UNIVERSITÄT BONN

Juergen Gall

Temporal Filtering

MA-INF 2201 - Computer Vision

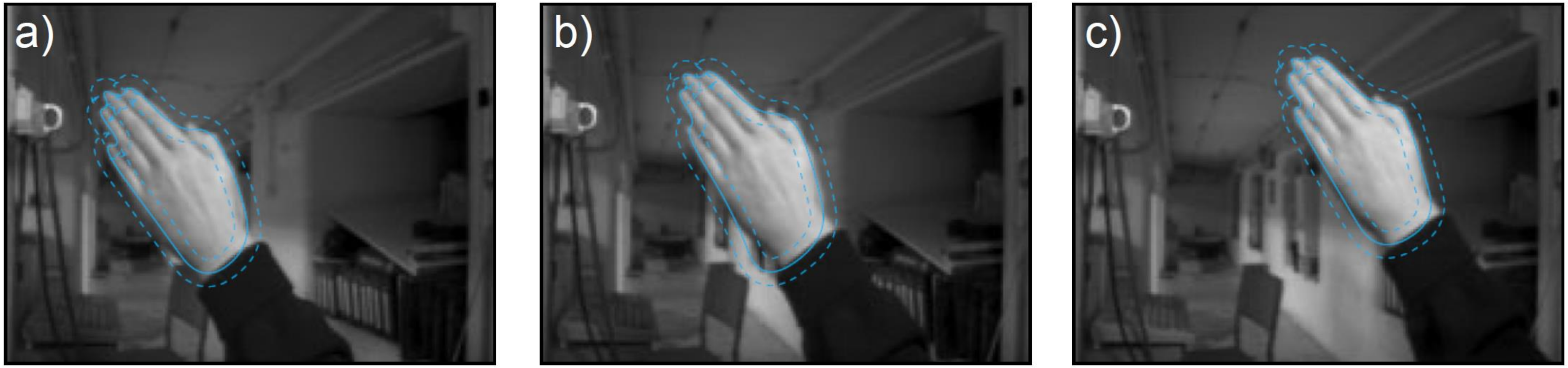
WS24/25

Literature

Chapter 19 Temporal Models

S. Prince. **Computer Vision: Models, Learning, and Inference.** Cambridge University Press 2012

Goal



To track object state from frame to frame in a video

Difficulties:

- Clutter (data association)
- One image may not be enough to fully define state
- Relationship between frames may be complicated

Source: S. Prince

Structure

- Temporal models
- Kalman filter
- Extended Kalman filter
- Unscented Kalman filter
- Particle filters
- Applications

Source: S. Prince

Temporal Models

- Consider an evolving system
- Represented by an unknown vector, **w**
- This is termed the state
- Examples:
  - 2D Position of tracked object in image
  - 3D Pose of tracked object in world
  - Joint positions of articulated model
- Goal: To compute the marginal posterior distribution over **w** at time t.

Source: S. Prince

Estimating State

Two contributions to estimating the state:

1. A set of measurements $\mathbf{x}_t$, which provide information about the state $\mathbf{w}_t$ at time t. This is a generative model: the measurements are derived from the state using a known probability relation $Pr(\mathbf{x}_t|\mathbf{w}_1\ldots\mathbf{w}_T)$
2. A time series model, which says something about the expected way that the system will evolve e.g., $Pr(\mathbf{w}_t|\mathbf{w}_1\ldots\mathbf{w}_{t-1},\mathbf{w}_{t+1}\ldots\mathbf{w}_T)$

Source: S. Prince

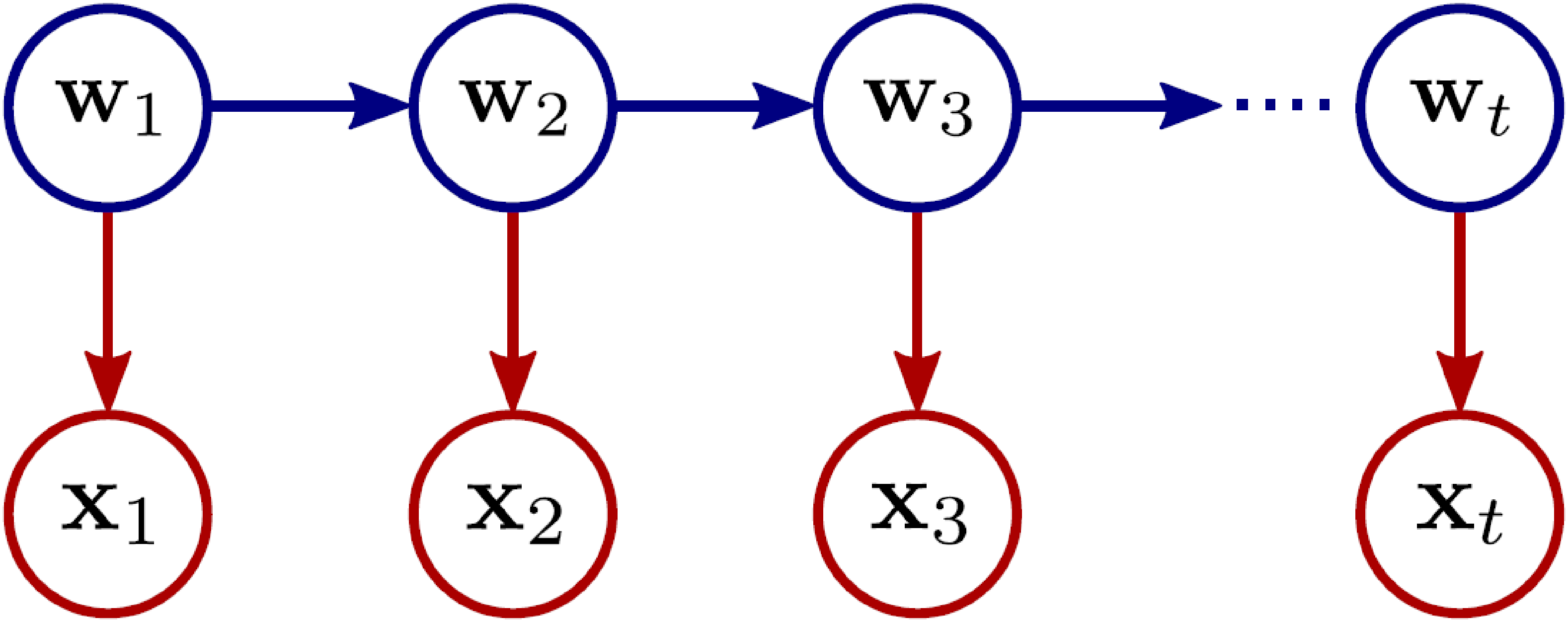
Assumptions

- Only the immediate past matters (Markov)
  - the probability of the state at time t is conditionally independent of states at times 1...t-2 given the state at time t-1, i.e., $P(w_t|w_{t-1}, \ldots, w_1) = P(w_t|w_{t-1})$

- Measurements depend on only the current state
  - the likelihood of the measurements at time t is conditionally independent of all of the other measurements and the states at times 1...t-1 given the state at time t, $P(x_t|w_t, \ldots, w_1, x_{t-1}, \ldots, x_1) = P(x_t|w_t)$

Source: S. Prince

Graphical Model



World states

Measurements

Source: S. Prince

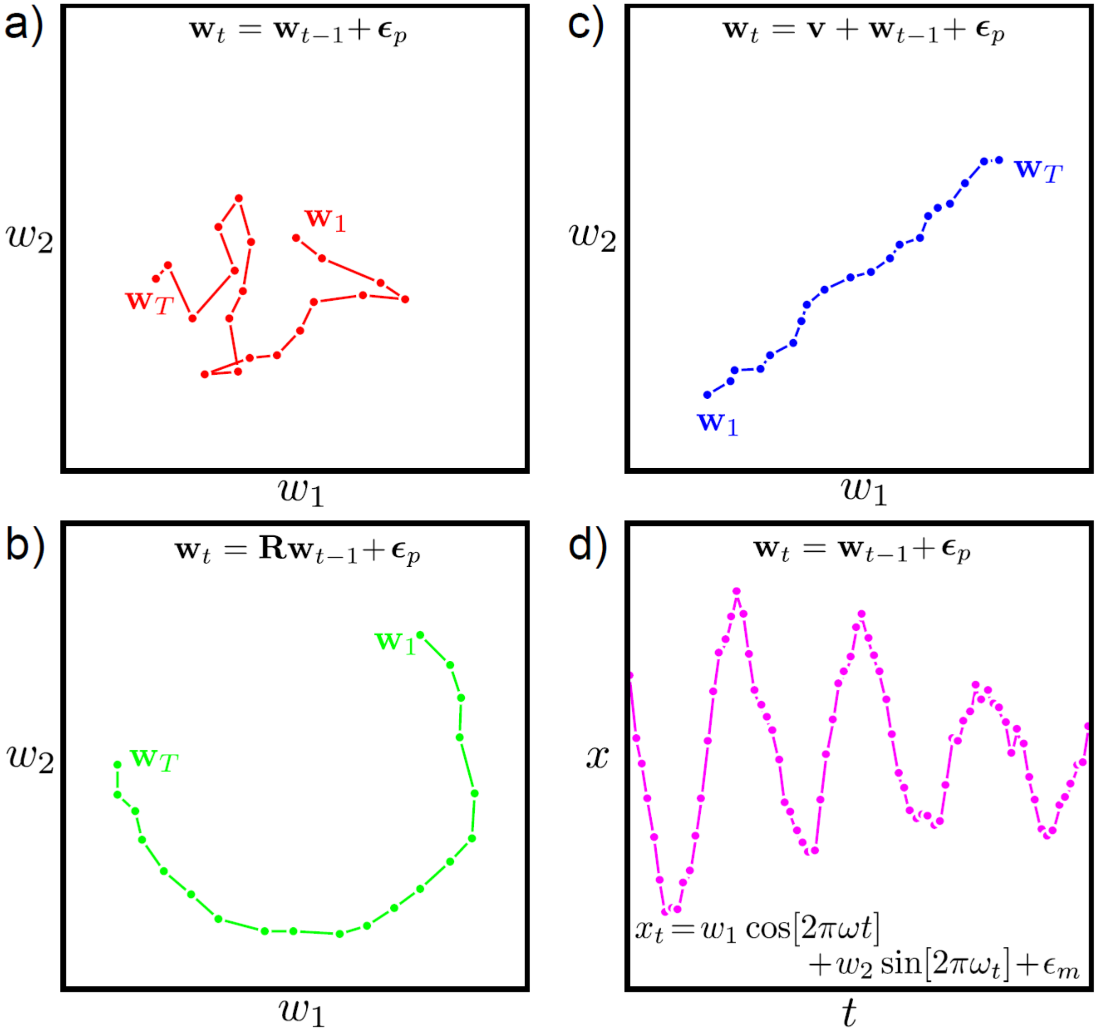
Temporal Models

Brownian motion

Constant velocity

Constant rotation

Brownian motion with oscillatory measurements



Source: S. Prince

Recursive Estimation

Time 1

$$Pr(w_1|x_1) = \frac{Pr(x_1|w_1)Pr(w_1)}{Pr(x_1)}$$

$$= \frac{Pr(x_1|w_1)Pr(w_1)}{\int Pr(x_1|w_1)Pr(w_1)dw_1}$$

Time 2

$$Pr(w_2|x_2,x_1) = \frac{Pr(x_2|w_2,x_1)Pr(w_2|x_1)}{Pr(x_2|x_1)}$$

$$= \frac{Pr(x_2|w_2)Pr(w_2|x_1)}{\int Pr(x_2|w_2,x_1)Pr(w_2|x_1)dw_2}$$

$$= \frac{Pr(x_2|w_2)Pr(w_2|x_1)}{\int Pr(x_2|w_2)Pr(w_2|x_1)dw_2}$$

Recursive Estimation

Time t

$$
\begin{aligned}
Pr(w_t|x_t, x_{1\ldots t-1}) &= \frac{Pr(x_t|w_t, x_{1\ldots t-1})Pr(w_t|x_{1\ldots t-1})}{Pr(x_t|x_{1\ldots t-1})} \\
&= \frac{Pr(x_t|w_t)Pr(w_t|x_{1\ldots t-1})}{\int Pr(x_t|w_t, x_{1\ldots t-1})Pr(w_t|x_{1\ldots t-1})dw_t} \\
&= \frac{\underbrace{Pr(x_t|w_t)}_{\text{Measurement model}}\underbrace{Pr(w_t|x_{1\ldots t-1})}_{\text{Prediction from temporal model}}}{\int Pr(x_t|w_t)Pr(w_t|x_{1\ldots t-1})dw_t}
\end{aligned}
$$

Computing the prior (time evolution)

Each time, the prior is based on the Chapman-Kolmogorov equation

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

Prior at time t

Temporal model

Posterior at time t-1

Source: S. Prince

Summary

Alternate between:

Temporal Evolution

Temporal model

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

Measurement Update

Measurement model

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

Source: S. Prince

What is the problem?

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

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

- No closed form solution
- Integrals expensive to compute
- Standard tricks:
  - Assume the world to be linear and Gaussian (Kalman filter)
  - Approximate non-linearity by Taylor expansion (Extended Kalman filter)
  - Sampling (Particle filter)

Structure

- Temporal models
- Kalman filter
- Extended Kalman filter
- Unscented Kalman filter
- Particle filters
- Applications

Source: S. Prince

Kalman Filter

The Kalman filter is just a special case of this type of recursive estimation procedure.

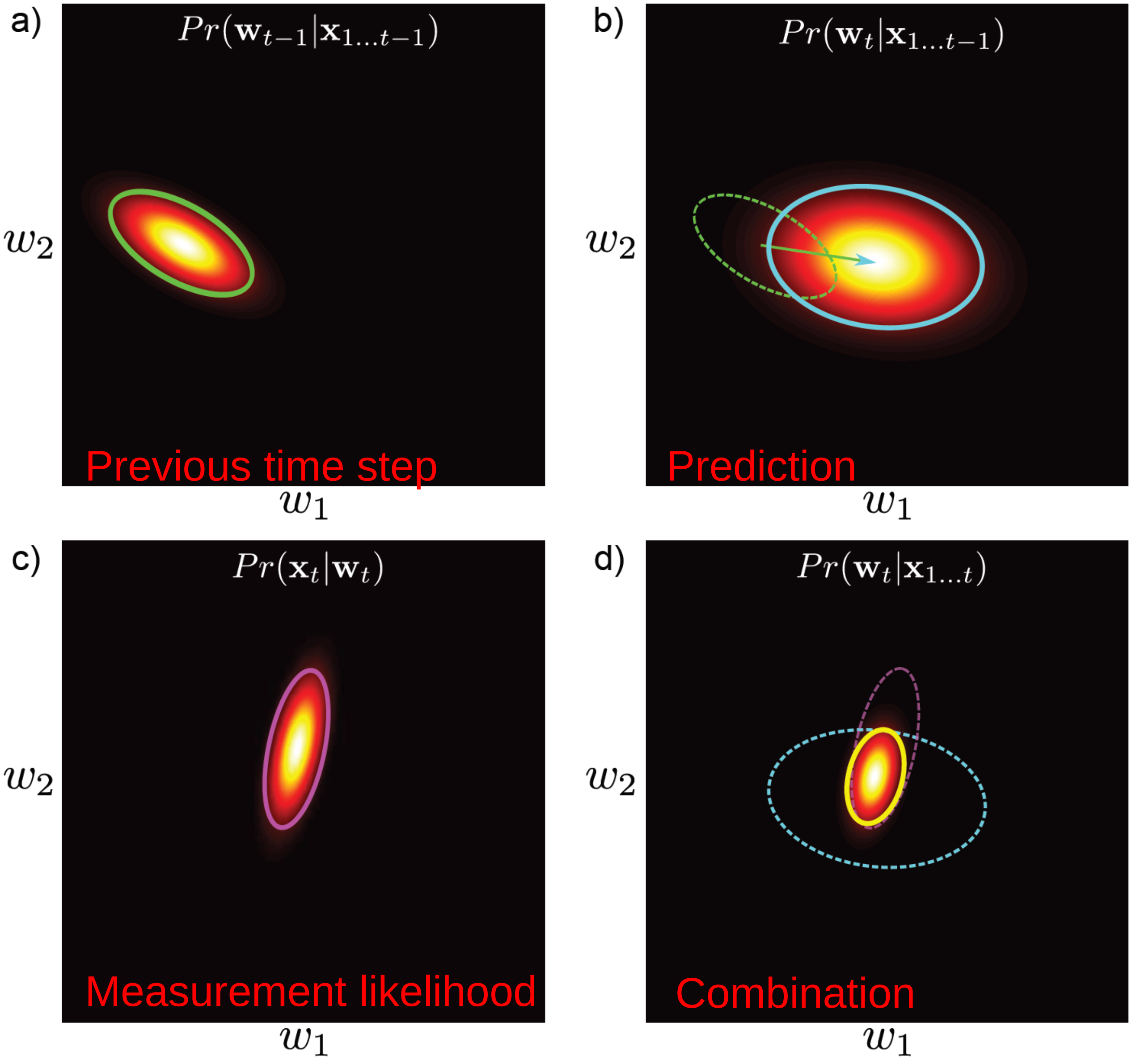
Temporal model and measurement model carefully chosen so that if the posterior at time t-1 was Gaussian then the

- prior at time t will be Gaussian
- posterior at time t will be Gaussian

The Kalman filter equations are rules for updating the means and covariances of these Gaussians

Source: S. Prince

The Kalman Filter


a) $Pr(\mathbf{w}_{t-1}|\mathbf{x}_{1\ldots t-1})$

Previous time step

b) $Pr(\mathbf{w}_t|\mathbf{x}_{1\ldots t-1})$

Prediction

c) $Pr(\mathbf{x}_t|\mathbf{w}_t)$

Measurement likelihood

d) $Pr(\mathbf{w}_t|\mathbf{x}_{1\ldots t})$

Combination

Source: S. Prince

Kalman Filter Definition

Time evolution equation

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

State transition matrix

Additive Gaussian noise

Measurement equation

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

Relates state and measurement

Additive Gaussian noise

Source: S. Prince

Kalman Filter Definition

Time evolution equation

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

State transition matrix

Additive Gaussian noise

Measurement equation

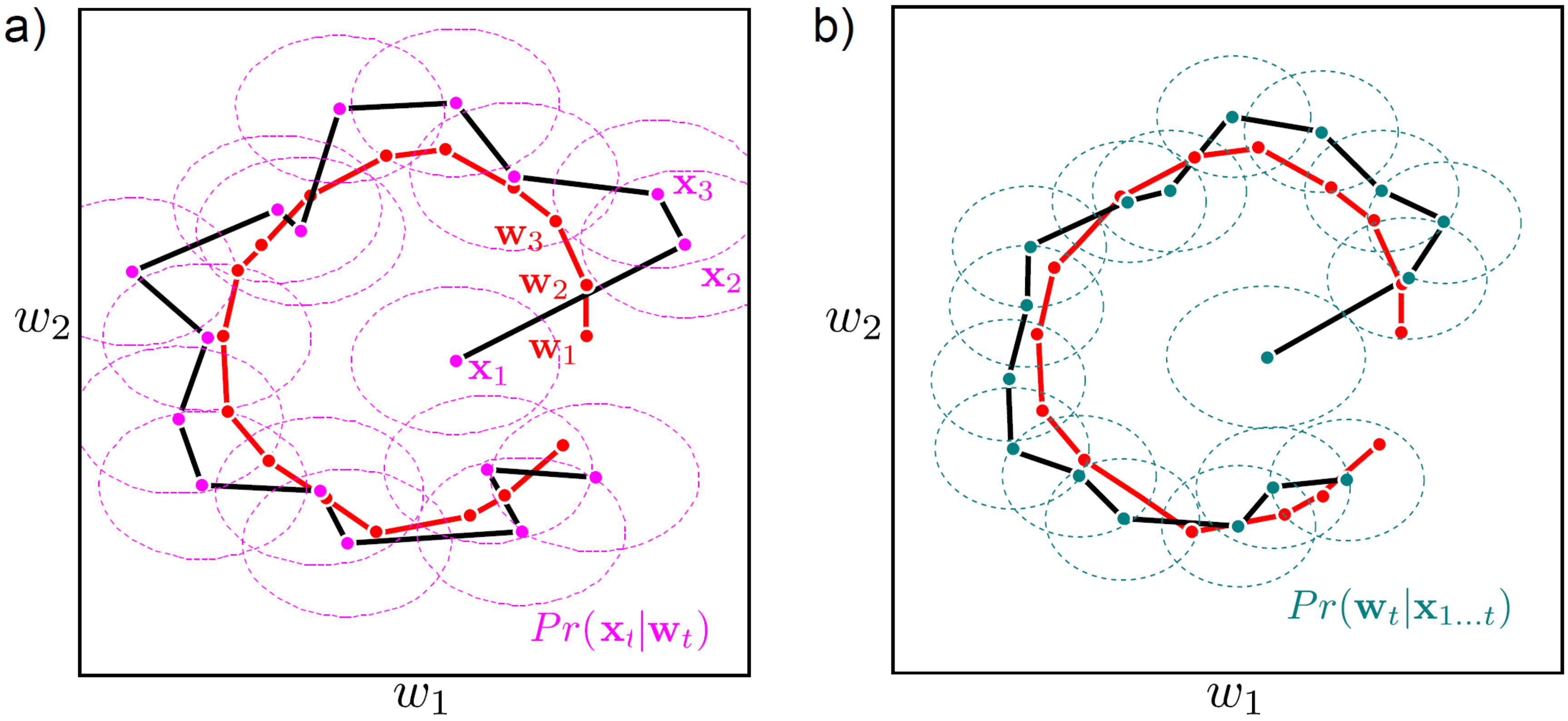
$$Pr(\mathbf{x}_t|\mathbf{w}_t) = \text{Norm}_{\mathbf{x}_t}[\boldsymbol{\mu}_m + \boldsymbol{\Phi}\mathbf{w}_t, \boldsymbol{\Sigma}_m]$$

Relates state and measurement

Additive Gaussian noise

Source: S. Prince

Kalman Filter Example



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

Not correct model!

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

Red: True State, Magenta: Observations, Green: Estimate (Kalman)

Benefit of temporal integration: Smoother and lower covariance

Source: S. Prince

Kalman Filter Example

General:

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

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

Example:

$$\begin{aligned} Pr(\mathbf{w}_t|\mathbf{w}_{t-1}) &= \text{Norm}_{\mathbf{w}_t}[\mathbf{w}_{t-1}, \sigma_p^2\mathbf{I}] \\ Pr(\mathbf{x}_t|\mathbf{w}_t) &= \text{Norm}_{\mathbf{x}_t}[\mathbf{w}_t, \boldsymbol{\Sigma}_m] \end{aligned}$$

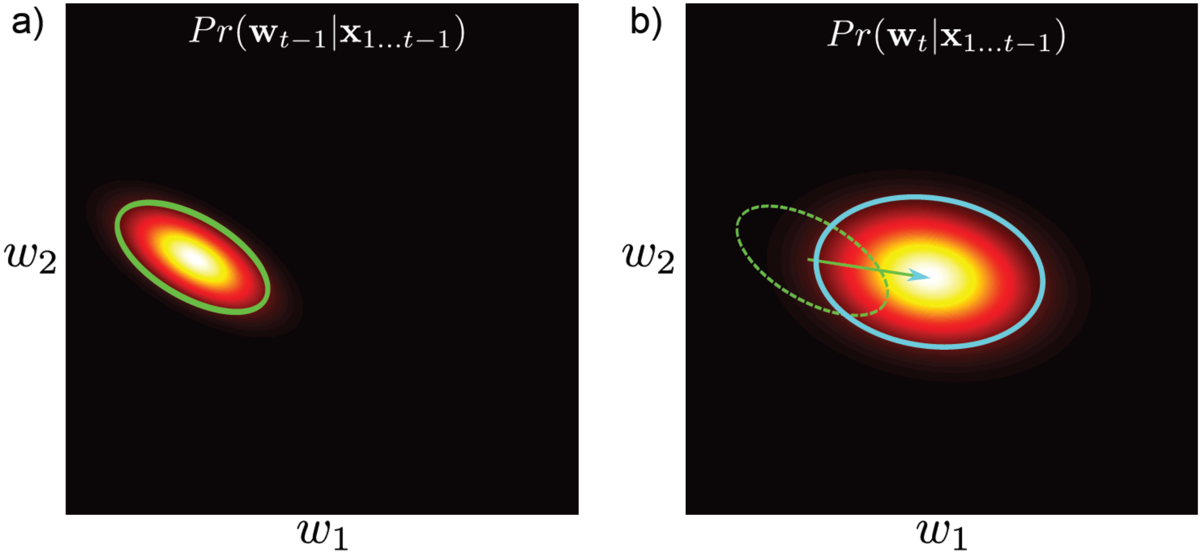
$$\mathbf{w}_t = \begin{pmatrix} w_{1,t} \\ w_{2,t} \end{pmatrix} \quad \mathbf{x}_t = \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix}$$

$$\mu_p = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \boldsymbol{\Psi} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \Sigma_p = \begin{pmatrix} \sigma_p^2 & 0 \\ 0 & \sigma_p^2 \end{pmatrix}$$

$$\mu_m = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \boldsymbol{\Phi} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \Sigma_m = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

Temporal evolution

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



Source: S. Prince

Temporal evolution

$$Pr(w_t|x_{1\ldots t-1})$$

$$= \int Pr(w_t|w_{t-1})Pr(w_{t-1}|x_{1\ldots t-1})dw_{t-1}$$

$$= \int \mathrm{Norm}_{w_t}[\mu_p + \Psi w_{t-1}, \Sigma_p]\,\mathrm{Norm}_{w_{t-1}}[\mu_{t-1}, \Sigma_{t-1}]dw_{t-1}$$

$$= \int \mathrm{Norm}_{w_t}[\mu_p + \Psi(\Psi^{-1}(w - \mu_p)), \Sigma_p]\,\mathrm{Norm}_{w}[\Psi\mu_{t-1} + \mu_p, \Psi\Sigma_{t-1}\Psi^T]dw$$

$$w = \Psi w_{t-1} + \mu_p$$

$$w_{t-1} = \Psi^{-1}(w - \mu_p)$$

Temporal evolution

$$Pr(w_t|x_{1\ldots t-1})$$

$$= \int Pr(w_t|w_{t-1})Pr(w_{t-1}|x_{1\ldots t-1})dw_{t-1}$$

$$= \int \mathrm{Norm}_{w_t}[\mu_p + \Psi w_{t-1}, \Sigma_p]\,\mathrm{Norm}_{w_{t-1}}[\mu_{t-1}, \Sigma_{t-1}]dw_{t-1}$$

$$= \int \mathrm{Norm}_{w_t}[\mu_p + \Psi(\Psi^{-1}(w - \mu_p)), \Sigma_p]\,\mathrm{Norm}_{w}[\Psi\mu_{t-1} + \mu_p, \Psi\Sigma_{t-1}\Psi^T]dw$$

$$= \int \mathrm{Norm}_{w}[w_t, \Sigma_p]\,\mathrm{Norm}_{w}[\Psi\mu_{t-1} + \mu_p, \Psi\Sigma_{t-1}\Psi^T]dw$$

$$= \mathrm{Norm}_{w_t}[\Psi\mu_{t-1} + \mu_p, \Sigma_p + \Psi\Sigma_{t-1}\Psi^T]$$

$$\cdot \int \mathrm{Norm}_{w}[\Sigma_*(\Sigma_p^{-1}w_t + (\Psi\Sigma_{t-1}\Psi^T)^{-1}(\Psi\mu_{t-1} + \mu_p)), \Sigma_*]dw$$

$$\boxed{\Sigma_* = (\Sigma_p^{-1} + (\Psi\Sigma_{t-1}\Psi^T)^{-1})^{-1}}$$

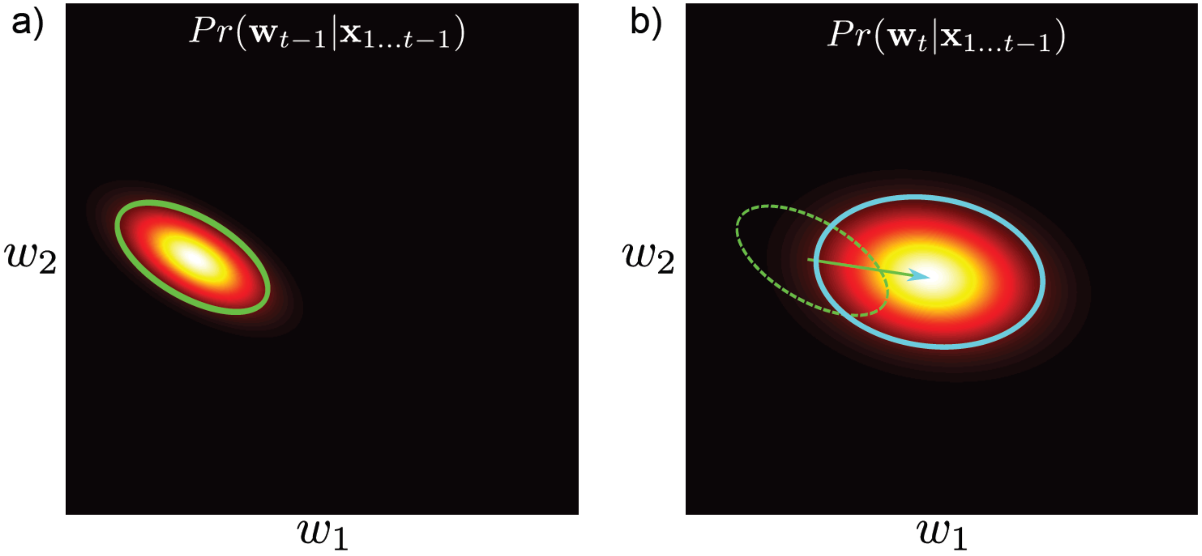
$$= \mathrm{Norm}_{w_t}[\Psi\mu_{t-1} + \mu_p, \Sigma_p + \Psi\Sigma_{t-1}\Psi^T]$$

$$= \mathrm{Norm}_{w_t}[\mu_+, \Sigma_+]$$

$$\boxed{\begin{array}{l}\int \mathrm{Norm}_x[a, A]\,\mathrm{Norm}_x[b, B]dx = \mathrm{Norm}_a[b, A + B] \\ \cdot \int \mathrm{Norm}_x[\Sigma_*(A^{-1}a + B^{-1}b), \Sigma_*]dx, \quad \Sigma_* = (A^{-1} + B^{-1})^{-1}\end{array}}$$

Temporal evolution

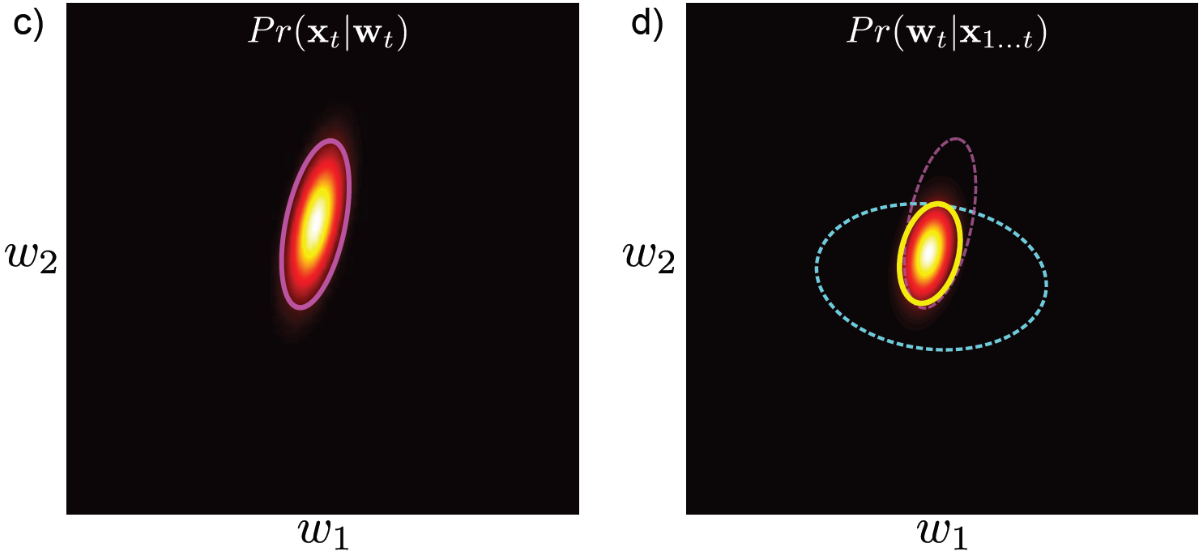
$$\begin{aligned} Pr(\mathbf{w}_t|\mathbf{x}_{1\ldots t-1}) &= \int Pr(\mathbf{w}_t|\mathbf{w}_{t-1})Pr(\mathbf{w}_{t-1}|\mathbf{x}_{1\ldots t-1})\, d\mathbf{w}_{t-1} \\ &= \int \mathrm{Norm}_{\mathbf{w}_t}[\boldsymbol{\mu}_p + \boldsymbol{\Psi}\mathbf{w}_{t-1}, \boldsymbol{\Sigma}_p]\mathrm{Norm}_{\mathbf{w}_{t-1}}[\boldsymbol{\mu}_{t-1},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



Source: S. Prince

Measurement incorporation

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



Source: S. Prince

Kalman Filter

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

This is not the usual way these equations are presented.

Part of the reason for this is the size of the inverses: ϕ is usually landscape and so $\phi^T\phi$ is inefficient.

Example:

- Measurement x: image coordinates (2d)
- State w: position + velocity (4d)
- $\Phi^T\Phi$ is 4x4 matrix
- $\Phi\Phi^T$ is 2x2 matrix

Source: S. Prince

Kalman Filter

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

Define Kalman gain:

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

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

Source: S. Prince

Matrix inversion relations

Consider the $d \times d$ matrix $\mathbf{A}$, the $k \times k$ matrix $\mathbf{C}$ and the $k \times d$ matrix $\mathbf{B}$ where $\mathbf{A}$ and $\mathbf{C}$ are symmetric, positive definite matrices. The following equality holds:

$$(\mathbf{A}^{-1} + \mathbf{B}^T\mathbf{C}^{-1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{C}^{-1} = \mathbf{A}\mathbf{B}^T(\mathbf{B}\mathbf{A}\mathbf{B}^T + \mathbf{C})^{-1}.$$

Proof:

$$\begin{aligned}
\mathbf{B}^T\mathbf{C}^{-1}\mathbf{B}\mathbf{A}\mathbf{B}^T + \mathbf{B}^T &= \mathbf{B}^T + \mathbf{B}^T\mathbf{C}^{-1}\mathbf{B}\mathbf{A}\mathbf{B}^T \\
\mathbf{B}^T\mathbf{C}^{-1}(\mathbf{B}\mathbf{A}\mathbf{B}^T + \mathbf{C}) &= (\mathbf{A}^{-1} + \mathbf{B}^T\mathbf{C}^{-1}\mathbf{B})\mathbf{A}\mathbf{B}^T.
\end{aligned}$$

Taking the inverse of both sides we get

$$(\mathbf{A}^{-1} + \mathbf{B}^T\mathbf{C}^{-1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{C}^{-1} = \mathbf{A}\mathbf{B}^T(\mathbf{B}\mathbf{A}\mathbf{B}^T + \mathbf{C})^{-1},$$

as required.

Matrix inversion relations

Consider the $d \times d$ matrix $\mathbf{A}$, the $k \times k$ matrix $\mathbf{C}$ and the $k \times d$ matrix $\mathbf{B}$ where $\mathbf{A}$ and $\mathbf{C}$ are symmetric, positive definite matrices. The following equality holds:

$$(\mathbf{A}^{-1} + \mathbf{B}^T \mathbf{C}^{-1} \mathbf{B})^{-1} = \mathbf{A} - \mathbf{A}\mathbf{B}^T (\mathbf{B}\mathbf{A}\mathbf{B}^T + \mathbf{C})^{-1} \mathbf{B}\mathbf{A}. \tag{C.61}$$

This is sometimes known as the *matrix inversion lemma.*

Proof:

$$\begin{aligned}
&(\mathbf{A}^{-1} + \mathbf{B}^T \mathbf{C}^{-1} \mathbf{B})^{-1} \\
&\quad = (\mathbf{A}^{-1} + \mathbf{B}^T \mathbf{C}^{-1} \mathbf{B})^{-1} (\mathbf{I} + \mathbf{B}^T \mathbf{C}^{-1} \mathbf{B}\mathbf{A} - \mathbf{B}^T \mathbf{C}^{-1} \mathbf{B}\mathbf{A}) \\
&\quad = (\mathbf{A}^{-1} + \mathbf{B}^T \mathbf{C}^{-1} \mathbf{B})^{-1} ((\mathbf{A}^{-1} + \mathbf{B}^T \mathbf{C}^{-1} \mathbf{B})\mathbf{A} - \mathbf{B}^T \mathbf{C}^{-1} \mathbf{B}\mathbf{A}) \\
&\quad = \mathbf{A} - (\mathbf{A}^{-1} + \mathbf{B}^T \mathbf{C}^{-1} \mathbf{B})^{-1} \mathbf{B}^T \mathbf{C}^{-1} \mathbf{B}\mathbf{A}. \qquad (C.62)
\end{aligned}$$

We know:

$$(\mathbf{A}^{-1} + \mathbf{B}^T \mathbf{C}^{-1} \mathbf{B})^{-1} \mathbf{B}^T \mathbf{C}^{-1} = \mathbf{A}\mathbf{B}^T (\mathbf{B}\mathbf{A}\mathbf{B}^T + \mathbf{C})^{-1}$$

Therefore:

$$(\mathbf{A}^{-1} + \mathbf{B}^T \mathbf{C}^{-1} \mathbf{B})^{-1} = \mathbf{A} - \mathbf{A}\mathbf{B}^T (\mathbf{B}\mathbf{A}\mathbf{B}^T + \mathbf{C})^{-1} \mathbf{B}\mathbf{A}$$

Mean Term

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

Kalman gain:

$$K = \Sigma_+\Phi^T(\Sigma_m + \Phi\Sigma_+\Phi^T)^{-1}$$

Matrix inversion relations:

$$(\Sigma_+^{-1} + \Phi^T\Sigma_m^{-1}\Phi)^{-1}\Phi^T\Sigma_m^{-1} = \Sigma_+\Phi^T(\Phi\Sigma_+\Phi^T + \Sigma_m)^{-1} = K$$

$$(\mathbf{A}^{-1}+\mathbf{B}^T\mathbf{C}^{-1}\mathbf{B})^{-1}\mathbf{B}^T\mathbf{C}^{-1} = \mathbf{A}\mathbf{B}^T(\mathbf{B}\mathbf{A}\mathbf{B}^T+\mathbf{C})^{-1}$$

Mean Term

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

Kalman gain:

$$K = \Sigma_+\Phi^T(\Sigma_m + \Phi\Sigma_+\Phi^T)^{-1}$$

Matrix inversion relations:

$$(\Sigma_+^{-1} + \Phi^T\Sigma_m^{-1}\Phi)^{-1}\Phi^T\Sigma_m^{-1} = \Sigma_+\Phi^T(\Phi\Sigma_+\Phi^T + \Sigma_m)^{-1} = K$$

$$(\Sigma_+^{-1} + \Phi^T\Sigma_m^{-1}\Phi)^{-1} = \Sigma_+ - \Sigma_+\Phi^T(\Phi\Sigma_+\Phi^T + \Sigma_m)^{-1}\Phi\Sigma_+$$

$$(\mathbf{A}^{-1} + \mathbf{B}^T\mathbf{C}^{-1}\mathbf{B})^{-1} = \mathbf{A} - \mathbf{A}\mathbf{B}^T(\mathbf{B}\mathbf{A}\mathbf{B}^T + \mathbf{C})^{-1}\mathbf{B}\mathbf{A}.$$

Mean Term

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

Kalman gain:

$$K = \Sigma_+\Phi^T(\Sigma_m + \Phi\Sigma_+\Phi^T)^{-1}$$

Matrix inversion relations:

$$(\Sigma_+^{-1} + \Phi^T\Sigma_m^{-1}\Phi)^{-1}\Phi^T\Sigma_m^{-1} = \Sigma_+\Phi^T(\Phi\Sigma_+\Phi^T + \Sigma_m)^{-1} = K$$

$$(\Sigma_+^{-1} + \Phi^T\Sigma_m^{-1}\Phi)^{-1} = \Sigma_+ - \Sigma_+\Phi^T(\Phi\Sigma_+\Phi^T + \Sigma_m)^{-1}\Phi\Sigma_+$$

Using Matrix inversion relations:

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

Covariance Term

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

Kalman gain:

$$K = \Sigma_+\Phi^T(\Sigma_m + \Phi\Sigma_+\Phi^T)^{-1}$$

Matrix inversion relations:

$$\begin{aligned}(\Sigma_+^{-1} + \Phi^T\Sigma_m^{-1}\Phi)^{-1} &= \Sigma_+ - \Sigma_+\Phi^T(\Phi\Sigma_+\Phi^T + \Sigma_m)^{-1}\Phi\Sigma_+ \\ &= \Sigma_+ - K\Phi\Sigma_+ \\ &= (I - K\Phi)\Sigma_+\end{aligned}$$

Final Kalman Filter Equation

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

Innovation (difference between actual and predicted measurements

Prior variance minus a term due to information from measurement

Source: S. Prince

Kalman Filter Summary

Time evolution equation

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

Measurement equation

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

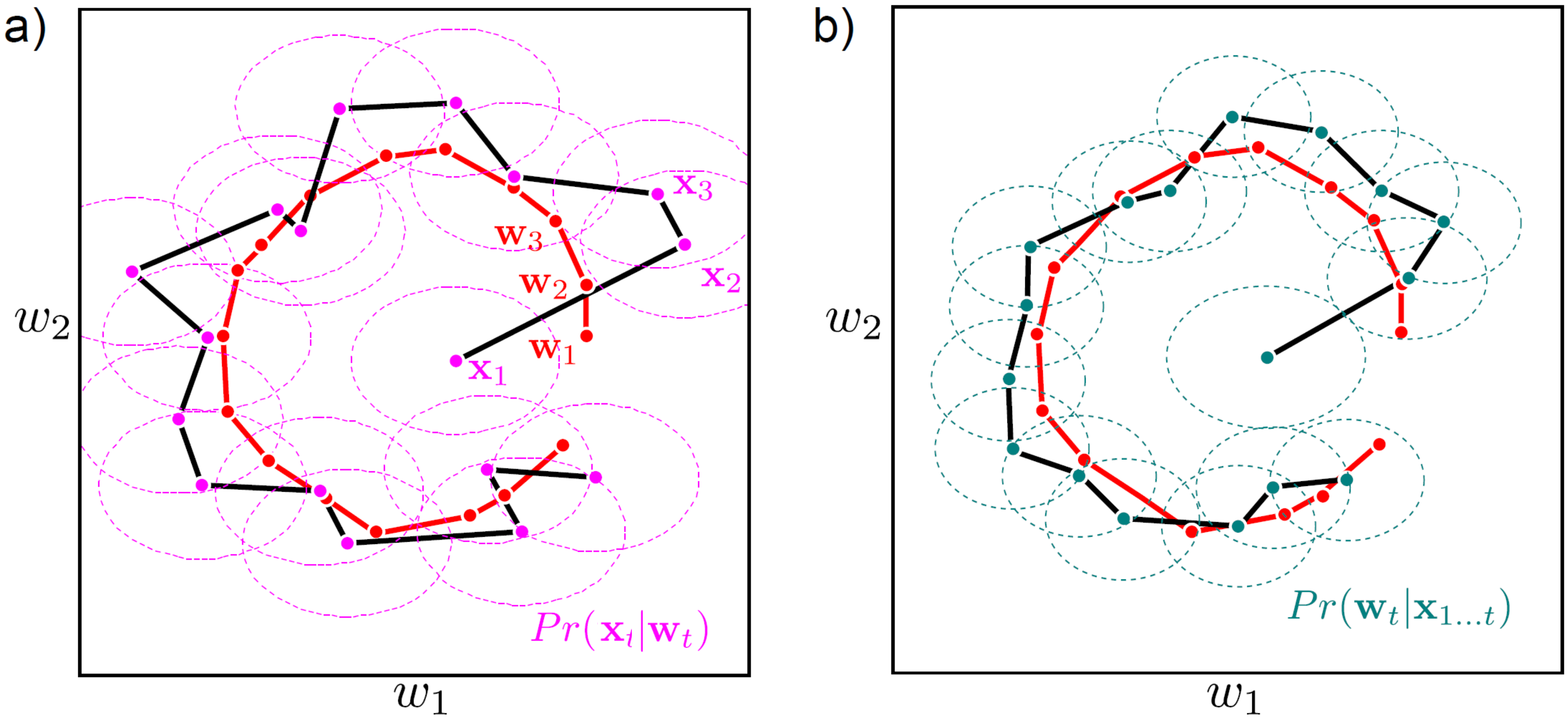
Inference

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

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

Source: S. Prince

Kalman Filter Example 1



$$\begin{aligned} Pr(\mathbf{w}_t|\mathbf{w}_{t-1}) &= \text{Norm}_{\mathbf{w}_t}[\mathbf{w}_{t-1}, \sigma_p^2\mathbf{I}] \\ Pr(\mathbf{x}_t|\mathbf{w}_t) &= \text{Norm}_{\mathbf{x}_t}[\mathbf{w}_t, \mathbf{\Sigma}_m] \end{aligned}$$

Not correct model!

Red: True State, Magenta: Observations, Green: Estimate (Kalman)

Benefit of temporal integration: Smoother and lower covariance

Source: S. Prince

Kalman Filter Example

General:

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

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

Example:

$$\begin{aligned} Pr(\mathbf{w}_t|\mathbf{w}_{t-1}) &= \text{Norm}_{\mathbf{w}_t}[\mathbf{w}_{t-1}, \sigma_p^2\mathbf{I}] \\ Pr(\mathbf{x}_t|\mathbf{w}_t) &= \text{Norm}_{\mathbf{x}_t}[\mathbf{w}_t, \boldsymbol{\Sigma}_m] \end{aligned}$$

$$\mathbf{w}_t = \begin{pmatrix} w_{1,t} \\ w_{2,t} \end{pmatrix} \quad \mathbf{x}_t = \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix}$$

$$\mu_p = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \boldsymbol{\Psi} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \Sigma_p = \begin{pmatrix} \sigma_p^2 & 0 \\ 0 & \sigma_p^2 \end{pmatrix}$$

$$\mu_m = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \boldsymbol{\Phi} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \Sigma_m = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

Kalman Filter Example

State Prediction: $\boldsymbol{\mu}_+ = \quad \boldsymbol{\mu}_p + \boldsymbol{\Psi}\boldsymbol{\mu}_{t-1}$

$$\begin{pmatrix} \mu_{1,+} \\ \mu_{2,+} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{1,t-1} \\ \mu_{2,t-1} \end{pmatrix} = \begin{pmatrix} \mu_{1,t-1} \\ \mu_{2,t-1} \end{pmatrix}$$

Covariance Prediction: $\boldsymbol{\Sigma}_+ = \quad \boldsymbol{\Sigma}_p + \boldsymbol{\Psi}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Psi}^T$

$$\begin{aligned} \begin{pmatrix} \Sigma_{11,+} & \Sigma_{12,+} \\ \Sigma_{21,+} & \Sigma_{22,+} \end{pmatrix} &= \begin{pmatrix} \sigma_p^2 & 0 \\ 0 & \sigma_p^2 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Sigma_{11,t-1} & \Sigma_{12,t-1} \\ \Sigma_{21,t-1} & \Sigma_{22,t-1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \Sigma_{11,t-1} + \sigma_p^2 & \Sigma_{12,t-1} \\ \Sigma_{21,t-1} & \Sigma_{22,t-1} + \sigma_p^2 \end{pmatrix} \end{aligned}$$

Kalman Gain: $\mathbf{K} = \boldsymbol{\Sigma}_+\boldsymbol{\Phi}^T(\boldsymbol{\Sigma}_m + \boldsymbol{\Phi}\boldsymbol{\Sigma}_+\boldsymbol{\Phi}^T)^{-1}$

$$\begin{aligned} \begin{pmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{pmatrix} &= \begin{pmatrix} \Sigma_{11,+} & \Sigma_{12,+} \\ \Sigma_{21,+} & \Sigma_{22,+} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \left(\begin{pmatrix} \Sigma_{11,m} & \Sigma_{12,m} \\ \Sigma_{21,m} & \Sigma_{22,m} \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Sigma_{11,+} & \Sigma_{12,+} \\ \Sigma_{21,+} & \Sigma_{22,+} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)^{-1} \\ &= \begin{pmatrix} \Sigma_{11,+} & \Sigma_{12,+} \\ \Sigma_{21,+} & \Sigma_{22,+} \end{pmatrix} \begin{pmatrix} \Sigma_{11,m} + \Sigma_{11,+} & \Sigma_{12,m} + \Sigma_{12,+} \\ \Sigma_{21,m} + \Sigma_{21,+} & \Sigma_{22,m} + \Sigma_{22,+} \end{pmatrix}^{-1} \end{aligned}$$

Kalman Filter Example

$$\text{State Update:} \qquad \boldsymbol{\mu}_t = \quad \boldsymbol{\mu}_+ + \mathbf{K}(\mathbf{x}_t - \boldsymbol{\mu}_m - \boldsymbol{\Phi}\boldsymbol{\mu}_+)$$

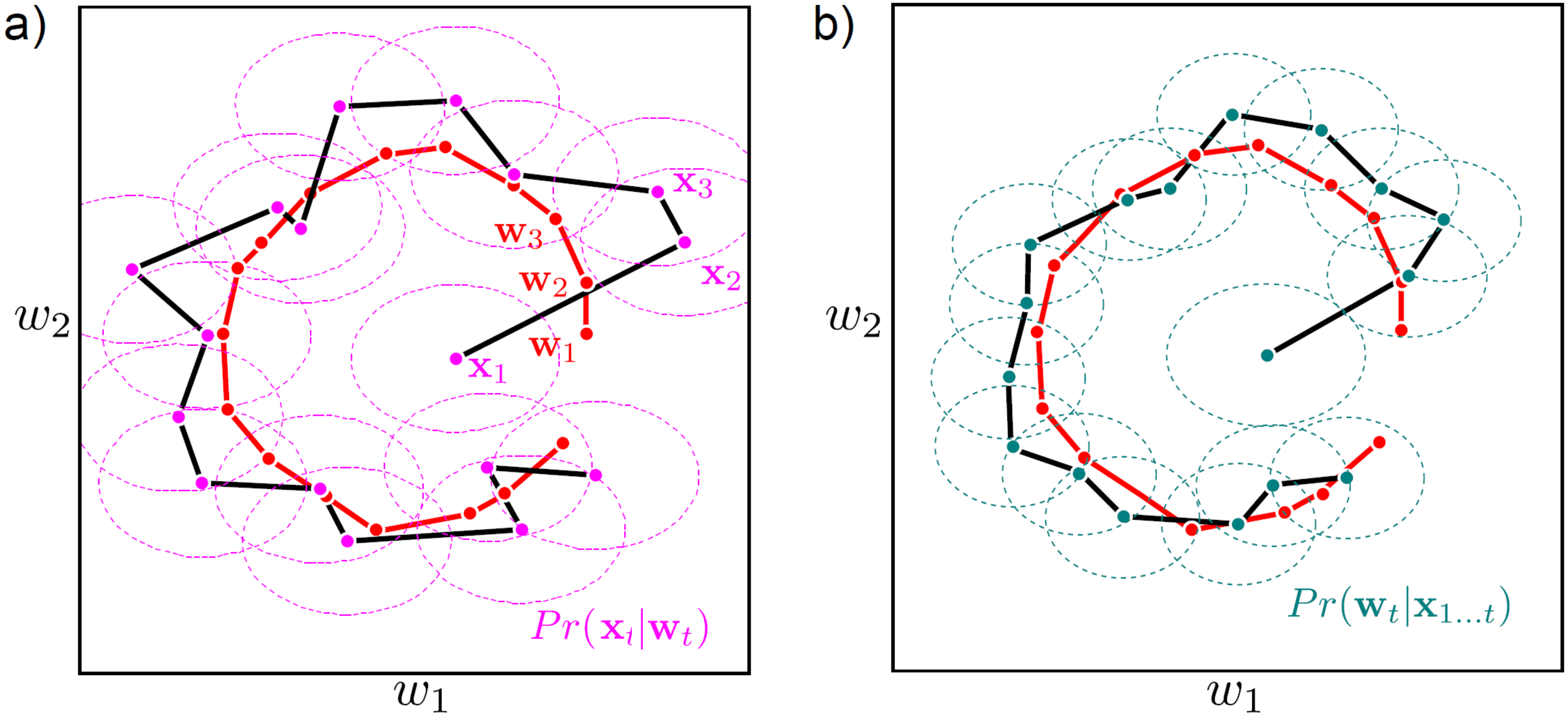
$$\begin{aligned}\begin{pmatrix}\mu_{1,t}\\ \mu_{2,t}\end{pmatrix} &= \begin{pmatrix}\mu_{1,+}\\ \mu_{2,+}\end{pmatrix} + \begin{pmatrix}K_{11} & K_{12}\\ K_{21} & K_{22}\end{pmatrix}\left(\begin{pmatrix}x_{1,t}\\ x_{2,t}\end{pmatrix} - \begin{pmatrix}0\\ 0\end{pmatrix} - \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}\mu_{1,+}\\ \mu_{2,+}\end{pmatrix}\right)\\ &= \begin{pmatrix}\mu_{1,+}\\ \mu_{2,+}\end{pmatrix} + \begin{pmatrix}K_{11} & K_{12}\\ K_{21} & K_{22}\end{pmatrix}\begin{pmatrix}x_{1,t} - \mu_{1,+}\\ x_{2,t} - \mu_{2,+}\end{pmatrix}\end{aligned}$$

$$\text{Covariance Update:} \qquad \boldsymbol{\Sigma}_t = \quad (\mathbf{I} - \mathbf{K}\boldsymbol{\Phi})\boldsymbol{\Sigma}_+,$$

$$\begin{aligned}\begin{pmatrix}\Sigma_{11,t} & \Sigma_{12,t}\\ \Sigma_{21,t} & \Sigma_{22,t}\end{pmatrix} &= \left(\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix} - \begin{pmatrix}K_{11} & K_{12}\\ K_{21} & K_{22}\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right)\begin{pmatrix}\Sigma_{11,+} & \Sigma_{12,+}\\ \Sigma_{21,+} & \Sigma_{22,+}\end{pmatrix}\\ &= \begin{pmatrix}1 - K_{11} & -K_{12}\\ -K_{21} & 1 - K_{22}\end{pmatrix}\begin{pmatrix}\Sigma_{11,+} & \Sigma_{12,+}\\ \Sigma_{21,+} & \Sigma_{22,+}\end{pmatrix}\end{aligned}$$

Except of one matrix inversion, very easy to compute!

Kalman Filter Example 1



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

Not correct model!

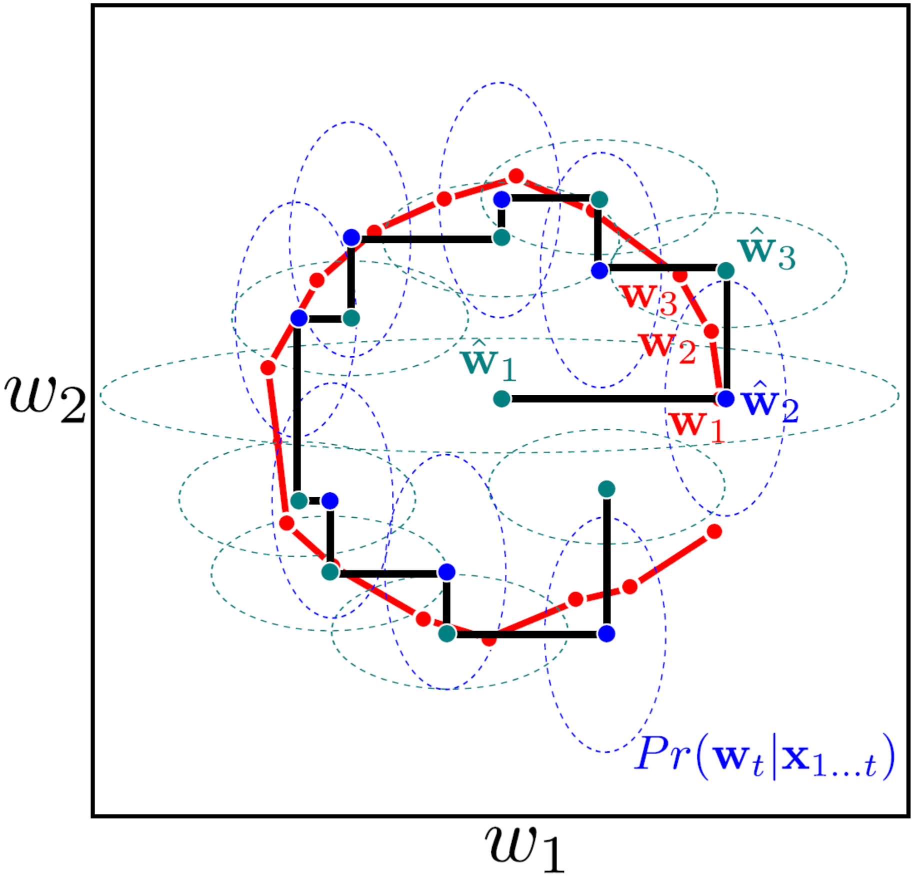
$$Pr(\mathbf{x}_t|\mathbf{w}_t) = \text{Norm}_{\mathbf{x}_t}[\mathbf{w}_t, \mathbf{\Sigma}_m]$$

Red: True State, Magenta: Observations, Green: Estimate (Kalman)

Benefit of temporal integration: Smoother and lower covariance

Source: S. Prince

Kalman Filter Example 2



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

At each time only one dimension is observed

Source: S. Prince

Smoothing

- Estimates depend only on measurements up to the current point in time.
- Sometimes want to estimate state based on future measurements as well

Fixed Lag Smoother:

This is an on-line scheme in which the optimal estimate for a state at time t - τ is calculated based on measurements up to time t, where τ is the time lag. i.e. we wish to calculate $Pr(w_{t-\tau} | x_1 ... x_t)$.

Fixed Interval Smoother:

We have a fixed time interval of measurements and want to calculate the optimal state estimate based on all of these measurements. In other words, instead of calculating $Pr(w_t | x_1 ... x_t)$ we now estimate $Pr(w_t | x_1 ... x_T)$ where T is the total length of the interval.

Source: S. Prince

Fixed lag smoother

State evolution equation

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

Estimate
delayed by τ

Measurement equation

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

Source: S. Prince

Kalman Filter Summary

Time evolution equation

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

Measurement equation

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

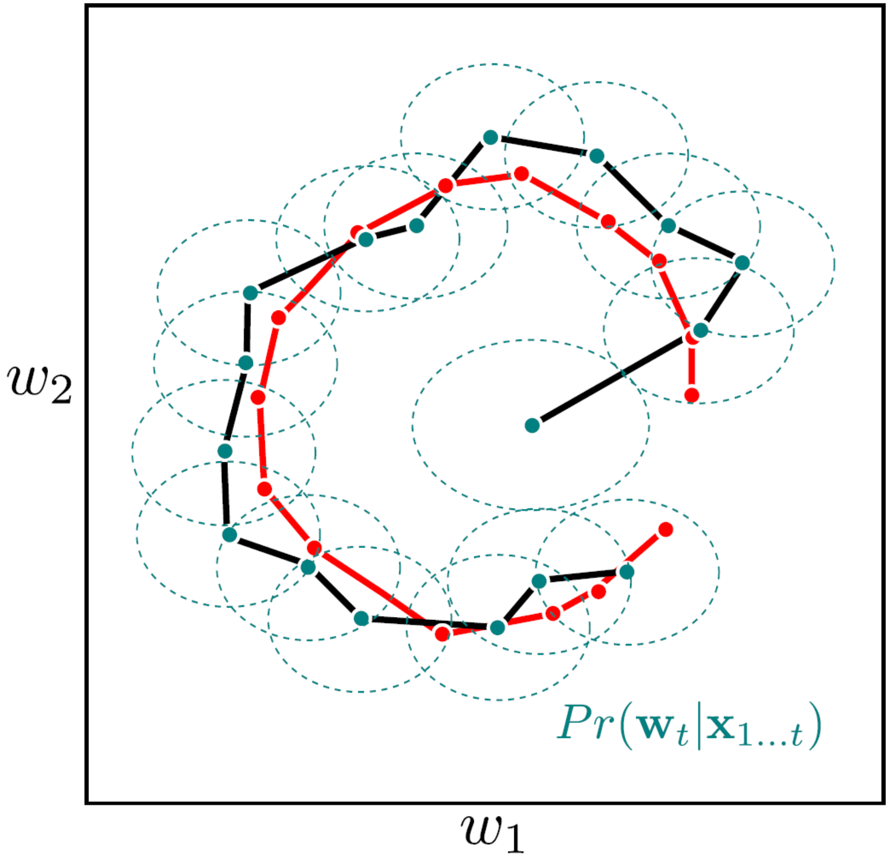
Inference

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

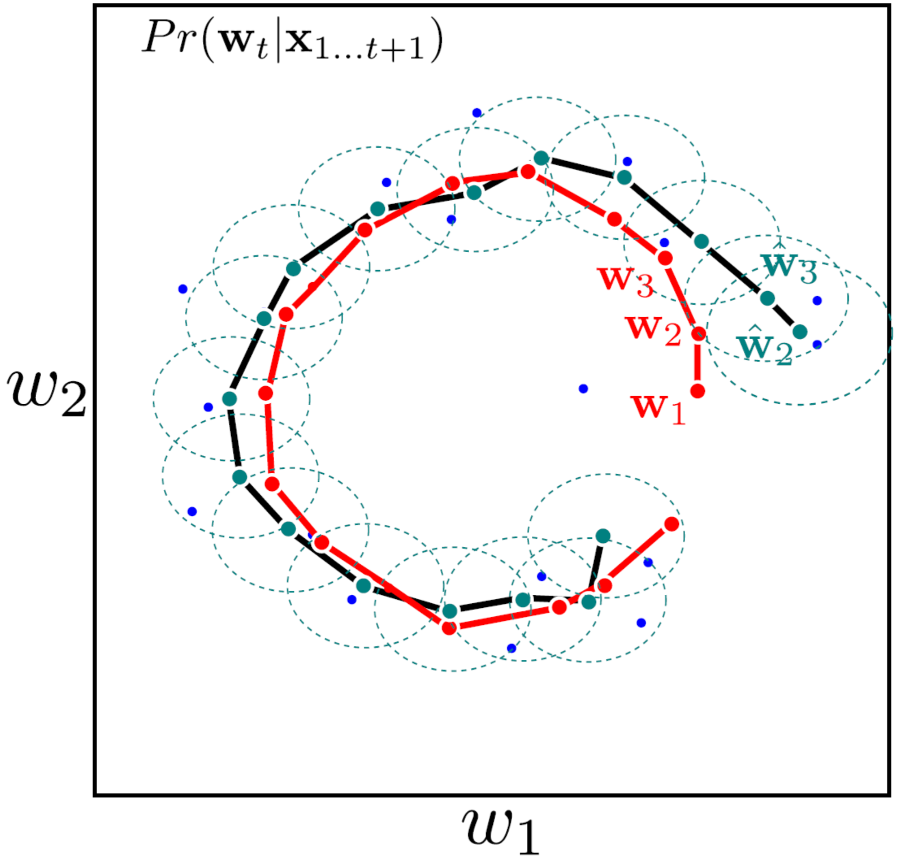
$$\mathbf{K} = \boldsymbol{\Sigma}_+\boldsymbol{\Phi}^T(\boldsymbol{\Sigma}_m + \boldsymbol{\Phi}\boldsymbol{\Sigma}_+\boldsymbol{\Phi}^T)^{-1}$$

Source: S. Prince

Fixed-lag Kalman Smoothing



Kalman filter



Kalman filter $\tau = 1$

Advantage: Better estimates and lower covariance
Disadvantage: Delay

Source: S. Prince

Fixed interval smoothing

Having μ_t, Σ_t, $\mu_{+|t}$, and $\Sigma_{+|t}$ from forward pass where $Pr(w_{t+1}|x_{1\ldots t}) = Norm(\mu_{+|t}, \Sigma_{+|t})$

Backward set of recursions

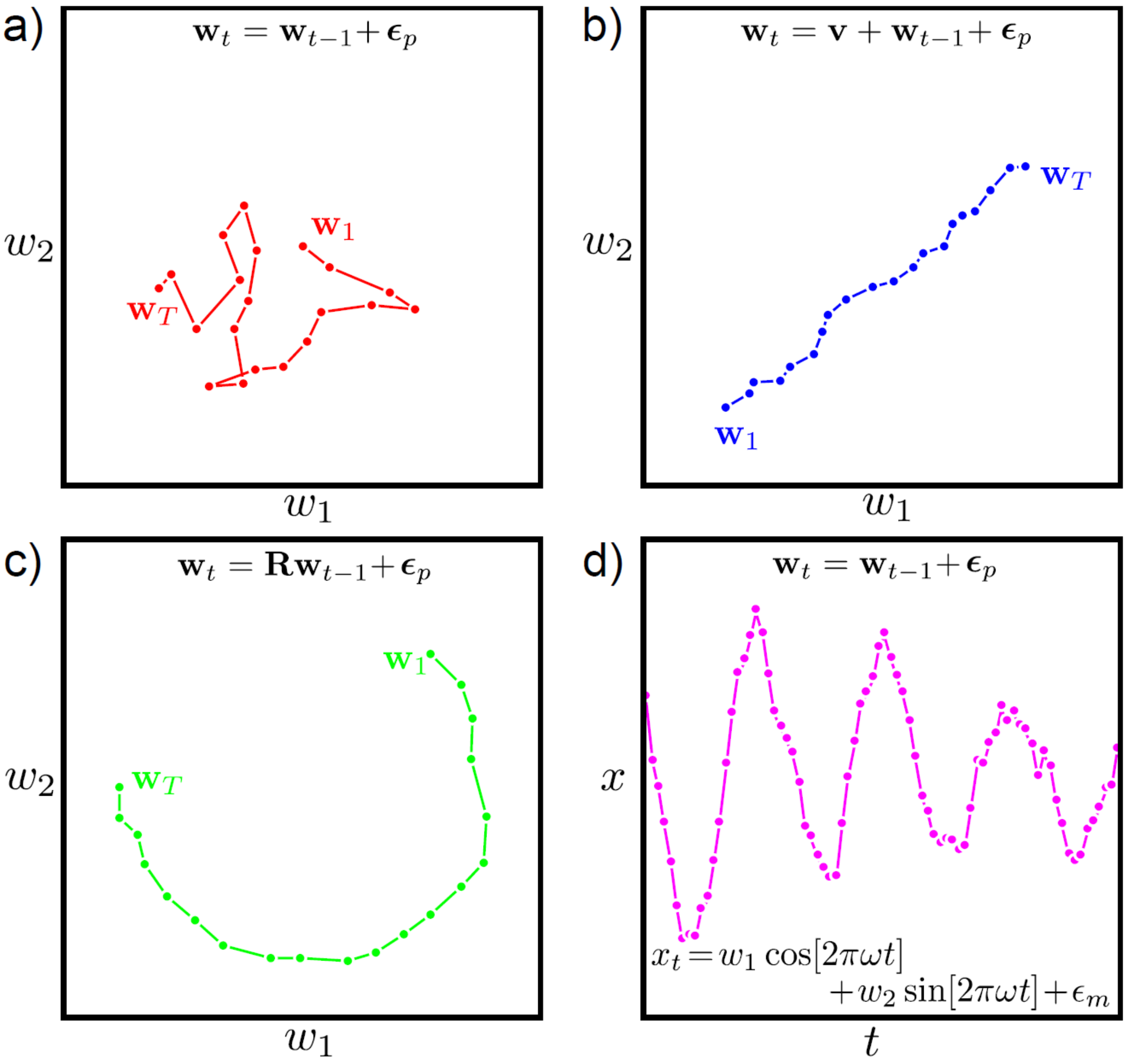
$$\begin{aligned} \boldsymbol{\mu}_{t|T} &= \boldsymbol{\mu}_t + \mathbf{C}_t(\boldsymbol{\mu}_{t+1|T} - \boldsymbol{\mu}_{+|t}) \\ \boldsymbol{\Sigma}_{t|T} &= \boldsymbol{\Sigma}_t + \mathbf{C}_t(\boldsymbol{\Sigma}_{t+1|T} - \boldsymbol{\Sigma}_{+|t})\mathbf{C}_t^T \end{aligned}$$

where

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

Source: S. Prince

Temporal Models



Source: S. Prince

Learning parameters

Parameters of equations:

$$w_t = \Psi w_{t-1} + \epsilon_p$$

$$x_t = \Phi w_t + \epsilon_m$$

$$\theta = \{\Psi, \Sigma_p, \Phi, \Sigma_m, \mu_0, \Sigma_0\}$$

Expectation-Maximization

- E step:
    - Fixed interval smoothing to get (w_t)
- M step:
    - Maximize log-likelihood to get θ

M step

Equations:

$$w_t = \Psi w_{t-1} + \epsilon_p$$

$$x_t = \Phi w_t + \epsilon_m$$

Solve where *μ* are estimates for *w* (from E step):

$$\hat{\theta} = \operatorname{argmax}_\theta \log P(\mu, x|\theta) \quad P(\mu, x) = P(\mu_1) \prod_{t=2}^{T} P(\mu_t|\mu_{t-1}) \prod_{t=1}^{T} P(x_t|\mu_t)$$

$$\begin{aligned} Q = \log P(\mu, x) = & -\sum_{t=1}^{T} \left(\frac{1}{2}(x_t - \Phi\mu_t)^T \Sigma_m^{-1} (x_t - \Phi\mu_t) \right) - \frac{T}{2} \log |\Sigma_m| \\ & -\sum_{t=2}^{T} \left(\frac{1}{2}(\mu_t - \Psi\mu_{t-1})^T \Sigma_p^{-1} (\mu_t - \Psi\mu_{t-1}) \right) - \frac{T-1}{2} \log |\Sigma_p| \\ & -\left(\frac{1}{2}(\mu_1 - \mu_0)^T \Sigma_0^{-1} (\mu_1 - \mu_0) \right) - \frac{1}{2} \log |\Sigma_0| + const. \end{aligned}$$

M step

$$\frac{\partial}{\partial \Phi} Q = \frac{\partial}{\partial \Phi} \left\{ -\sum_{t=1}^{T} \left(\frac{1}{2} (x_t - \Phi \mu_t)^T \Sigma_m^{-1} (x_t - \Phi \mu_t) \right) \right\}$$

$$= \sum_{t=1}^{T} \Sigma_m^{-1} (x_t - \Phi \mu_t) \mu_t^T$$

$$\frac{\partial (\mathbf{X}\mathbf{b} + \mathbf{c})^T \mathbf{D} (\mathbf{X}\mathbf{b} + \mathbf{c})}{\partial \mathbf{X}} = (\mathbf{D} + \mathbf{D}^T)(\mathbf{X}\mathbf{b} + \mathbf{c})\mathbf{b}^T.$$

M step

$$\frac{\partial}{\partial \Phi} Q = \frac{\partial}{\partial \Phi} \left\{ - \sum_{t=1}^{T} \left(\frac{1}{2} (x_t - \Phi \mu_t)^T \Sigma_m^{-1} (x_t - \Phi \mu_t) \right) \right\}$$

$$= \sum_{t=1}^{T} \Sigma_m^{-1} (x_t - \Phi \mu_t) \mu_t^T = \Sigma_m^{-1} \left(\sum_{t=1}^{T} x_t \mu_t^T - \Phi \sum_{t=1}^{T} \mu_t \mu_t^T \right) = 0$$

$$\Phi = \frac{\sum_{t=1}^{T} x_t \mu_t^T}{\sum_{t=1}^{T} \mu_t \mu_t^T}$$

In the same way:

$$\Psi = \frac{\sum_{t=2}^{T} \mu_t \mu_{t-1}^T}{\sum_{t=2}^{T} \mu_{t-1} \mu_{t-1}^T}$$

M step

$$\frac{\partial}{\partial \Sigma_m^{-1}} Q = \frac{\partial}{\partial \Sigma_m^{-1}} \left\{ -\sum_{t=1}^{T} \left(\frac{1}{2} (x_t - \Phi\mu_t)^T \Sigma_m^{-1} (x_t - \Phi\mu_t) \right) + \frac{T}{2} \log |\Sigma_m^{-1}| \right\}$$

$$= -\sum_{t=1}^{T} \left(\frac{1}{2} (x_t - \Phi\mu_t)(x_t - \Phi\mu_t)^T \right) + \frac{T}{2} \Sigma_m$$

$$\frac{\partial \mathbf{a}^T \mathbf{X} \mathbf{b}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T \qquad \frac{\partial}{\partial A} \log(|A|) = (A^{-1})^T$$

M step

$$\frac{\partial}{\partial \Sigma_m^{-1}} Q = \frac{\partial}{\partial \Sigma_m^{-1}} \left\{ - \sum_{t=1}^{T} \left(\frac{1}{2} (x_t - \Phi \mu_t)^T \Sigma_m^{-1} (x_t - \Phi \mu_t) \right) + \frac{T}{2} \log |\Sigma_m^{-1}| \right\}$$

$$= - \sum_{t=1}^{T} \left(\frac{1}{2} (x_t - \Phi \mu_t)(x_t - \Phi \mu_t)^T \right) + \frac{T}{2} \Sigma_m$$

$$= - \left(\sum_{t=1}^{T} x_t x_t^T - 2 \sum_{t=1}^{T} \Phi \mu_t x_t^T + \Phi \left(\sum_{t=1}^{T} \mu_t \mu_t^T \right) \Phi^T \right) + T \Sigma_m$$

$$= - \left(\sum_{t=1}^{T} x_t x_t^T - 2 \sum_{t=1}^{T} \Phi \mu_t x_t^T + \Phi \sum_{t=1}^{T} \mu_t x_t^T \right) + T \Sigma_m = 0$$

$$\boxed{\Phi = \frac{\sum_{t=1}^{T} x_t \mu_t^T}{\sum_{t=1}^{T} \mu_t \mu_t^T}}$$

$$\Sigma_m = \frac{1}{T} \sum_{t=1}^{T} \left(x_t x_t^T - \Phi \mu_t x_t^T \right)$$

$$\Sigma_p = \frac{1}{T-1} \sum_{t=2}^{T} \left(\mu_t \mu_t^T - \Psi \mu_{t-1} \mu_t^T \right)$$

Problems with the Kalman filter

- Requires linear temporal and measurement equations

- Represents result as a normal distribution: what if the posterior is genuinely multi-modal?

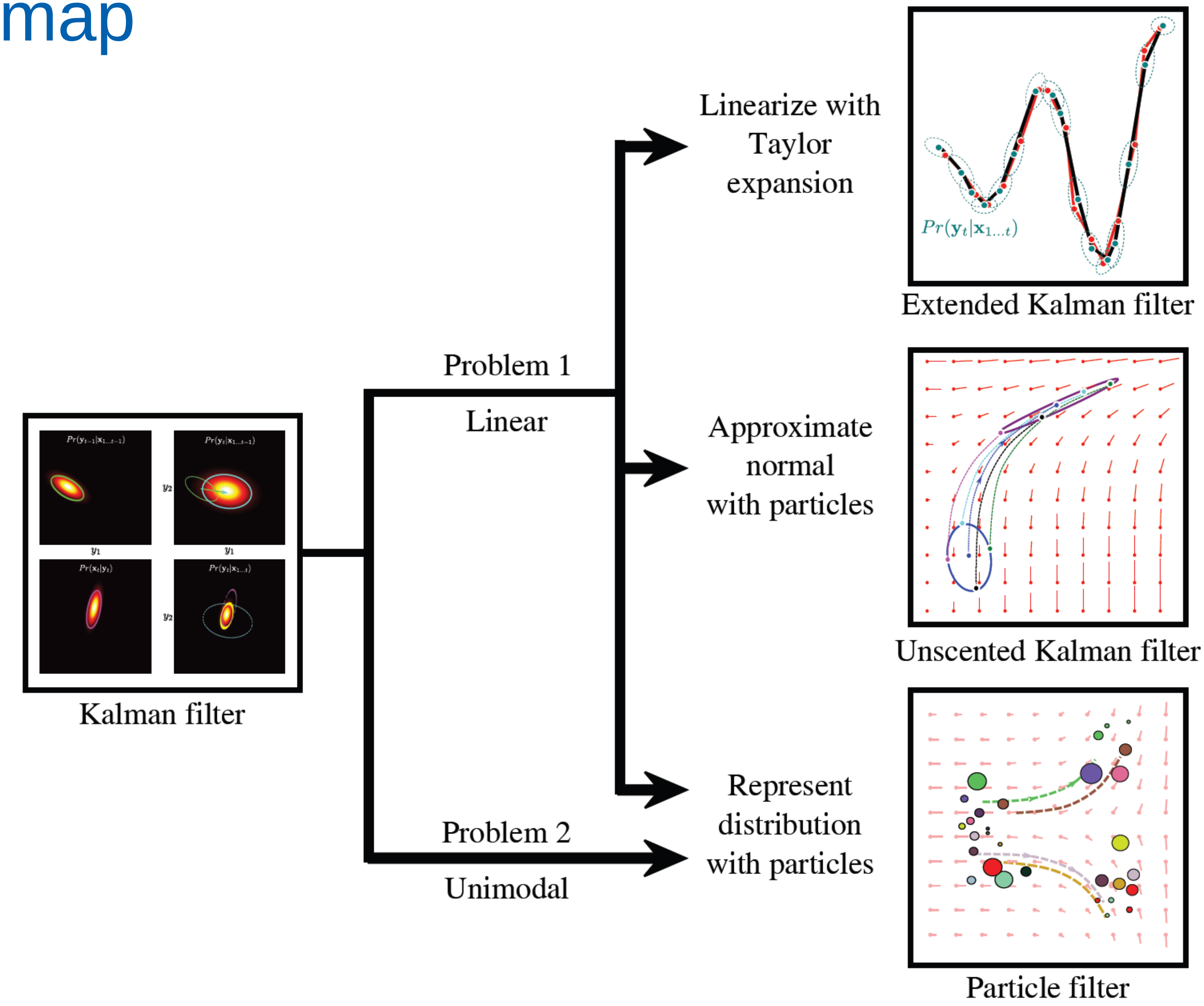
Source: S. Prince

Structure

- Temporal models
- Kalman filter
- Extended Kalman filter
- Unscented Kalman filter
- Particle filters
- Applications

Source: S. Prince

Roadmap



Kalman filter

Problem 1

Linear

Problem 2

Unimodal

Linearize with Taylor expansion

Extended Kalman filter

Approximate normal with particles

Unscented Kalman filter

Represent distribution with particles

Particle filter

Source: S. Prince

Extended Kalman Filter

Allows non-linear measurement and temporal equations

$$\mathbf{w}_t = \mathbf{f}[\mathbf{w}_{t-1}, \boldsymbol{\epsilon}_p]$$
$$\mathbf{x}_t = \mathbf{g}[\mathbf{w}_t, \boldsymbol{\epsilon}_m]$$

Key idea: take Taylor expansion and treat as locally linear

Source: S. Prince

Jacobians

Based on Jacobians matrices of derivatives

$$\begin{aligned}
\boldsymbol{\Psi} &= \left.\frac{\partial \mathbf{f}[\mathbf{w}_{t-1}, \boldsymbol{\epsilon}_p]}{\partial \mathbf{w}_{t-1}}\right|_{\boldsymbol{\mu}_{t-1},\mathbf{0}} \\
\boldsymbol{\Upsilon}_p &= \left.\frac{\partial \mathbf{f}[\mathbf{w}_{t-1}, \boldsymbol{\epsilon}_p]}{\partial \boldsymbol{\epsilon}_p}\right|_{\boldsymbol{\mu}_{t-1},\mathbf{0}} \\
\boldsymbol{\Phi} &= \left.\frac{\partial \mathbf{g}[\mathbf{w}_t, \boldsymbol{\epsilon}_m]}{\partial \mathbf{w}_t}\right|_{\boldsymbol{\mu}_+,\mathbf{0}} \\
\boldsymbol{\Upsilon}_m &= \left.\frac{\partial \mathbf{g}[\mathbf{w}_t, \boldsymbol{\epsilon}_m]}{\partial \boldsymbol{\epsilon}_m}\right|_{\boldsymbol{\mu}_+,\mathbf{0}},
\end{aligned}$$

Source: S. Prince

Extended Kalman Filter Equations

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

where

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

Source: S. Prince

Comparison: Kalman Filter

Time evolution equation

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

Measurement equation

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

Inference

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

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

Source: S. Prince

Extended Kalman filter

$$\mathbf{w}_t = f(\mathbf{w}_{t-1}, \epsilon_p) = \begin{pmatrix} w_{1,t-1} + \epsilon_{1,p} \\ w_{1,t-1}\sin(w_{1,t-1}) + \epsilon_{2,p} \end{pmatrix}$$

$$\mathbf{x}_t = g(\mathbf{w}_t, \boldsymbol{\epsilon}_m) = \mathbf{w}_t + \boldsymbol{\epsilon}_m$$

$$\boldsymbol{\Psi} = \frac{\partial}{\partial \mathbf{w}} \begin{pmatrix} w_1 + \epsilon_{1,p} \\ w_1 \sin(w_1) + \epsilon_{2,p} \end{pmatrix} \Bigg|_{\mu_{t-1},0}$$

$$\boldsymbol{\Upsilon}_p = \frac{\partial}{\partial \boldsymbol{\epsilon}_p} \begin{pmatrix} w_1 + \epsilon_{1,p} \\ w_1 \sin(w_1) + \epsilon_{2,p} \end{pmatrix} \Bigg|_{\mu_{t-1},0}$$

$$\boldsymbol{\Phi} = \frac{\partial}{\partial \mathbf{w}} \begin{pmatrix} w_1 + \epsilon_{1,m} \\ w_2 + \epsilon_{2,m} \end{pmatrix} \Bigg|_{\mu_{+},0}$$

$$\boldsymbol{\Upsilon}_m = \frac{\partial}{\partial \boldsymbol{\epsilon}_m} \begin{pmatrix} w_1 + \epsilon_{1,m} \\ w_2 + \epsilon_{2,m} \end{pmatrix} \Bigg|_{\mu_{+},0}$$

$$\boxed{\mathbf{y} = f(\mathbf{x}) \qquad \frac{\partial f}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_M} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_N}{\partial x_1} & \cdots & \frac{\partial f_N}{\partial x_M} \end{pmatrix}}$$

Extended Kalman filter

$$\mathbf{w}_t = f(\mathbf{w}_{t-1}, \epsilon_p) = \begin{pmatrix} w_{1,t-1} + \epsilon_{1,p} \\ w_{1,t-1}\sin(w_{1,t-1}) + \epsilon_{2,p} \end{pmatrix}$$

$$\mathbf{x}_t = g(\mathbf{w}_t, \boldsymbol{\epsilon}_m) = \mathbf{w}_t + \boldsymbol{\epsilon}_m$$

$$\boldsymbol{\Psi} = \frac{\partial}{\partial \mathbf{w}} \begin{pmatrix} w_1 + \epsilon_{1,p} \\ w_1 \sin(w_1) + \epsilon_{2,p} \end{pmatrix}\Bigg|_{\mu_{t-1},0} = \begin{pmatrix} 1 & 0 \\ \sin(\mu_{1,t-1}) + \mu_{1,t-1}\cos(\mu_{1,t-1}) & 0 \end{pmatrix}$$

$$\boldsymbol{\Upsilon}_p = \frac{\partial}{\partial \boldsymbol{\epsilon}_p} \begin{pmatrix} w_1 + \epsilon_{1,p} \\ w_1 \sin(w_1) + \epsilon_{2,p} \end{pmatrix}\Bigg|_{\mu_{t-1},0}$$

$$\boldsymbol{\Phi} = \frac{\partial}{\partial \mathbf{w}} \begin{pmatrix} w_1 + \epsilon_{1,m} \\ w_2 + \epsilon_{2,m} \end{pmatrix}\Bigg|_{\mu_+,0}$$

$$\boldsymbol{\Upsilon}_m = \frac{\partial}{\partial \boldsymbol{\epsilon}_m} \begin{pmatrix} w_1 + \epsilon_{1,m} \\ w_2 + \epsilon_{2,m} \end{pmatrix}\Bigg|_{\mu_+,0}$$

$$\boxed{\mathbf{y} = f(\mathbf{x}) \qquad \frac{\partial f}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_M} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_N}{\partial x_1} & \cdots & \frac{\partial f_N}{\partial x_M} \end{pmatrix}}$$

Extended Kalman filter

$$\mathbf{w}_t = f(\mathbf{w}_{t-1}, \epsilon_p) = \begin{pmatrix} w_{1,t-1} + \epsilon_{1,p} \\ w_{1,t-1}\sin(w_{1,t-1}) + \epsilon_{2,p} \end{pmatrix}$$

$$\mathbf{x}_t = g(\mathbf{w}_t, \boldsymbol{\epsilon}_m) = \mathbf{w}_t + \boldsymbol{\epsilon}_m$$

$$\mathbf{\Psi} = \frac{\partial}{\partial \mathbf{w}} \begin{pmatrix} w_1 + \epsilon_{1,p} \\ w_1\sin(w_1) + \epsilon_{2,p} \end{pmatrix} \Bigg|_{\mu_{t-1},0} = \begin{pmatrix} 1 & 0 \\ \sin(\mu_{1,t-1}) + \mu_{1,t-1}\cos(\mu_{1,t-1}) & 0 \end{pmatrix}$$

$$\mathbf{\Upsilon}_p = \frac{\partial}{\partial \boldsymbol{\epsilon}_p} \begin{pmatrix} w_1 + \epsilon_{1,p} \\ w_1\sin(w_1) + \epsilon_{2,p} \end{pmatrix} \Bigg|_{\mu_{t-1},0} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\mathbf{\Phi} = \frac{\partial}{\partial \mathbf{w}} \begin{pmatrix} w_1 + \epsilon_{1,m} \\ w_2 + \epsilon_{2,m} \end{pmatrix} \Bigg|_{\mu_{+},0} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\mathbf{\Upsilon}_m = \frac{\partial}{\partial \boldsymbol{\epsilon}_m} \begin{pmatrix} w_1 + \epsilon_{1,m} \\ w_2 + \epsilon_{2,m} \end{pmatrix} \Bigg|_{\mu_{+},0} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

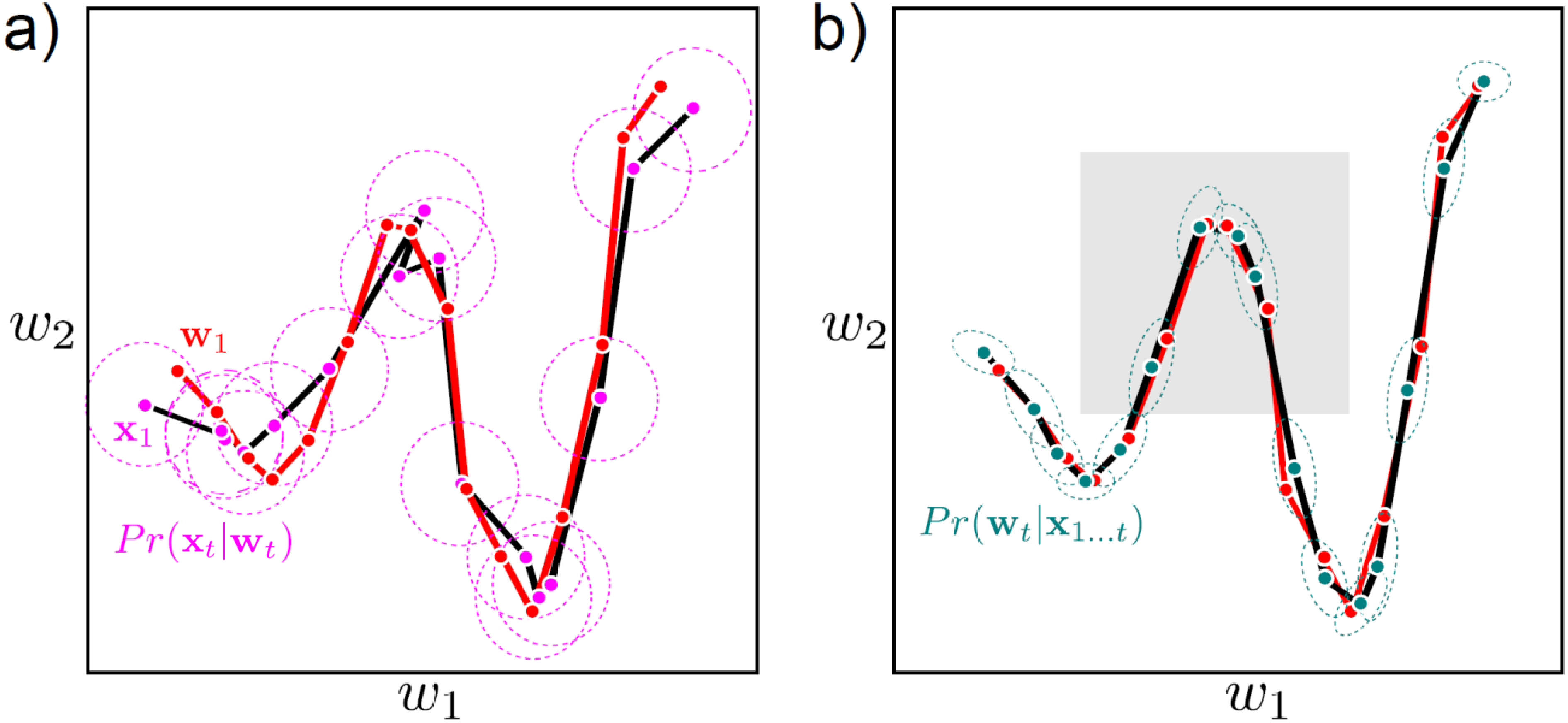
$$\boxed{\mathbf{y} = f(\mathbf{x}) \quad \frac{\partial f}{\partial \mathbf{x}} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_M} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_N}{\partial x_1} & \cdots & \frac{\partial f_N}{\partial x_M} \end{pmatrix}}$$

Extended Kalman Filter

$$\mathbf{w}_t = f(\mathbf{w}_{t-1}, \epsilon_p) = \begin{pmatrix} w_{1,t-1} + \epsilon_{1,p} \\ w_{1,t-1}\sin(w_{1,t-1}) + \epsilon_{2,p} \end{pmatrix}$$

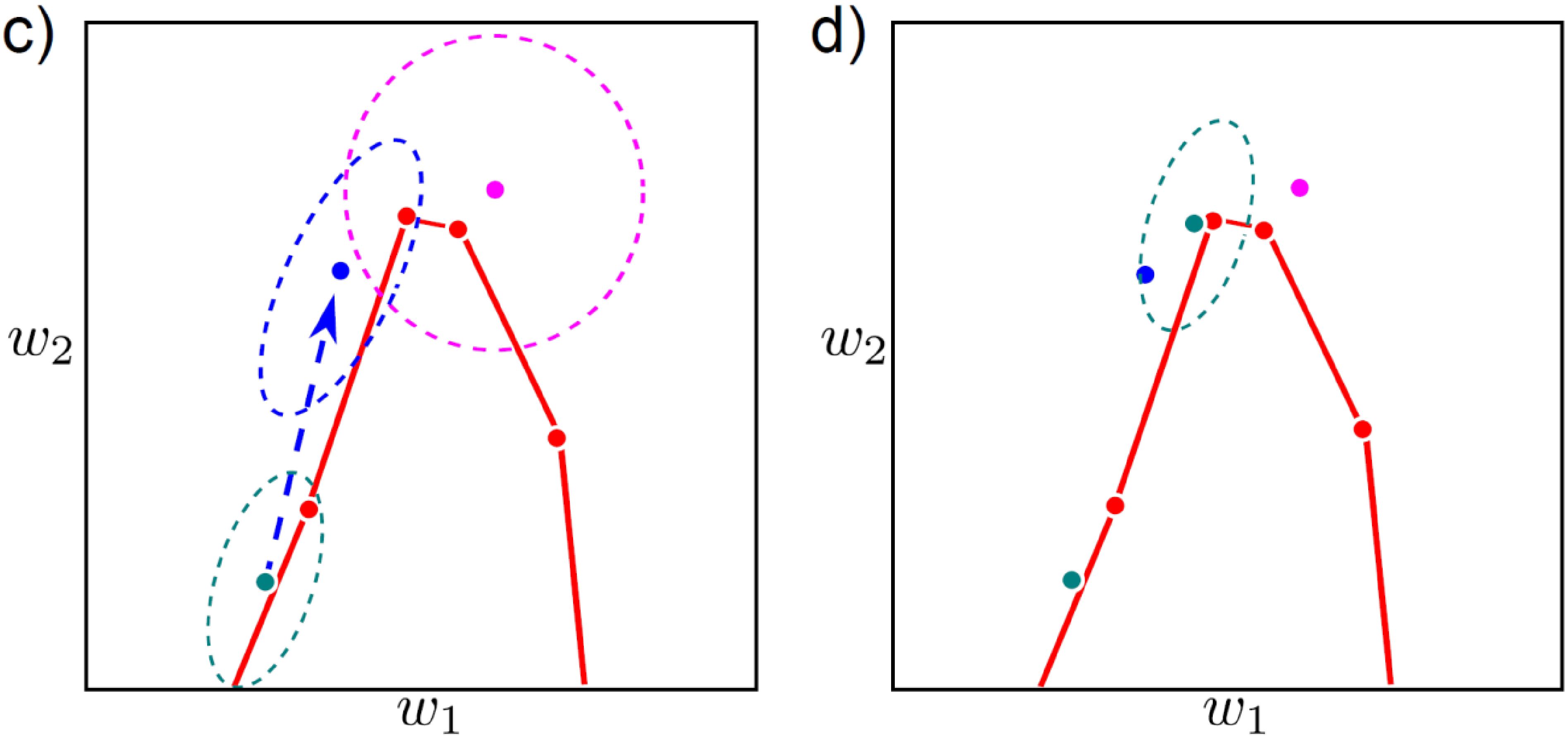
$$\mathbf{x}_t = g(\mathbf{w}_t, \boldsymbol{\epsilon}_m) = \mathbf{w}_t + \boldsymbol{\epsilon}_m$$

Red: True State, Magenta: Observations, Green: Estimate (Extended Kalman)



Extended Kalman Filter

Red: True State, Magenta: Likelihood, Green: Posterior, Blue: Prediction



Iterated extended Kalman filter

Q passes, first pass

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

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

q-th pass

$$\boldsymbol{\Phi}^q = \left.\frac{\partial \mathbf{g}[\mathbf{w}_t,\boldsymbol{\epsilon}_m]}{\partial \mathbf{w}_t}\right|_{\boldsymbol{\mu}_t^{q-1},\mathbf{0}} \qquad \boldsymbol{\Upsilon}_m^q = \left.\frac{\partial \mathbf{g}[\mathbf{w}_t,\boldsymbol{\epsilon}_m]}{\partial \boldsymbol{\epsilon}_m}\right|_{\boldsymbol{\mu}_t^{q-1},\mathbf{0}}$$

Backward passes can be included

Iterated extended Kalman filter

Algorithm 19.4: The iterated extended Kalman filter

```
Input : Measurements $\{\mathbf{x}\}_{t=1}^{T}$, temporal function $\mathbf{f}[\bullet,\bullet]$, measurement function $\mathbf{g}[\bullet,\bullet]$
Output: Means $\{\boldsymbol{\mu}_t\}_{t=1}^{T}$ and covariances $\{\boldsymbol{\Sigma}_t\}_{t=1}^{T}$ of marginal posterior distributions
begin
    // Initialize mean and covariance
    $\boldsymbol{\mu}_0 = \mathbf{0}$
    $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}_0$                    // Typically set to large multiple of identity
    // For each time step
    for $t$=1 to $T$ do
        // State prediction
        $\boldsymbol{\mu}_+ = \mathbf{f}[\boldsymbol{\mu}_{t-1}, \mathbf{0}]$
        // Covariance prediction
        $\boldsymbol{\Sigma}_+ = \boldsymbol{\Psi}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Psi}^T + \boldsymbol{\Upsilon}_p\boldsymbol{\Sigma}_p\boldsymbol{\Upsilon}_p^T$
        // For each iteration
        for $q$=0 to $Q$ do
            // Compute Kalman gain
            $\mathbf{K} = \boldsymbol{\Sigma}_+\boldsymbol{\Phi}^{qT}(\boldsymbol{\Upsilon}_m^q\boldsymbol{\Sigma}_m\boldsymbol{\Upsilon}_m^{qT} + \boldsymbol{\Phi}^q\boldsymbol{\Sigma}_+\boldsymbol{\Phi}^{qT})^{-1}$
            // State update
            $\boldsymbol{\mu}_t^q = \boldsymbol{\mu}_+ + \mathbf{K}(\mathbf{x}_t - \mathbf{g}[\boldsymbol{\mu}_+, \mathbf{0}])$
            // Covariance update
            $\boldsymbol{\Sigma}_t^q = (\mathbf{I} - \mathbf{K}\boldsymbol{\Phi}^q)\boldsymbol{\Sigma}_+$
        end
    end
end
```

Problems with EKF

Consider flow field (red) of function f

- Mean is updated by non-linear function without noise

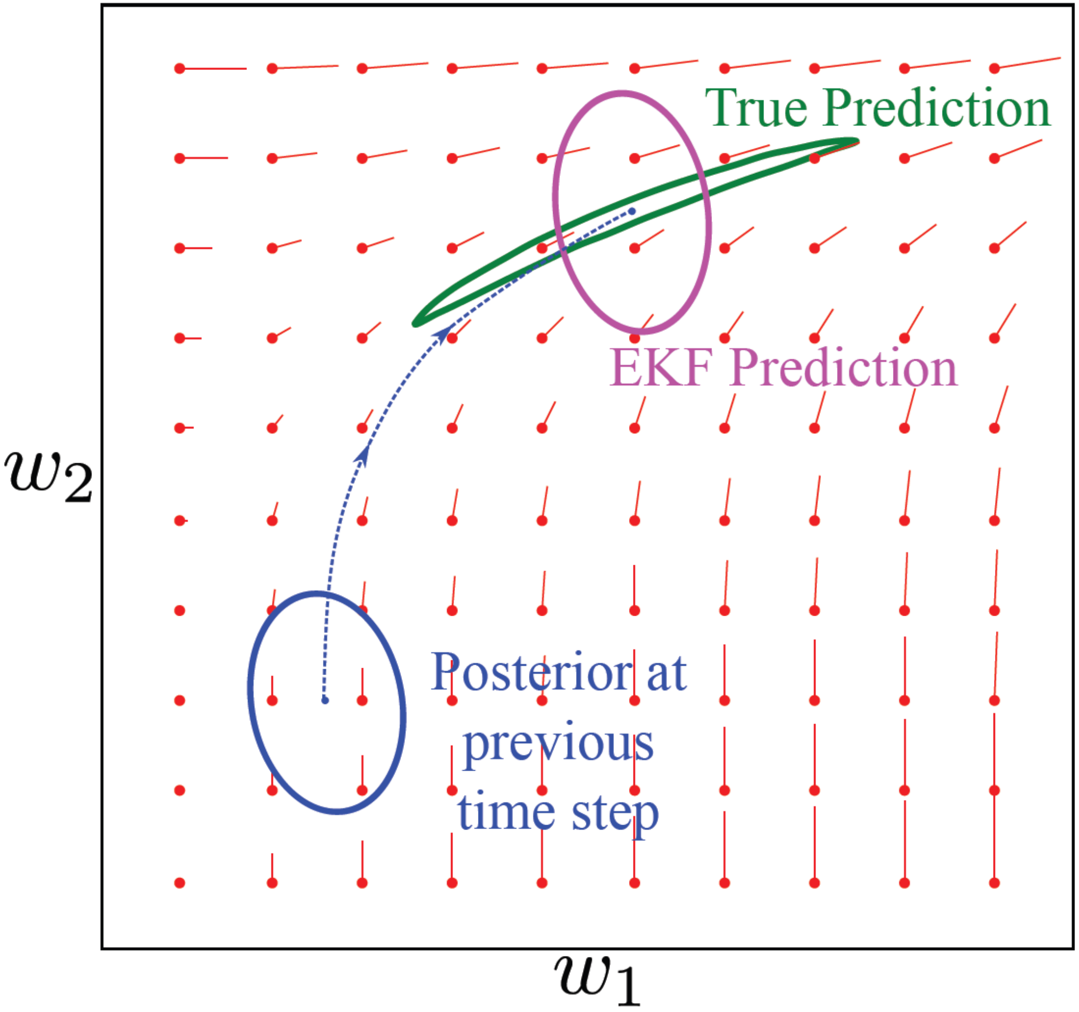
$$\boldsymbol{\mu}_+ = \mathbf{f}[\boldsymbol{\mu}_{t-1}, \mathbf{0}]$$

- Covariance is linearly approximated

$$\boldsymbol{\Sigma}_+ = \boldsymbol{\Psi}\boldsymbol{\Sigma}_{t-1}\boldsymbol{\Psi}^T + \boldsymbol{\Upsilon}_p\boldsymbol{\Sigma}_p\boldsymbol{\Upsilon}_p^T$$

- Predicted covariance is poor

How can this be fixed?



Source: S. Prince

Structure

- Temporal models
- Kalman filter
- Extended Kalman filter
- Unscented Kalman filter
- Particle filters
- Applications

Source: S. Prince

Unscented Kalman Filter

Key ideas:

- Approximate distribution as a sum of weighted particles with correct mean and covariance

- Pass particles through non-linear function of the form

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

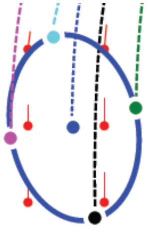
- Compute mean and covariance of transformed variables

Source: S. Prince

Unscented Kalman Filter

Approximate with particles where D_w is dimensionality of state:

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



weights

particles or sigma points

Choose so that

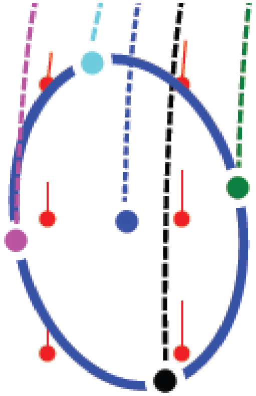
$$\sum_{j=0}^{2D_w} a_j = 1 \qquad \delta(x) = \begin{cases} 1 & x = 0 \\ 0 & \text{otherwise} \end{cases}$$

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

Source: S. Prince

One possible scheme

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



With:

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

Parameter

$$a_0 \in [0,1]$$

Source: S. Prince

One possible scheme

$$\sum_{j=0}^{2D_w} a_j \hat{w}^j = a_0 \mu_{t-1} + \frac{1-a_0}{2D_w} 2D_w \mu_{t-1} = \mu_{t-1}$$

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

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

$$\sum_{j=0}^{2D_w} a_j (\hat{w}^j - \mu_{t-1})(\hat{w}^j - \mu_{t-1})^T$$

$$= \frac{1-a_0}{2D_w} 2 \sum_{j=1}^{D_w} \left(\sqrt{\frac{D_w}{1-a_0}} \Sigma_{t-1}^{1/2} e_j \right) \left(\sqrt{\frac{D_w}{1-a_0}} \Sigma_{t-1}^{1/2} e_j \right)^T$$

$$= \Sigma_{t-1}$$

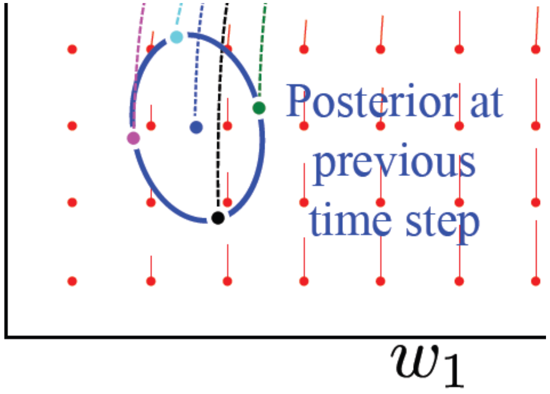
Parameter

Parameter a_0

$$\hat{\mathbf{w}}^{[0]} = \boldsymbol{\mu}_{t-1}$$

$$\hat{\mathbf{w}}^{[j]} = \boldsymbol{\mu}_{t-1} + \sqrt{\frac{D_{\mathbf{w}}}{1-a_0}}\boldsymbol{\Sigma}_{t-1}^{1/2}\mathbf{e}_j$$

$$\hat{\mathbf{w}}^{[D_{\mathbf{w}}+j]} = \boldsymbol{\mu}_{t-1} - \sqrt{\frac{D_{\mathbf{w}}}{1-a_0}}\boldsymbol{\Sigma}_{t-1}^{1/2}\mathbf{e}_j$$



a_0 = 0.00: 68%

a_0 = 0.75: 95.5%

$$P(\mu - \sigma < X \leq \mu + \sigma) \approx 68\%$$

$$P(\mu - 2\sigma < X \leq \mu + 2\sigma) \approx 95.5\%$$

$$P(\mu - 3\sigma < X \leq \mu + 3\sigma) \approx 99.7\%$$

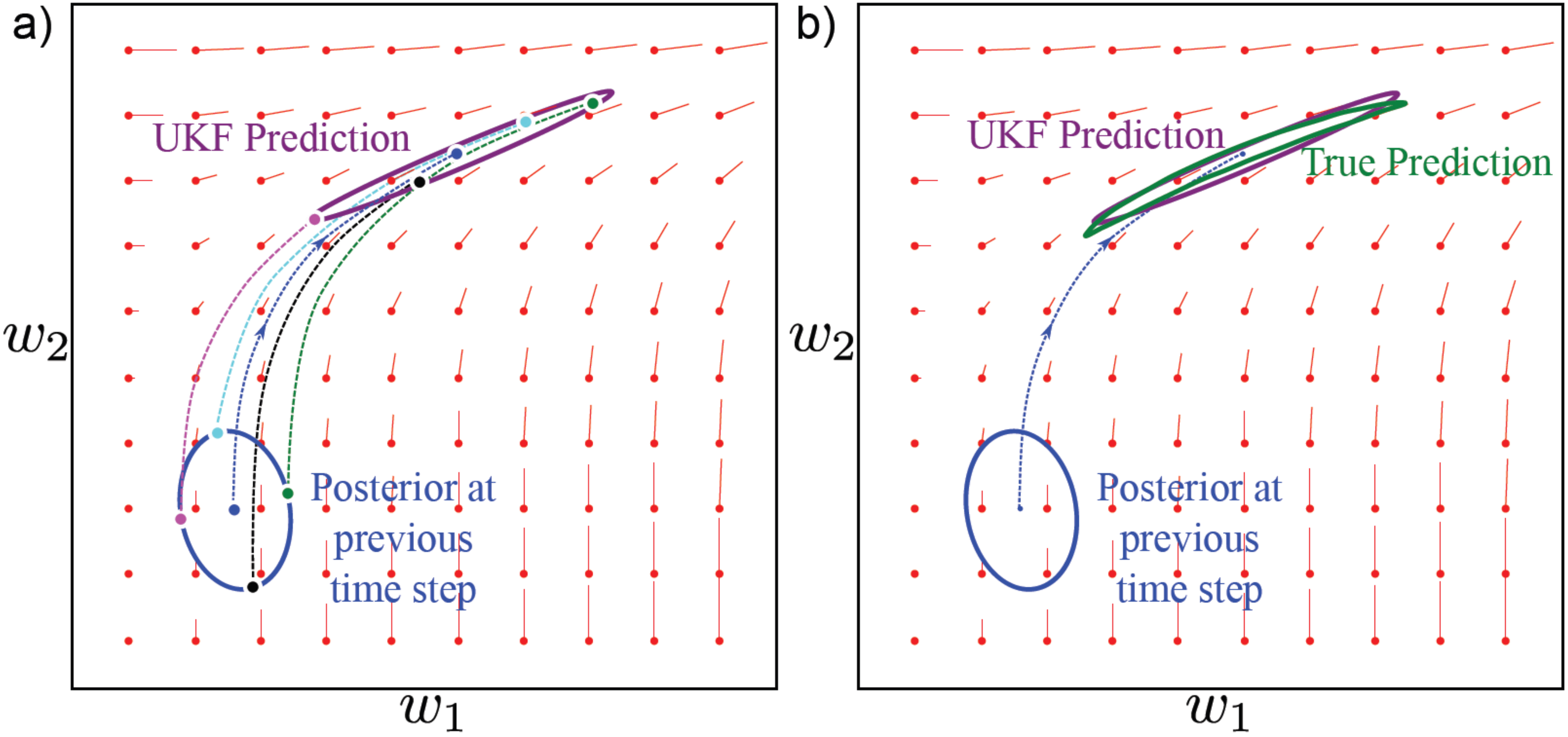
Reconstitution

Non-linear prediction of sigma points

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

Source: S. Prince

Unscented Kalman Filter



Source: S. Prince

Measurement incorporation

Measurement incorporation works in a similar way: Approximate predicted distribution by set of particles

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

Particles chosen so that mean and covariance the same

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

Source: S. Prince

Measurement incorporation

Pass particles through measurement equation and recompute mean and variance:

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

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

Measurement update equations:

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_+ + \mathbf{K}\left(\mathbf{x}_t - \boldsymbol{\mu}_x\right)$$

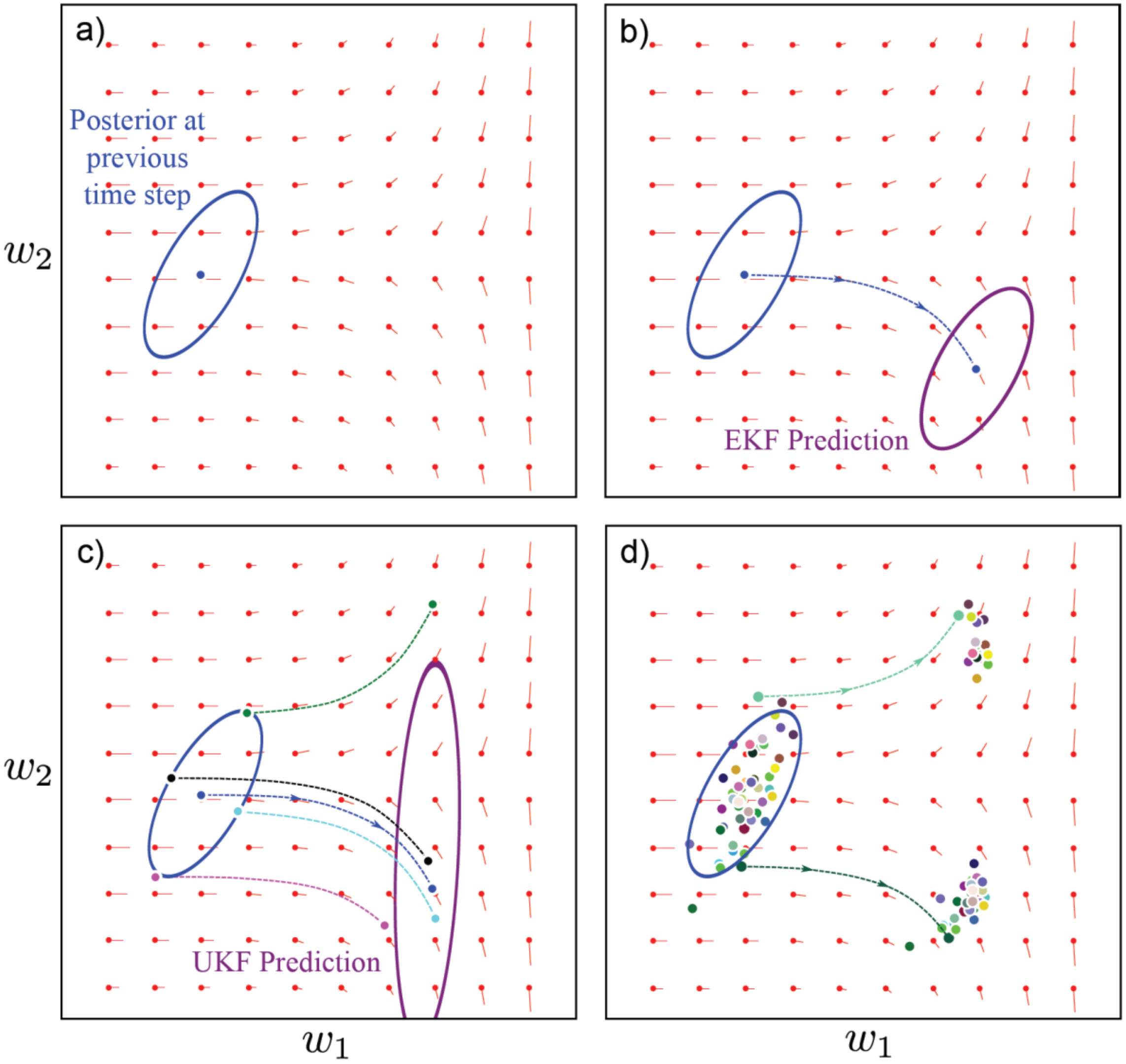
$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_+ - \mathbf{K}\boldsymbol{\Sigma}_x\mathbf{K}^T,$$

Kalman gain now computed from particles:

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

Source: S. Prince

Problems with UKF



Source: S. Prince

Structure

- Temporal models
- Kalman filter
- Extended Kalman filter
- Unscented Kalman filter
- Particle filters
- Applications

Source: S. Prince

Particle filter

Key idea:

- Represent probability distribution as a set of weighted particles

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

Advantages and disadvantages:

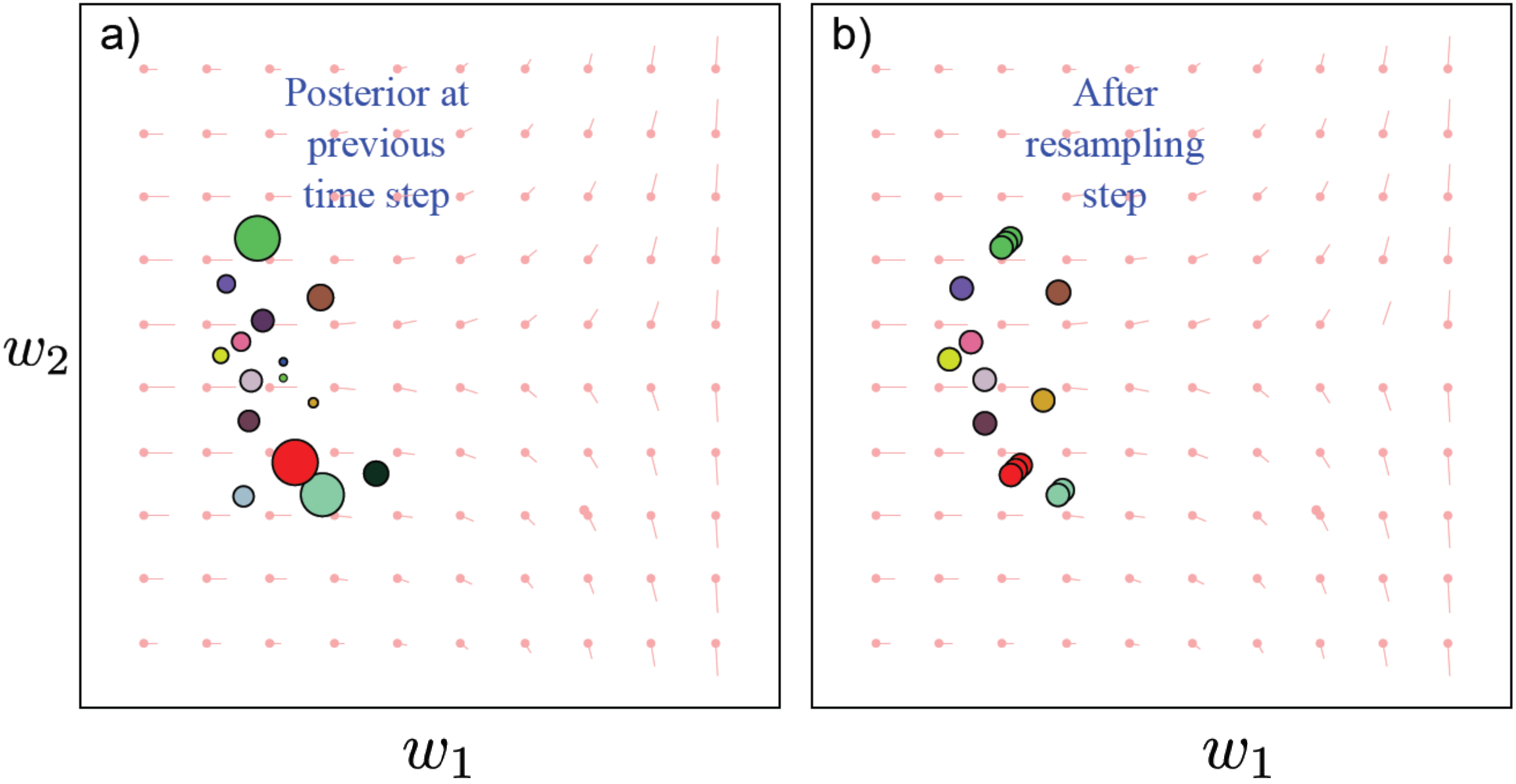
+ Can represent non-Gaussian multimodal densities
- Expensive

N. Gordon et al. **Novel approach to nonlinear/non-Gaussian Bayesian state estimation.** 1993
M. Isard and A. Blake. **Contour tracking by stochastic propagation of conditional density.** ECCV 1996

Source: S. Prince

Particle filter

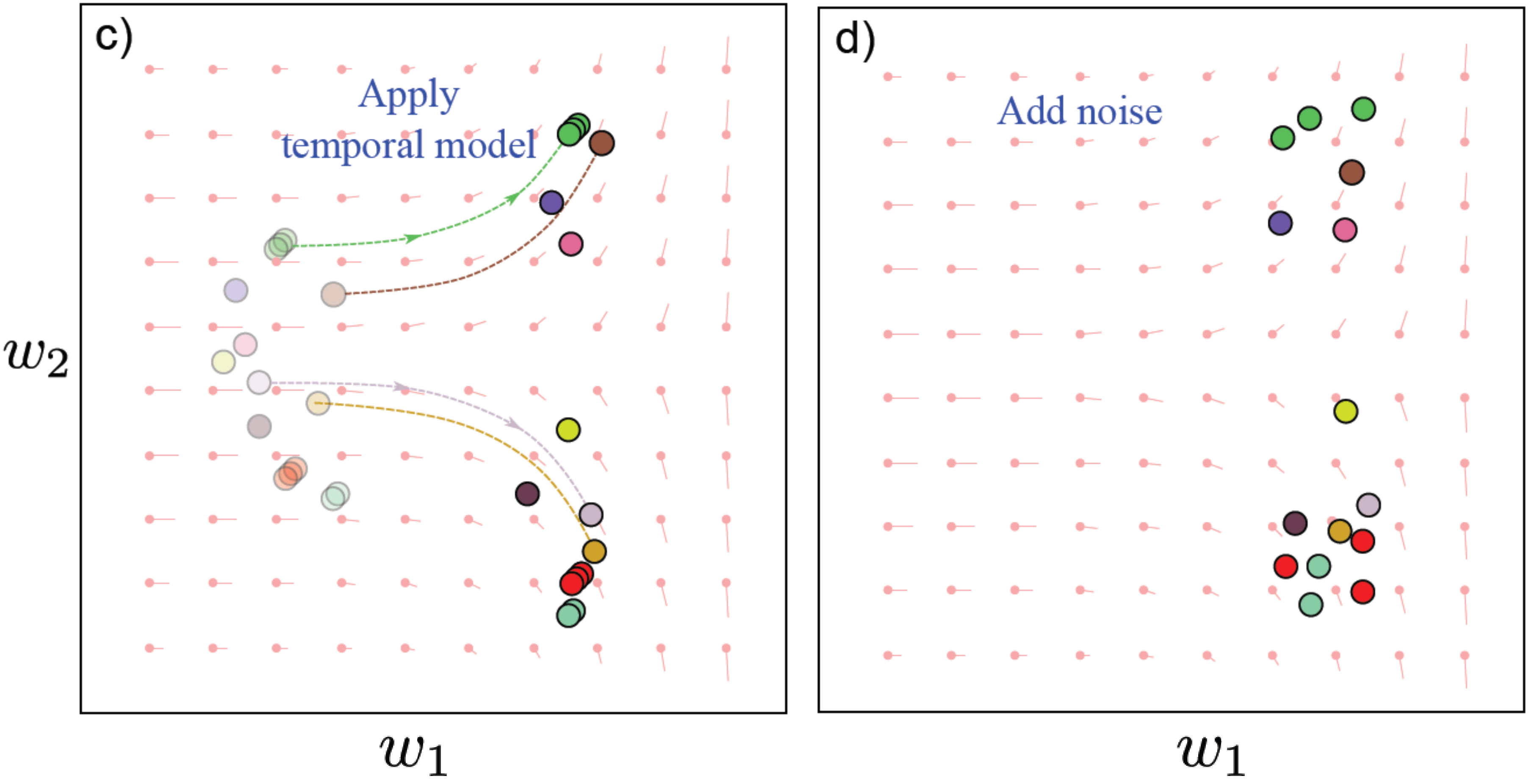
Stage 1: Resample from weighted particles according to their weight to get unweighted particles



Source: S. Prince

Particle filter

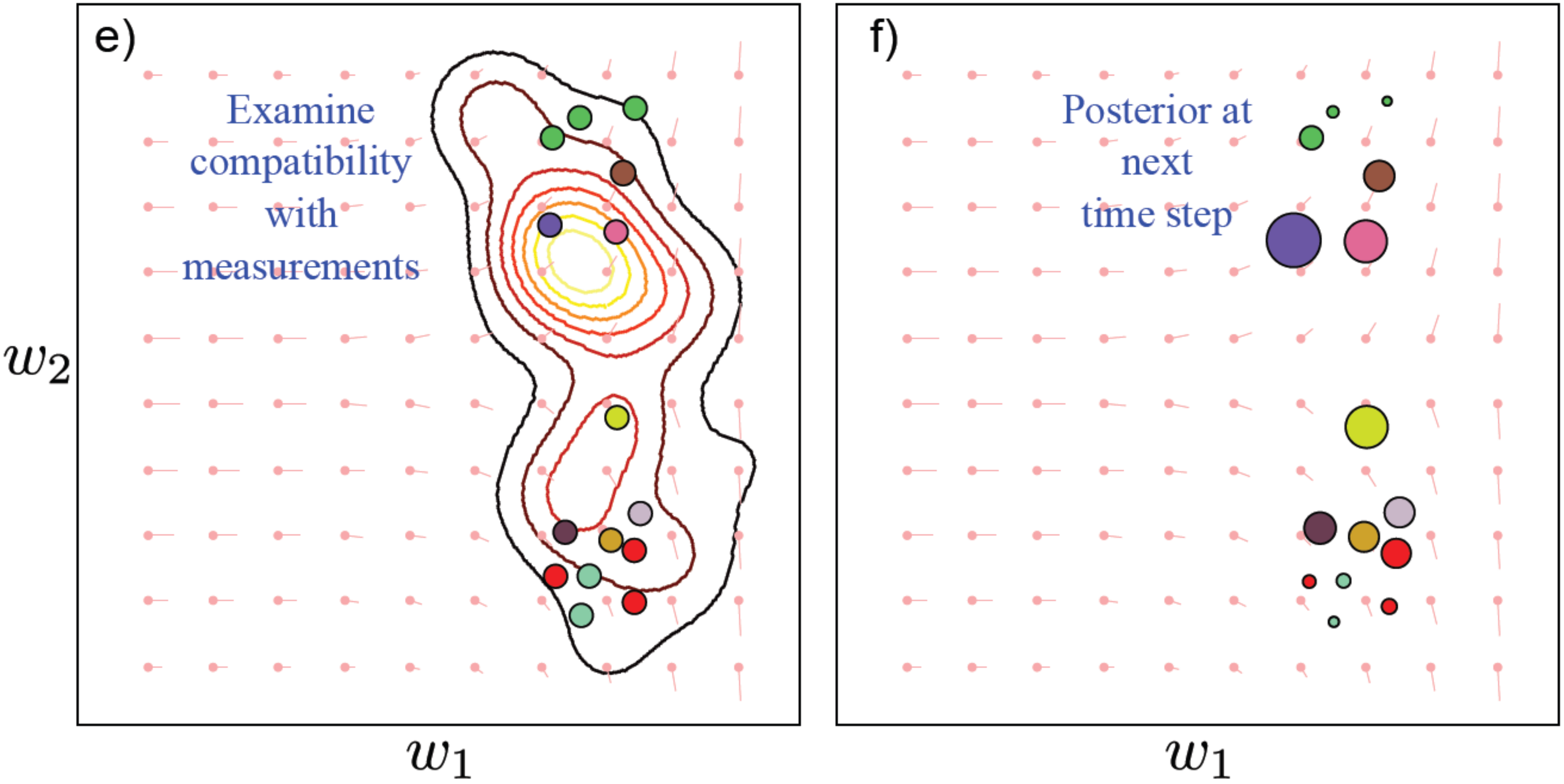
Stage 2: Pass unweighted samples through temporal model and add noise



Source: S. Prince

Particle filter

Stage 3: Weight samples by measurement density



Source: S. Prince

Filter problem as stochastic processes

- X_t Markov process
- Dynamics by kernels K_t
- Observation process with independent noise W_t (density function g_t)

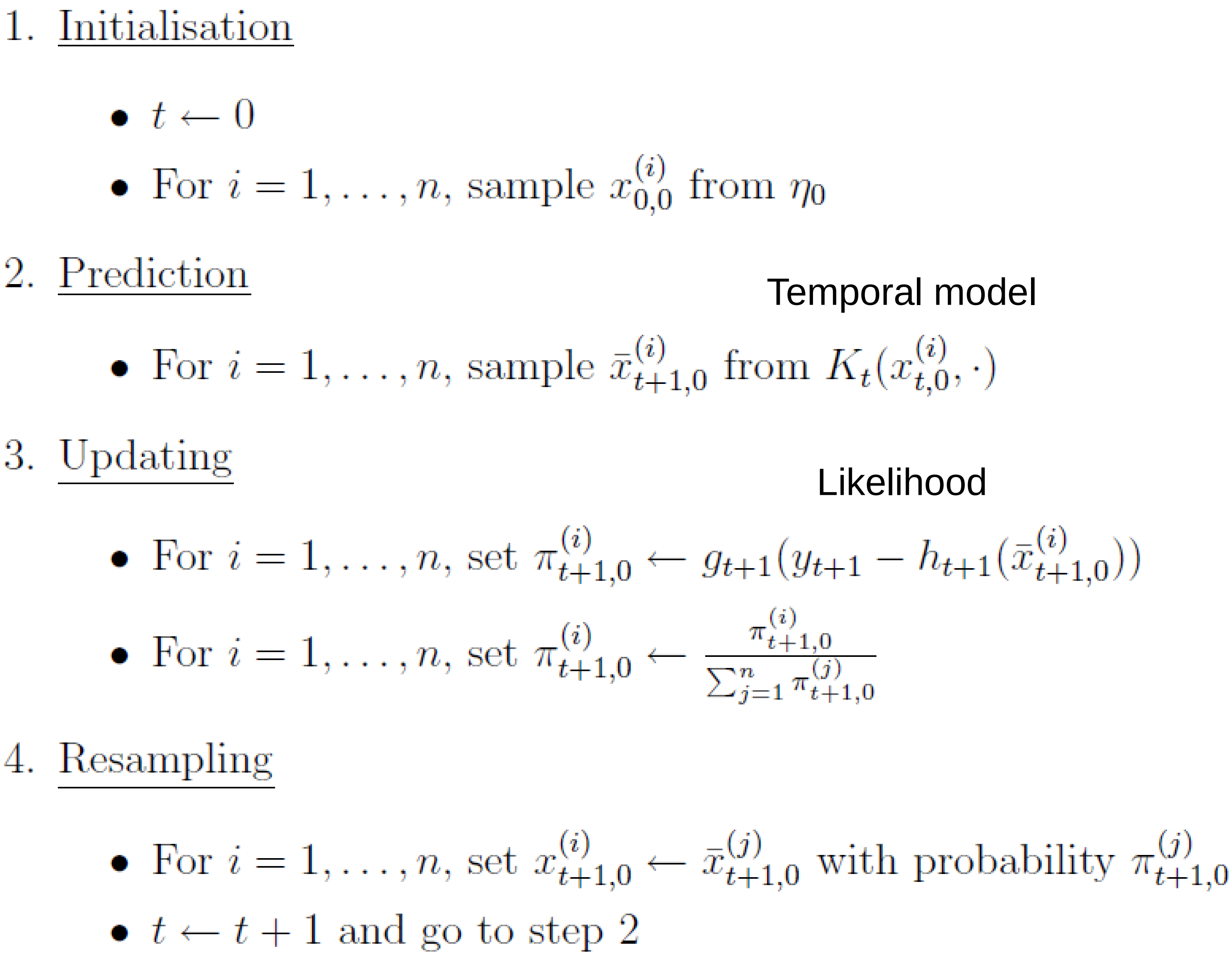
$$Y_t = h_t(X_t) + W_t$$

- Posterior:

$$\eta_t(B) := P\left(X_t \in B \mid \mathcal{G}_t\right) \quad \mathcal{G}_t \ := \ \sigma(Y_t, \ldots, Y_0)$$

- As before, only slight change of notation

Generic particle filter



1. Initialisation
   - $t \leftarrow 0$
   - For $i = 1, \ldots, n$, sample $x_{0,0}^{(i)}$ from η_0
2. Prediction
   - For $i = 1, \ldots, n$, sample $\bar{x}_{t+1,0}^{(i)}$ from $K_t(x_{t,0}^{(i)}, \cdot)$

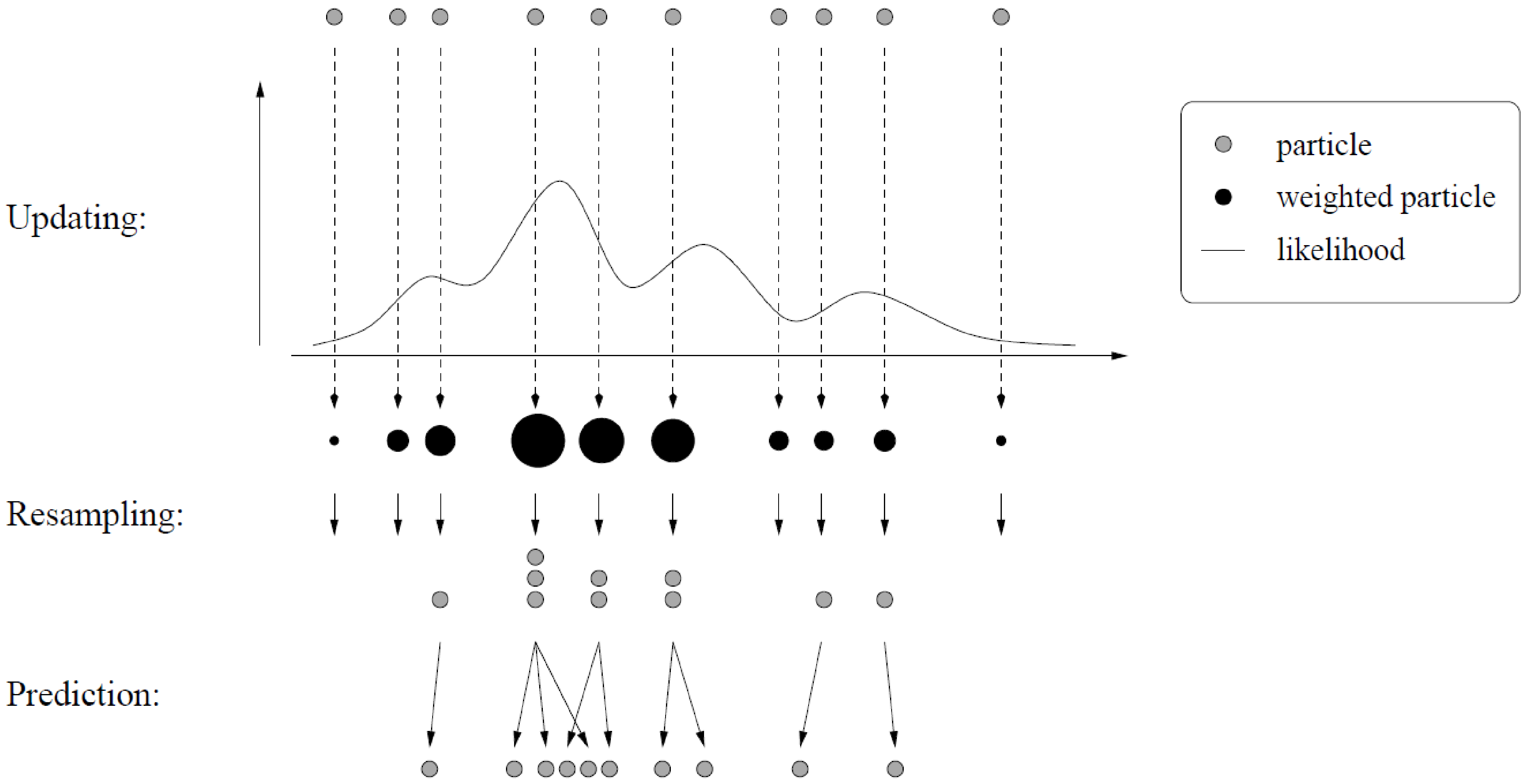
Temporal model

3. Updating
   - For $i = 1, \ldots, n$, set $\pi_{t+1,0}^{(i)} \leftarrow g_{t+1}(y_{t+1} - h_{t+1}(\bar{x}_{t+1,0}^{(i)}))$
   - For $i = 1, \ldots, n$, set $\pi_{t+1,0}^{(i)} \leftarrow \frac{\pi_{t+1,0}^{(i)}}{\sum_{j=1}^{n} \pi_{t+1,0}^{(j)}}$

Likelihood

4. Resampling
   - For $i = 1, \ldots, n$, set $x_{t+1,0}^{(i)} \leftarrow \bar{x}_{t+1,0}^{(j)}$ with probability $\pi_{t+1,0}^{(j)}$
   - $t \leftarrow t + 1$ and go to step 2

Generic particle filter

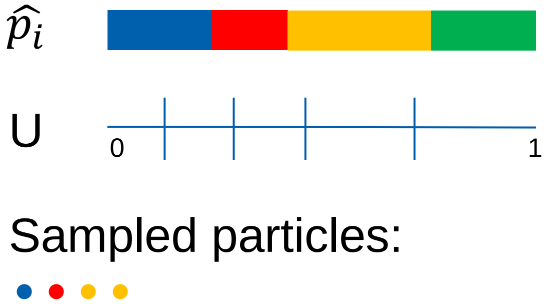


Resampling

Sample n-times with replacement using probabilities $\sum \pi_i = 1$

- Accumulate weights $\widehat{p_i} = \sum_{j \le i} \pi_j$
- Generate n random numbers [0,1)
- Sort random numbers U
- Collect samples:

```
j=0;
for(i=0; i<n; i++)
    while(U[i] > P[j]) j++;
    Particle[i] = j;
end for
```



Convergence

The approximation of the posterior by n particles:

$$\eta_t^n(\omega) := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_{t,0}^{(i)}(\omega)}$$

Converges (almost surely) to the true posterior as n goes to infinity:

$$P\left(\omega : \eta_t^n(\omega) \xrightarrow{w} \eta_t(\omega)\right) = 1$$

→ Probability 1 does not mean always!

J. Gall. **Generalised Annealed Particle Filter - Mathematical Framework, Algorithms and Applications.** 2005

Rate of convergence

Convergence rate 1/n for any bounded function *φ*

$$E\left[\left(\langle \eta_t^n, \varphi\rangle - \langle \eta_t, \varphi\rangle\right)^2\right] \leq c(\varepsilon)\frac{\|\varphi\|_\infty^2}{n}$$

If temporal model allows to reach any state in a finite number of steps, e.g.,

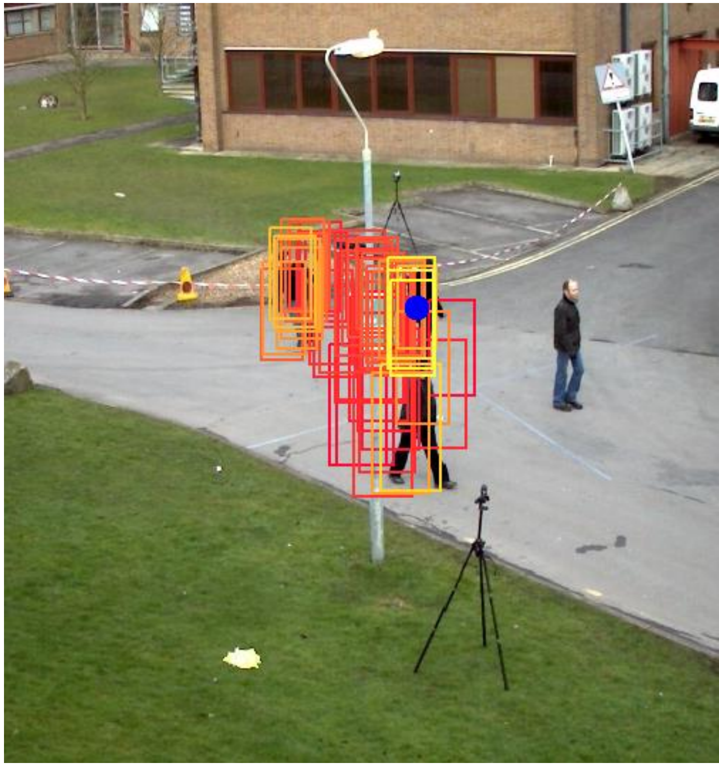
- True for bounded state space
- Not true for Gaussian noise on infinite state space

Does not depend on dimension of state space, but c(ε) is very large.

In practice, the number of particles is empirical set.

Tracking pedestrians

- State space 3D (x,y,scale)
- Use Hough transform to compute likelihood for pedestrians
- Initialization by manual marked bounding box



J. Gall et al. **On-line Adaption of Class-specific Codebooks for Instance Tracking.** BMVC 2010
J. Gall et al. **Hough Forests for Object Detection, Tracking, and Action Recognition.** TPAMI 2011

Results

Method

Results

UNIVERSITÄT BONN