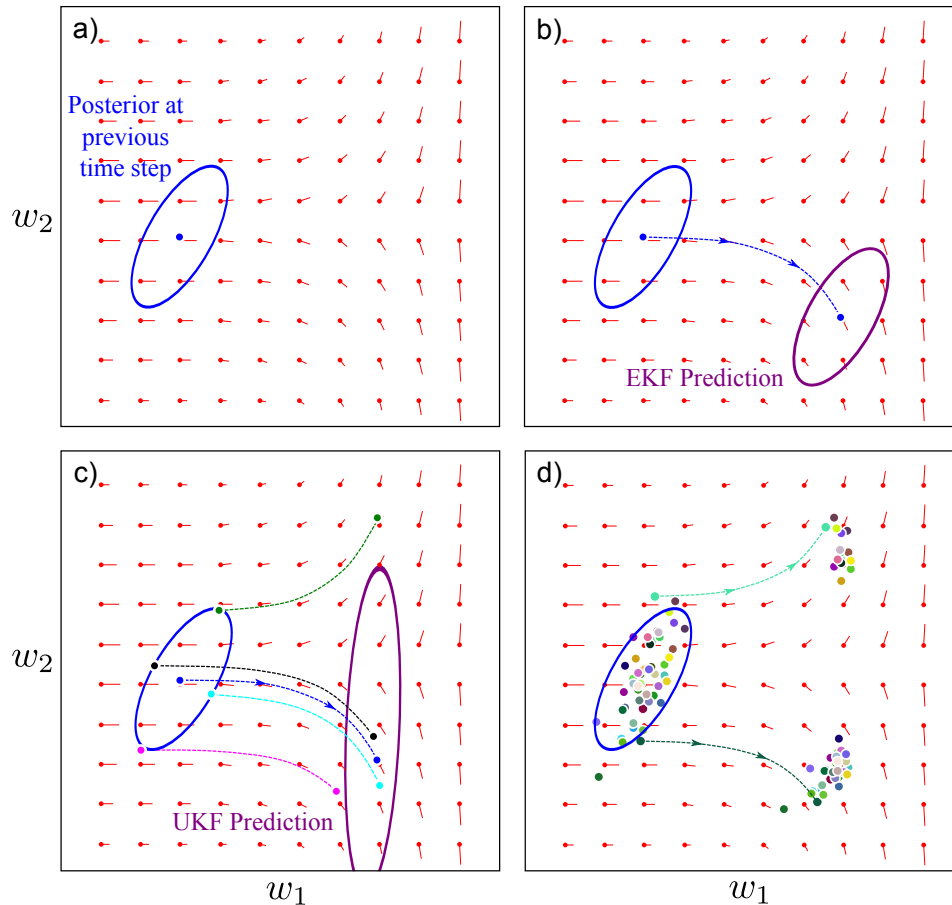
Figure 19.12 The need for particle filtering. a) Consider this new temporal update function (change of state with time indicated by red pointers), which bifurcates around the horizontal midline. b) With the EKF, the predicted distribution after the time update is toward the bottom, as the initial mean was below the bifurcation. The linear approximation is not good here, and so the covariance estimate is inaccurate. c) In the UKF, one of the sigma points that approximates the prior is above the midline, so it is moved upward (green dashed line), whereas the others are below the midline and are moved downward (other dashed lines). The estimated covariance is very large. d) We can get an idea of the true predicted distribution by sampling the posterior at the previous time-step, and passing the samples through the nonlinear temporal model. It is clear that the resulting distribution is bimodal and can never be well approximated with a normal distribution. The particle filter represents the distribution in terms of particles throughout the tracking process, and so it can describe multimodal distributions like this.

hypothesis. Our goal is to compute the probability distribution $Pr(\mathbf{w}_t|\mathbf{x}_{1\ldots t})$ at the next time step, which will be represented in a similar fashion. As usual, this process is divided into a time evolution and a measurement incorporation step, which we now consider in turn.

19.5.1 Time evolution

In the time evolution step, we create J predictions $\hat{\mathbf{w}}_+^{[j]}$ for the time evolved state. Each is represented by an *unweighted particle*. We create each prediction by resampling so that we

- choose an index $n \in \{1\ldots J\}$ of the original weighted particles where the probability is according to the weights. In other words we draw a sample from $\mathrm{Cat}_n[\mathbf{a}]$, and

- draw the sample $\hat{\mathbf{w}}_+^{[j]}$ from the temporal update distribution $Pr(\mathbf{w}_t|\mathbf{w}_{t-1} = \hat{\mathbf{w}}_{t-1}^{[n]})$ (equation 19.37).

In this process, the final unweighted particles $\hat{\mathbf{w}}_+^{[j]}$ are created from the original weighted particles $\hat{\mathbf{w}}_{t-1}^{[j]}$ according to the weights $\mathbf{a} = [a_1, a_2 \ldots a_J]$. Hence, the highest weighted original particles may contribute repeatedly to the final set, and the lowest weighted ones may not contribute at all.

19.5.2 Measurement incorporation

In the measurement incorporation step we weight the new set of particles according to how well they agree with the observed data. To this end, we

- pass the particles through the measurement model $\hat{\mathbf{x}}_+^{[j]} = \mathbf{g}[\hat{\mathbf{w}}_+^{[j]}]$.

- weight the particles according to their agreement with the observation density. For example, with a Gaussian measurement model, we could use

$$a_j \propto \mathrm{Norm}_{\mathbf{x}_t}[\hat{\mathbf{x}}_+^{[j]}, \boldsymbol{\Sigma}_m]. \tag{19.49}$$

- normalize the resulting weights $\{a_j\}_{j=1}^J$ to sum to one.

- Finally, we set the new states $\hat{\mathbf{w}}_t^{[j]}$ to the predicted states $\hat{\mathbf{w}}_+^{[j]}$ and the new weights to a_j.

Figure 19.13 demonstrates the action of the particle filter. The filter copes elegantly with the multi-modal probability distribution. It is also suitable for situations where it is not obvious which aspect of the data is the true measurement. This is known as the *data association* problem (figure 19.14).

The main disadvantage of particle filters is their cost: in high dimensions, a very large number of particles may be required to get an accurate representation of the true distribution over the state.
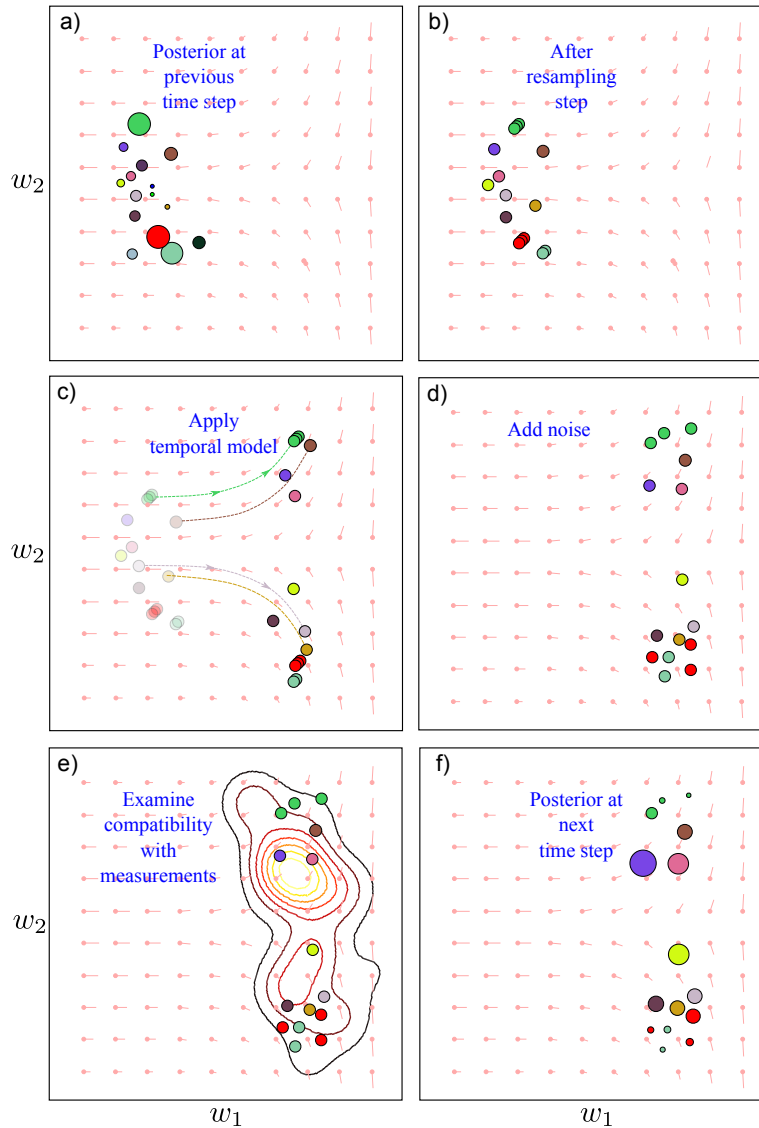
Figure 19.13 The condensation algorithm. a) The posterior at the previous step is represented as a set of weighted particles. b) The particles are re-sampled according to their weights to produce a new set of unweighted particles. c) These particles are passed through the nonlinear temporal function. d) Noise is added according to the temporal model. e) The particles are passed through the measurement model and compared to the measurement density. f) The particles are re-weighted according to their compatibility with the measurements, and the process can begin again.
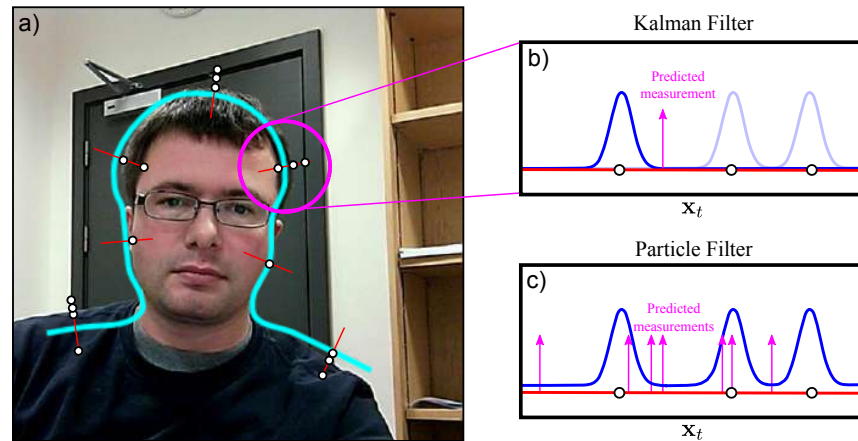
Figure 19.14 Data association. a) Consider the problem of tracking this
contour. The contour from the previous frame is shown in cyan, and the
problem is to incorporate the measurements in this frame. These measure-
ments take the form of edges along one-dimensional slices perpendicular to
the contour. In most cases, there are multiple possible edges that might
be due to the contour. b) In the Kalman filter, the measurement density
is constrained to be normal: we are forced to choose between the different
possible hypotheses. A sensible way to do this is to select the one that is
closest to the predicted measurement. This is known as data association.
c) In the particle filter, there are multiple predicted measurements and it is
not necessary that the measurement density is normal. We can effectively
take into account all of the possible measurements.

19.5.3   Extensions

There are many variations and extensions to the particle filter. In many schemes,
the particles are not re-sampled on each iteration, but only occasionally so the
main scheme operates with weighted particles. The re-sampling process can also be
improved by applying *importance sampling*. Here, the new samples are generated
in concert with the measurement process, so that particles that agree with the
measurements are more likely to be produced. This helps prevent the situation
where none of the unweighted samples after the prediction stage agree with the
measurements.

The process of *Rao-Blackwellization* partitions the state into two subsets of
variables. The first subset is tracked using a particle filter, but the other subset is
conditioned on the first subset and evaluated analytically using a process more like
the Kalman filter. Careful choice of these subsets can result in a tracking algorithm
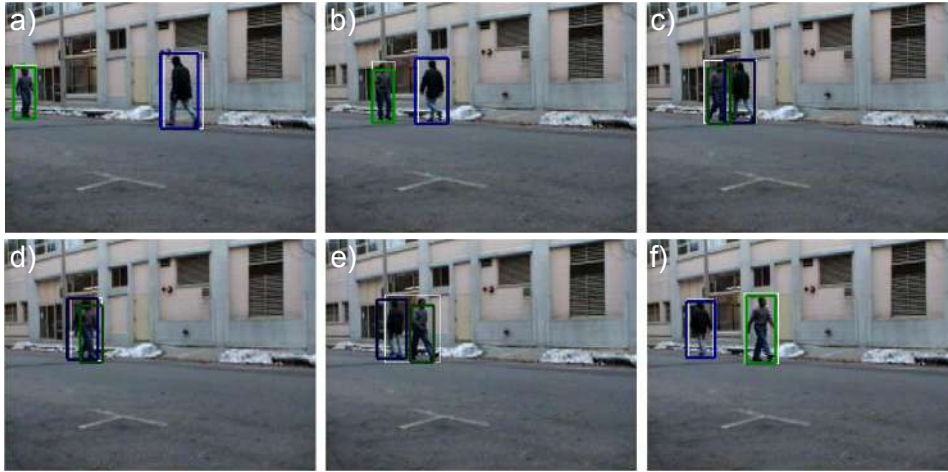that is both efficient and accurate.

Figure 19.15 Pedestrian tracking results. a-f) Six frames from a sequence in which a 3D bounding plane of constant depth was used to track each person. Adapted from Rosales & Sclaroff (1999). ©1999 IEEE.

19.6   Applications

Tracking algorithms can be used in combination with any model that is being re-estimated over a time sequence. For example, they are often used in combination with 3D body models to track the pose of a person in video footage. In this section, we will describe three example applications. First we will consider tracking the 3D position of a pedestrian in a scene. Second, we will describe *simultaneous localization and mapping (SLAM)* in which both a 3D representation of a scene and the pose of the camera viewing it are estimated through a time sequence. Finally, we will consider contour tracking for an object with complex motion against a cluttered background.

19.6.1   Pedestrian Tracking

Rosales & Sclaroff (1999) described a system to track pedestrians in 2D from a fixed static camera using the extended Kalman filter. For each frame, the measurements $\mathbf{x}$ consisted of a 2D bounding box $\mathbf{x} = \{x_1, y_1, x_2, y_2\}$ around the pedestrian. This was found by segmenting the pedestrian from the background using a background subtraction model, and finding a connected region of foreground pixels.

The state of the world $\mathbf{w}$ was considered to be the 3D position of a bounding box of constant depth $\{u_1, v_1, u_2, v_2, w\}$ and the velocities associated with these five quantities. The state update equation assumed first-order Newtonian dynamics in 3D space. The relationship between the state and measurements was

$$\begin{bmatrix} x_1 \\ y_1 \\ x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \end{bmatrix} \frac{1}{1+w} + \boldsymbol{\epsilon}, \tag{19.50}$$

which mimics the nonlinear action of the pinhole camera.

Results from the tracking procedure are illustrated in figure 19.15. The system successfully tracks pedestrians and was extended to cope with occlusions (which are predicted by the EKF based on the estimated trajectory of the objects). In the full system, the estimated bounding boxes are used to warp the images of the pedestrians to a fixed size, and the resulting registered images were used as input to an action recognition algorithm

19.6.2   Monocular SLAM

The goal of *simultaneous localization and mapping* or *SLAM* is to build a map of an unknown environment based on measurements from a moving sensor (usually attached to a robot), and to establish where the sensor is in the world at any given time. Hence, the system must simultaneously estimate the structure of the world and its own position and orientation within that model. SLAM was originally based on range sensors, but in the last decade it has become possible to build systems that work in real-time based on a monocular camera. In essence, it is a real-time version of the sparse 3D reconstruction algorithms discussed in chapter 16.

Davison *et al.* (2007) presented one such system which was based on the extended Kalman filter. Here the world state $\mathbf{w}$ contains

- the camera position (u, v, w),
- the orientation of the camera as represented by a 4D quaternion $\mathbf{q}$,
- the velocity and angular velocity, and
- a set of 3D points $\{\mathbf{p}_k\}$ where $\mathbf{p}_k = [p_{uk}, p_{vk}, p_{wk}]$. This set generally expands during the sequence as more parts of the world are seen.

The state update equation modifies the position and orientation according to their respective velocities, and allows these velocities themselves to change. The observation equation maps each 3D point through a pinhole camera model to create a predicted 2D image position (i.e., similar to equation 19.50). The actual measurements consist of interest points in the image, which are uniquely identified and associated with the predicted measurements by considering their surrounding region. The system is very efficient as it is only necessary to search the image in regions close to the 2D point predicted by the measurement equation from the current state.

The full system is more complex, and includes special procedures for initializing new feature points, modeling the local region of each point as a plane, selecting which features to measure in this frame, and managing the resulting map by deleting extraneous features to reduce the complexity. A series of frames from the system is illustrated in figure 19.16.
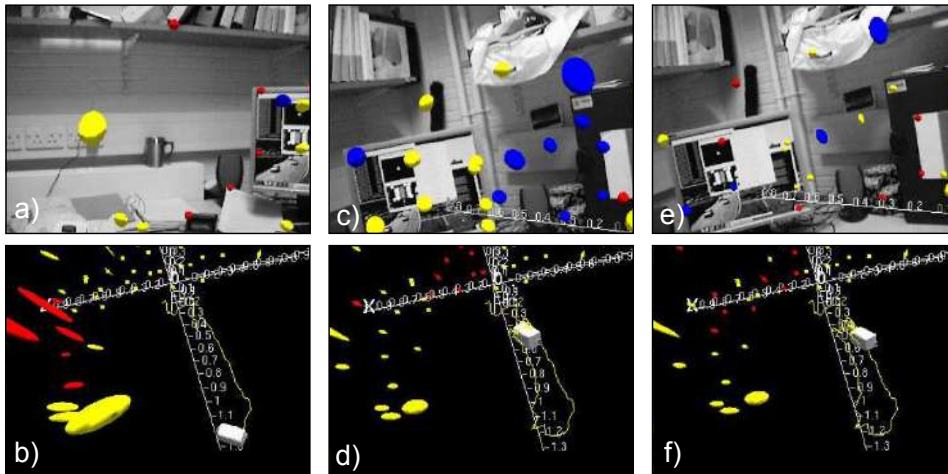
Figure 19.16 Monocular SLAM model of Davison *et al.* (2007). a) Current view of world with visual features and their 3D covariance ellipse superimposed. Yellow points are features in the model that were measured. Blue points are features in the model that were not detected in this frame. Red points are new features that can potentially be incorporated into the model. b) Model of world at the time this frame was captured, showing the position of the camera and the position and uncertainty of the visual features. c-f) Two more pairs of images and associated models from later in the same time sequence. Notice that the uncertainty in the feature positions gradually decreases over time. Adapted from Davison *et al.* (2007). ©2007 IEEE.

19.6.3   Tracking a contour through clutter

In section 17.3 we discussed fitting a template shape model to an image. It was noted that this was a challenging problem as the template frequently fell into local minima when the object was observed in clutter. In principle, this problem should be easier in a temporal sequence; if we know the template position in the current frame, then we can hypothesize quite accurately where it is in the next frame. Nonetheless, tracking methods based on the Kalman filter still fail to maintain a lock on the object; sooner or later, part of the contour becomes erroneously associated with the background and the tracking fails.

Blake & Isard (1998) describe a system based on the condensation algorithm that can cope with tracking a contour undergoing rapid motions against a cluttered background. The state of the system consists of the parameters of an affine transform that maps the template into the scene and the measurements are similar to those illustrated in figure 19.14. The state is represented as 100 weighted particles, each of which represents a different hypothesis about the current position of the template. This system was used to track the head of a child dancing in front of a cluttered background (figure 19.17).
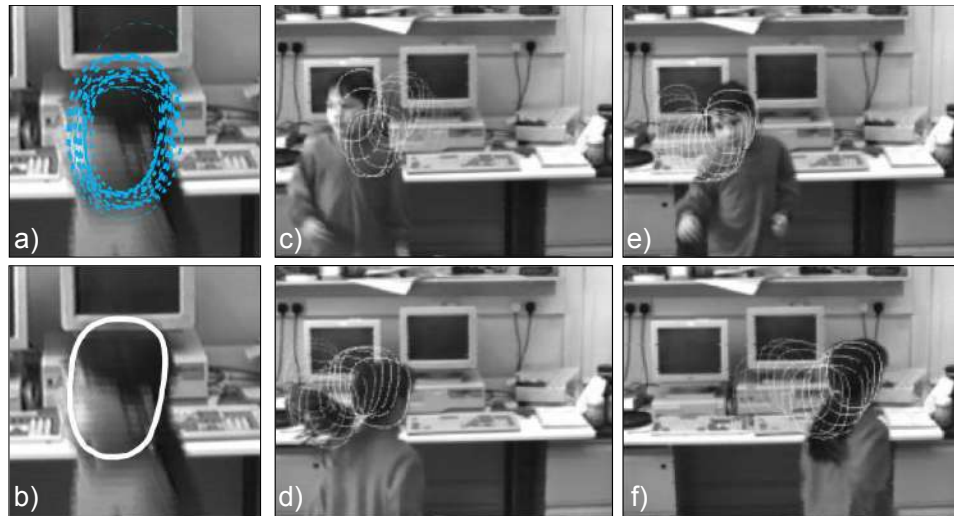
Figure 19.17 Tracking a contour through clutter. a) The representation of uncertainty about the contour consists of a set of weighted particles, each of which represents a hypothesis about the current position. In this panel, their weight indicates our relative belief that the system takes this state. b) The overall estimate of the current position can be illustrated by taking a weighted mean of the hypotheses. c-f) Four frames of sequence tracked by condensation algorithm showing the current estimate and the estimates from previous frames. Adapted from Blake & Isard (1998). ©1998 Springer.

Discussion

In this chapter we have discussed a family of models that allow us to estimate a set of continuous parameters throughout a time sequence, by exploiting the temporal coherence of the underlying state. In principle, these methods can be applied to any model in this book that estimates a set of parameters from a single frame. These models have a close relationship with the chain-based models discussed in chapter 11. The latter models used discrete parameters and generally estimated the state from the whole sequence rather than just the observed data up until the current time.

Notes

Applications of tracking: Tracking models are used for a variety of tasks in vision including tracking pedestrians (Rosales & Sclaroff 1999; Beymer & Konolige 1999), contours (Terzopolous & Szeliski 1992; Blake *et al.* 1993; Blake *et al.* 1995; Blake & Isard 1996; Blake & Isard 1998), points (Broida & Chellappa 1986), 3D hand models (Stenger *et al.* 2001b) and 3D body models (Wang *et al.* 2008). They have also been used for activity recognition (Vaswani *et al.* 2003), estimating depth (Matthies *et al.* 1989), SLAM (Davison *et al.* 2007), and object recognition (Zhou *et al.* 2004). Reviews of tracking methods and applications can be found in Blake (2006) and Yilmaz *et al.* (2006). Approaches to tracking the human body are reviewed in Poppe (2007).

Tracking models: The Kalman filter was originally developed by Kalman (1960) and Kalman & Bucy (1961). The unscented Kalman filter was developed by Julier & Uhlmann (1997). The condensation algorithm is due to Blake & Isard (1996). For more information about the Kalman filter and its variants, consult Maybeck (1990), Gelb (1974) and Jazwinski (1970). Roweis & Ghahramani (1999) provide a unifying review of linear models which provides information about how the Kalman filter relates to other methods. Arulampalam *et al.* (2002) provide a detailed review of the use of particle filters. A summary of tracking models and different approaches to inference within them can be found in Minka (2002).

There are a large number of variants on the standard Kalman filter. Many of these involve switching between different state space models or propagating mixtures (Shumway & Stoffer 1991; Ghahramani & Hinton 1996b; Murphy 1998; Chen & Liu 2000; Isard & Blake 1998). A notable recent extension has been to develop a nonlinear tracking algorithm based on the GPLVM (Wang *et al.* 2008).

One topic that has not been discussed in this chapter is how to learn the parameters of the tracking models. In practice, it is not uncommon to set these parameters by hand. However, information about learning in these temporal models can be found in Shumway & Stoffer (2007), Ghahramani & Hinton (1996a), Roweis & Ghahramani (2001), and Oh *et al.* (2005).

Simultaneous localization and mapping (SLAM): Simultaneous localization and mapping has its roots in the robotics community who were concerned with the representation of spatial uncertainty by vehicles exploring an environment (Durrant-Whyte 1988; Smith & Cheeseman 1987) although the term SLAM was coined much later by Durrant-Whyte *et al.* (1996). Smith *et al.* (1990) had the important insight that the errors in mapped positions were correlated due to uncertainty in the camera position. The roots of vision-based SLAM are to be found in the pioneering work of Harris & Pike (1987), Ayache (1991), and Beardsley *et al.* (1995).

SLAM systems are usually based on either the extended Kalman filter (Guivant & Nebot 2001; Leonard & Feder 2000; Davison *et al.* 2007) or a Rao-Blackwellized particle filter (Montemerlo *et al.* 2002; Montemerlo *et al.* 2003; Sim *et al.* 2005). However, it currently the subject of some debate as to whether a tracking method of this type is necessary at all, or whether repeated bundle adjustments on tactically chosen subsets of 3D points will suffice (Strasdat *et al.* 2010).

Recent examples of efficient visual SLAM systems can be found in Nistér *et al.* (2004), Davison *et al.* (2007), Klein & Murray (2007), Mei *et al.* (2009), and Newcombe *et al.* (2011), and most of these include a bundle adjustment procedure into the algorithm. Current research issues in SLAM include how to efficiently match features in the image with features in the current model (Handa *et al.* 2010) and how to close loops (i.e.,

recognize that the robot has returned to a familiar place) in a map (Newman & Ho 2005). A review of SLAM techniques can be found in Durrant-Whyte & Bailey (2006) and Bailey & Durrant-Whyte (2006).

Problems

Problem 19.1 Prove the Chapman-Kolmogorov relation:

$$
\begin{aligned}
Pr(\mathbf{w}_t|\mathbf{x}_{1...t-1}) &= \int Pr(\mathbf{w}_t|\mathbf{w}_{t-1})Pr(\mathbf{w}_{t-1}|\mathbf{x}_{1...t-1})\,d\mathbf{w}_{t-1} \\
&= \int \mathrm{Norm}_{\mathbf{w}_t}[\boldsymbol{\mu}_p + \boldsymbol{\Psi}\mathbf{w}_{t-1}, \boldsymbol{\Sigma}_p]\mathrm{Norm}_{\mathbf{w}_{t-1}}[\boldsymbol{\mu}_{t-1}, \boldsymbol{\Sigma}_{t-1}]\,d\mathbf{w}_{t-1} \\
&= \mathrm{Norm}_{\mathbf{w}_t}[\boldsymbol{\mu}_p + \boldsymbol{\Psi}\boldsymbol{\mu}_{t-1}, \boldsymbol{\Sigma}_p + \boldsymbol{\Psi}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Psi}^T]
\end{aligned}
$$

Problem 19.2 Derive the measurement incorporation step for the Kalman filter. In other words, show that

$$
\begin{aligned}
Pr(\mathbf{w}_t|\mathbf{x}_{1...t}) &= \frac{Pr(\mathbf{x}_t|\mathbf{w}_t)Pr(\mathbf{w}_t|\mathbf{w}_{1...t-1}, \mathbf{x}_{1...t})}{Pr(\mathbf{x}_{1...t})} \\
&= \frac{\mathrm{Norm}_{\mathbf{x}_t}[\boldsymbol{\mu}_m + \boldsymbol{\Phi}\mathbf{w}_t, \boldsymbol{\Sigma}_m]\mathrm{Norm}_{\mathbf{w}_t}[\boldsymbol{\mu}_+, \boldsymbol{\Sigma}_+]}{Pr(\mathbf{x}_{1...t})} \\
&= \mathrm{Norm}_{\mathbf{w}_t}\left[\left(\boldsymbol{\Phi}^T\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\Phi} + \boldsymbol{\Sigma}_+^{-1}\right)^{-1}\left(\boldsymbol{\Phi}^T\boldsymbol{\Sigma}_m^{-1}(\mathbf{x}_t - \boldsymbol{\mu}_m) + \boldsymbol{\Sigma}_+^{-1}\boldsymbol{\mu}_+\right),\right. \\
&\qquad\qquad\qquad\qquad\qquad\left.\left(\boldsymbol{\Phi}^T\boldsymbol{\Sigma}_m^{-1}\boldsymbol{\Phi} + \boldsymbol{\Sigma}_+^{-1}\right)^{-1}\right].
\end{aligned}
$$

Problem 19.3 Consider a system similar to the Kalman filter but where the prior based on the previous time step is a mixture of K Gaussians

$$
Pr(\mathbf{w}_t|\mathbf{x}_{1...t-1}) = \sum_{k=1}^{K} \lambda_k \mathrm{Norm}_{\mathbf{w}_t}[\boldsymbol{\mu}_{+k}, \boldsymbol{\Sigma}_{+k}].
$$

What will happen in the subsequent measurement incorporation step? What will happen in the next time update step?

Problem 19.4 Consider a model where there are two possible temporal update equations, represented by state transition matrices $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_2$ and the system periodically switches from one regime to the other. Write a set of equations that describe this model and discuss a strategy for max-marginals inference.

Problem 19.5 In a Kalman filter model, discuss how you would compute the joint posterior distribution $Pr(\mathbf{w}_{1...T}|\mathbf{x}_{1...T})$ over all of the unknown world states. What form will this posterior distribution take? In the Kalman filter we choose to compute the marginal posteriors instead. Why is this?

Problem 19.6 Apply the sum-product algorithm (section 11.4.3) to the Kalman filter model and show that the result is equivalent to applying the Kalman filter recursions.

Problem 19.7 Prove the Kalman smoother recursions:

$$
\begin{aligned}
\boldsymbol{\mu}_{t|T} &= \boldsymbol{\mu}_t + \mathbf{C}_t(\boldsymbol{\mu}_{t+1|T} - \boldsymbol{\mu}_{+|t}) \\
\boldsymbol{\Sigma}_{t|T} &= \boldsymbol{\Sigma}_t + \mathbf{C}_t(\boldsymbol{\Sigma}_{t+1|T} - \boldsymbol{\Sigma}_{+|t})\mathbf{C}_t^T,
\end{aligned}
$$

where

$$
\mathbf{C}_t = \boldsymbol{\Sigma}_{t|t}\boldsymbol{\Psi}^T\boldsymbol{\Sigma}_{+|t}^{-1}.
$$

Hint: It may help to examine the proof of the forward-backward algorithm for HMMs (section 11.4.2).

Problem 19.8 Discuss how you would learn the parameters of the Kalman filter model given training sequences consisting of (i) both the known world state and the observed data, and (ii) just the observed data alone.

Problem 19.9 In the unscented Kalman filter we represented a Gaussian with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ with a set of weights

$$
\begin{aligned}
\hat{\mathbf{w}}^{[0]} &= \boldsymbol{\mu}_{t-1} \\
\hat{\mathbf{w}}^{[j]} &= \boldsymbol{\mu}_{t-1} + \sqrt{\frac{D_{\mathbf{w}}}{1-a_0}}\boldsymbol{\Sigma}_{t-1}^{1/2}\mathbf{e}_j && \text{for all } j \in \{1 \ldots D_{\mathbf{w}}\} \\
\hat{\mathbf{w}}^{[D_{\mathbf{w}}+j]} &= \boldsymbol{\mu}_{t-1} - \sqrt{\frac{D_{\mathbf{w}}}{1-a_0}}\boldsymbol{\Sigma}_{t-1}^{1/2}\mathbf{e}_j && \text{for all } j \in \{1 \ldots D_{\mathbf{w}}\},
\end{aligned}
$$

with associated weights

$$
a_j = \frac{1-a_0}{2D_{\mathbf{w}}}.
$$

Show that the mean and covariance of these points are indeed $\boldsymbol{\mu}_{t-1}$ and $\boldsymbol{\Sigma}_{t-1}$ so that

$$
\begin{aligned}
\boldsymbol{\mu}_{t-1} &= \sum_{j=0}^{2D_{\mathbf{w}}} a_j\hat{\mathbf{w}}^{[j]} \\
\boldsymbol{\Sigma}_{t-1} &= \sum_{j=0}^{2D_{\mathbf{w}}} a_j(\hat{\mathbf{w}}^{[j]} - \boldsymbol{\mu}_{t-1})(\hat{\mathbf{w}}^{[j]} - \boldsymbol{\mu}_{t-1})^T.
\end{aligned}
$$

Problem 19.10 The extended Kalman filter requires the Jacobian matrix describing how small changes in the data create small changes in the measurements. Compute the Jacobian matrix for the measurement model for the pedestrian-tracking application (equation 19.50).