Faculty of SET / School of Computer and Mathematical Sciences

COMP SCI 3007&7059 Artificial Intelligence

Probability Reasoning Over Time 3 – Kalman Filter

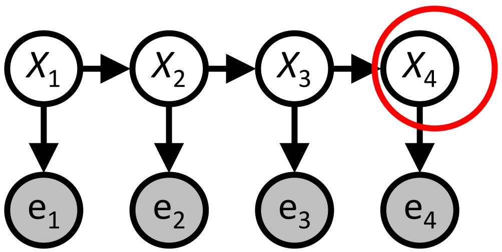
Acknowledgement of Country

We acknowledge and pay our respects to the Kaurna people, the traditional custodians whose ancestral lands we gather on.

We acknowledge the deep feelings of attachment and relationship of the Kaurna people to the country and we respect and value their past, present and ongoing connection to the land and cultural beliefs.

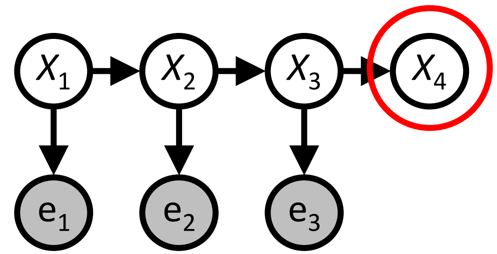
Inference tasks

Filtering: $P(X_t | e_{1:t})$

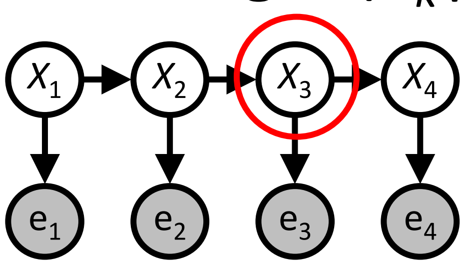


Forward

Prediction: $P(X_{t+k} | e_{1:t})$

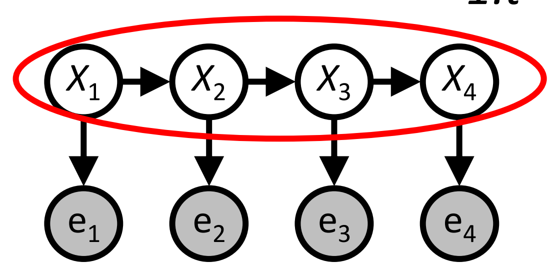


Smoothing: $P(X_k | e_{1:t})$, $k<t$



Forward- backward

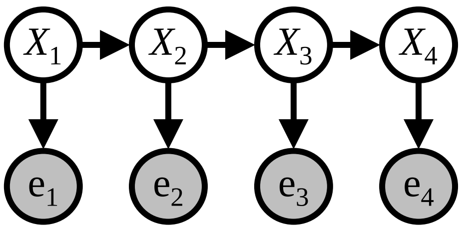
Explanation: $P(X_{1:t} | e_{1:t})$



Viterbi

Filtering for Continuous Random Variables

Filtering: $P(X_t|e_{1:t})$



Discrete

$$\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1})$$
$$= \underbrace{\alpha\, \mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1})\, \underline{\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t})}}_{\text{Update}} \quad \text{Prediction.}$$

- HMM - discrete states
  - ✓ Transition model: $P(\mathbf{X}_{t+1}|\mathbf{X}_t)$
  - ✓ Emission model: $P(\mathbf{E}_{t+1}|\mathbf{X}_{t+1})$
  - ✓ Initial state: $P(\mathbf{X}_0)$
- If the random variables are **continuous,** rather than discrete as in HMM, the number of states become **infinite**.

BN with Continuous Variables

- Continuous variables have an infinite number of possible values, so it is impossible to specify conditional probabilities explicitly for each value.
  - Discretization- dividing up the possible values into a fixed set of intervals.
  - Probability density function, e.g. Gaussian distribution

    $N(\mu, \sigma^2)(x)$

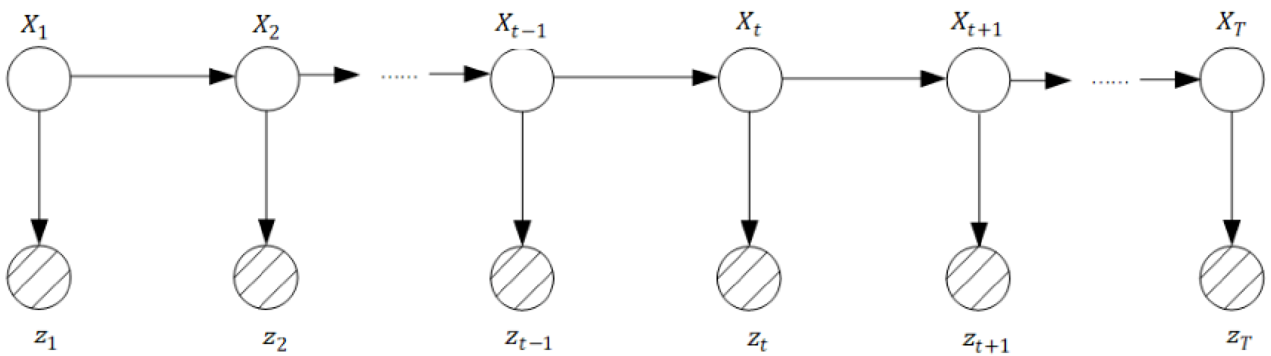
PDF: $$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

μ is mean

σ is standard deviation

σ^2 is variance

Filtering for Continuous Random Variables



- One algorithm to solve filtering problem is **Kalman Filters**.
- Applications: any system characterized by continuous state variables and noisy measurements: autonomous car location, nuclear reactors states…

Kalman Filters - Example

Z_{t+1}: GPS location

$X_t: \; P_{t,} V_t$ $\qquad$ $X_{t+1}: \; P_{t+1,} V_{t+1}$

In an ideal world, the car is running with a constant speed,

$$P_{t+1} = P_t + \Delta t \, V_t$$
$$V_{t+1} = V_t$$

$$\Rightarrow \quad X_{t+1} = AX_t$$

$$X_{t+1} = AX_t + \varepsilon \qquad \varepsilon \sim N(\,0,\, Q)$$

$$Z_{t+1} = HX_{t+1} + \delta \qquad \delta \sim N(\,0,\, R)$$

Kalman Filters - Example

Z_{t+1}: GPS location

X_t : $P_{t,}$ V_t

X_{t+1} : $P_{t+1,}$ V_{t+1}

$$X_{t+1} = AX_t + \varepsilon \qquad \varepsilon \sim N(0, Q)$$

$$Z_{t+1} = HX_{t+1} + \delta \qquad \delta \sim N(0, R)$$

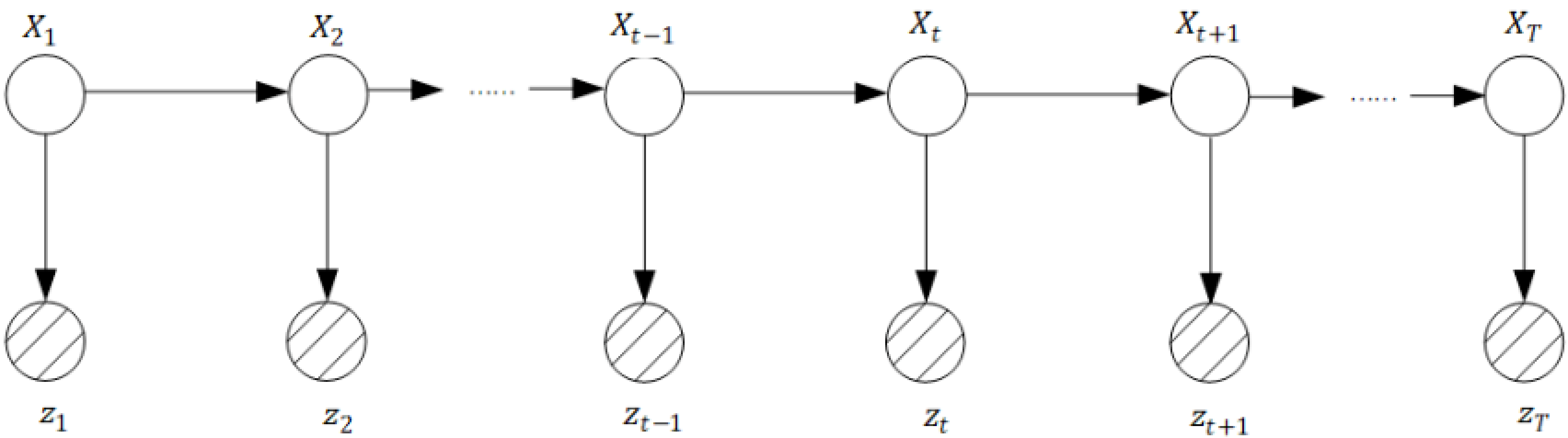
Z_t

X_t → Predict → X'_{t+1} → Update → X_{t+1}

Σ_t → Predict → Σ'_{t+1} → Update → Σ_{t+1}

$=Q + A\Sigma_t A^T$

Kalman Filters - Example



X_t: continuous states, Z_t: measurements.

- The linear assumption

$$X_{t+1} = AX_t + \varepsilon$$

$$Z_{t+1} = HX_{t+1} + \delta$$

- The Gaussian noise Assumption

$$\epsilon \sim N(0, Q) \quad , \quad \delta \sim N(0, R)$$

Kalman Filters

$$X_{t+1} = AX_t + \varepsilon$$
$$Z_{t+1} = HX_{t+1} + \delta$$

$$\epsilon \sim N(0, Q) \quad , \quad \delta \sim N(0, R)$$

- Transition model

  $P(X_{t+1} | X_t) \sim N(AX_t, Q)$

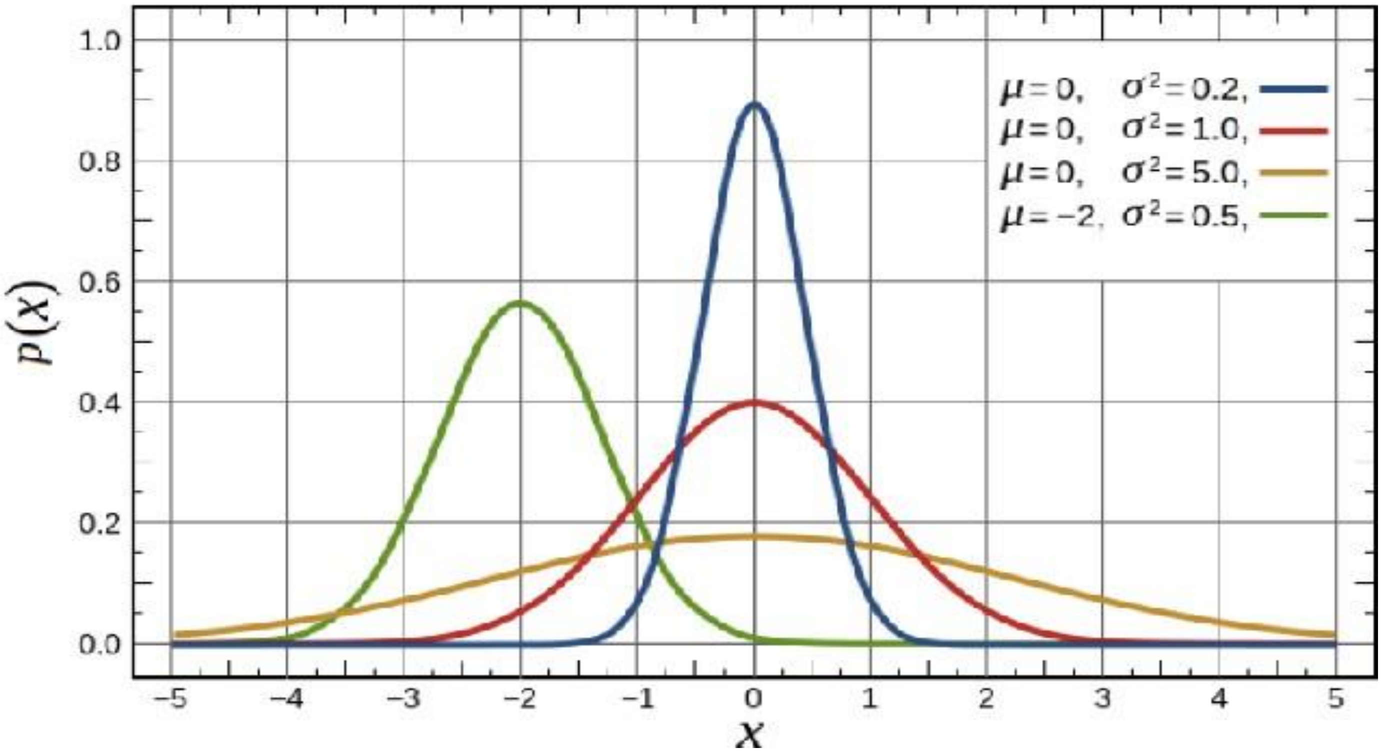
- Sensor/emission model

  $P(Z_{t+1} | X_{t+1}) \sim N(HX_{t+1}, R)$

Gaussian Distribution

The μ specifies the "location" of the Gaussian, while the σ controls the spread.

Example:



Bivariate Gaussian Distribution

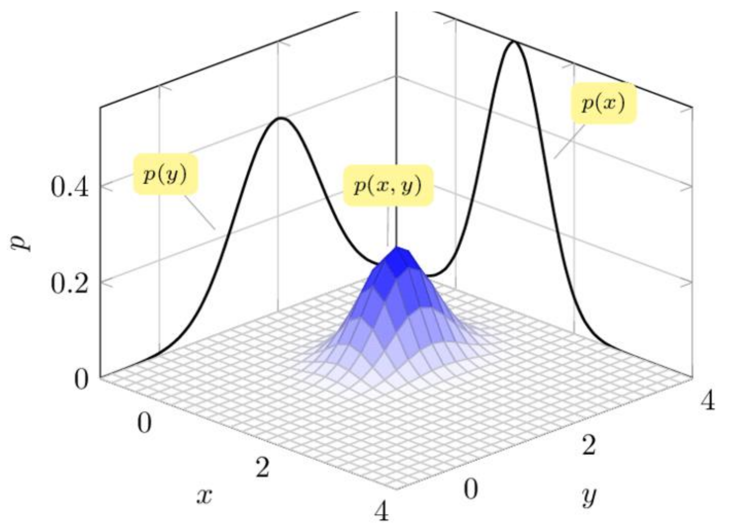
- Univariate Gaussian Distribution:
  - PDF: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
- Bivariate Gaussian Distribution:
  - PDF:

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]}$$

ρ is the correlation of X and Y.

$$\rho = \frac{cov(X,Y)}{\sigma_X\sigma_Y} \qquad -1 < \rho < 1$$

$$f(x,y) = f_x(x)f_y(y), \quad if \quad \rho = 0$$



Marginal Distribution of Bivariate Gaussian Distribution

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]} dy \\
&= \cdots \\
&= \frac{1}{\sqrt{2\pi}\sigma_X} e^{-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2} \qquad -\infty < x < \infty
\end{aligned}
$$

$$X \sim N\left(\mu_X, \sigma_X^2\right) \quad \text{and} \quad Y \sim N\left(\mu_Y, \sigma_Y^2\right)$$

Conditional Distribution of Bivariate Gaussian Distribution

$$f_{Y|X=x}(y \mid X = x) = \frac{f(x, y)}{f_X(x)} \qquad -\infty < y < \infty$$

$$= \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_Y} e^{-\frac{1}{2(1-\rho^2)\sigma_Y^2}\left(y-\mu_Y-\frac{\rho\sigma_Y(X-\mu_X)}{\sigma_X}\right)^2}$$

$$P(Y \mid X = x) \sim N\left(\mu_Y + \frac{\rho\sigma_Y(x-\mu_X)}{\sigma_X}, (1-\rho^2)\,\sigma_Y^2\right)$$

$$P(X \mid Y = y) \sim N\left(\mu_X + \frac{\rho\sigma_X(y-\mu_Y)}{\sigma_Y}, (1-\rho^2)\,\sigma_X^2\right)$$

Multivariate Gaussian Distribution

- Multivariate Gaussian Distribution:

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}\left((\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}$$

$\mathbf{x}$ is a vector, $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ is the **covariance matrix**

$$\boldsymbol{\Sigma}_{i,j} = \operatorname{cov}(X_i, X_j) = \mathrm{E}\left[(X_i - \mathrm{E}[X_i])(X_j - \mathrm{E}[X_j])^\top\right]$$

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

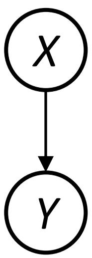
- In the bivariate case: $\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$

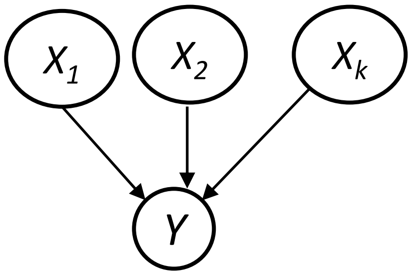
Determinant of covariance matrix: $|\Sigma| = \sigma_1^2\sigma_2^2(1-\rho^2)$

Inverse of covariance matrix: $\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)}\begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix}$

Linear Gaussian Distribution

- In BN, **Linear Gaussian distribution** is the Child random variable has a Gaussian distribution with **mean μ varies linearly** with the value of its parent, but **standard deviation σ is fixed**, i.e., Mean of Y is a linear combination of means of Gaussian parents.





$$P(Y \mid X) \sim N\left(\beta_0 + \beta X; \sigma^2\right)$$

i.e., $Y = \beta X + noise, noise \sim N\left(\beta_0; \sigma^2\right)$

$$P(Y \mid X_l, ..X_k) \sim N\left(\beta_0 + \beta_l X_l + ..\beta_k X_k; \sigma^2\right)$$

i.e., $Y = \beta_l X_l + ..\beta_k X_k + noise, \; noise \sim N\left(\beta_0; \sigma^2\right)$

- All variables are Gaussian and all conditional probability distributions are linear Gaussian.

Kalman Filters

$$X_{t+1} = AX_t + \varepsilon \qquad \epsilon \sim N(0, Q) \quad , \quad \delta \sim N(0, R)$$

$$Z_{t+1} = HX_{t+1} + \delta$$

- Transition model

  $P(X_{t+1} | X_t) \sim N(AX_t , Q)$

- Sensor/emission model

  $P(Z_{t+1} | X_{t+1}) \sim N(HX_{t+1}, R)$

Kalman Filters – Two steps

- Prediction

$$\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{z}_{1:t}) = \int_{\mathbf{x}_t} \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t) P(\mathbf{x}_t \mid \mathbf{z}_{1:t}) \, d\mathbf{x}_t$$

Recall in LE16 $\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t}) = \sum_{\mathbf{x}_t} P(\mathbf{X}_{t+1} | \mathbf{x}_t) P(\mathbf{X}_t | \mathbf{e}_{1:t})$

- If the current distribution $P(\mathbf{X}_t \mid \mathbf{z}_{1:t})$ is Gaussian and the transition model $\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{x}_t)$ is linear–Gaussian, then **prediction** is Gaussian distribution.

Kalman Filters – Two steps

- Update

$$\mathbf{P}\left(\mathbf{X}_{t+1} \mid \mathbf{z}_{1:t+1}\right) = \alpha \mathbf{P}\left(\mathbf{z}_{t+1} \mid \mathbf{X}_{t+1}\right) \mathbf{P}\left(\mathbf{X}_{t+1} \mid \mathbf{z}_{1:t}\right)$$

Recall in LE16 $\mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t+1}) = \alpha \underbrace{\mathbf{P}(\mathbf{e}_{t+1} \mid \mathbf{X}_{t+1}) \mathbf{P}(\mathbf{X}_{t+1} \mid \mathbf{e}_{1:t})}_{\text{Update}}$

  - If the prediction is Gaussian and the sensor model is linear–Gaussian, then, after conditioning on the new evidence, the **updated distribution** is also Gaussian.

Kalman Filters

- Given these properties,
  - If we start with a **Gaussian** prior, filtering with a **linear–Gaussian** model produces a **Gaussian** state distribution for all time.
  - The mapping from one Gaussian to another is computing a new mean and covariance matrix from previous mean and covariance matrix.

A One-dimensional Example

- Given continuous state variable X_t, and a noisy observation variable Z_t

Transition model:

$$P(X_{t+1} | X_t) \sim N(AX_t, Q)$$

$$P(X_{t+1}|X_t) \sim N\left(X_t, Q\right)$$

When A = I, Q reduce to X's variance.

$$P(x_{t+1}|x_t) = \alpha e^{-\frac{1}{2}\left(\frac{(x_{t+1}-x_t)^2}{\sigma_x^2}\right)}$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

A One-dimensional Example

- Given continuous state variable $X_{t,}$ and a noisy observation variable Z_t

Sensor model:

$$P(Z_{t+1} | X_{t+1}) \sim N(HX_{t+1}, R)$$

$$P(Z_{t+1}|X_{t+1}) \sim N\left(X_{t+1}, R\right)$$

When H= I, R reduce to Z's variance.

$$P(z_t|x_t) = \alpha e^{-\frac{1}{2}\left(\frac{(z_t-x_t)^2}{\sigma_z^2}\right)}$$

A One-dimensional Example

Transition model: $P(x_{t+1}|x_t) = \alpha e^{-\frac{1}{2}\left(\frac{(x_{t+1}-x_t)^2}{\sigma_x^2}\right)}$

Sensor model: $P(z_t|x_t) = \alpha e^{-\frac{1}{2}\left(\frac{(z_t-x_t)^2}{\sigma_z^2}\right)}$

Prior: $\mathbf{X}_0 \sim N(\mu_0, \sigma_0)$ $P(x_0) = \alpha e^{-\frac{1}{2}\left(\frac{(x_0-\mu_0)^2}{\sigma_0^2}\right)}$

$$P(x_1) = \int_{-\infty}^{\infty} P(x_1 \mid x_0) P(x_0)\, dx_0 = \alpha \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{(x_1-x_0)^2}{\sigma_x^2}\right)} e^{-\frac{1}{2}\left(\frac{(x_0-\mu_0)^2}{\sigma_0^2}\right)} dx_0$$

$$= \alpha\, e^{-\frac{1}{2}\left(c-\frac{b^2}{4a}\right)} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(a(x_0-\frac{-b}{2a})^2\right)} dx_0$$

$$= \alpha\, e^{-\frac{1}{2}\left(c-\frac{b^2}{4a}\right)} \qquad = \alpha\, e^{-\frac{1}{2}\left(\frac{(x_1-\mu_0)^2}{\sigma_0^2+\sigma_x^2}\right)}$$

$a = (\sigma_0^2 + \sigma_x^2)/(\sigma_0^2\sigma_x^2)$, $b = -2(\sigma_0^2 x_1 + \sigma_x^2\mu_0)/(\sigma_0^2\sigma_x^2)$ $c = (\sigma_0^2 x_1^2 + \sigma_x^2\mu_0^2)/(\sigma_0^2\sigma_x^2)$

A One-dimensional Example

$$P(x_1|z_1) = \alpha P(z_1|x_1)P(x_1)$$

$$= \alpha e^{-\frac{1}{2}\left(\frac{(z_1-x_1)^2}{\sigma_z^2}\right)} e^{-\frac{1}{2}\left(\frac{(x_1-\mu_0)^2}{\sigma_0^2+\sigma_x^2}\right)}$$

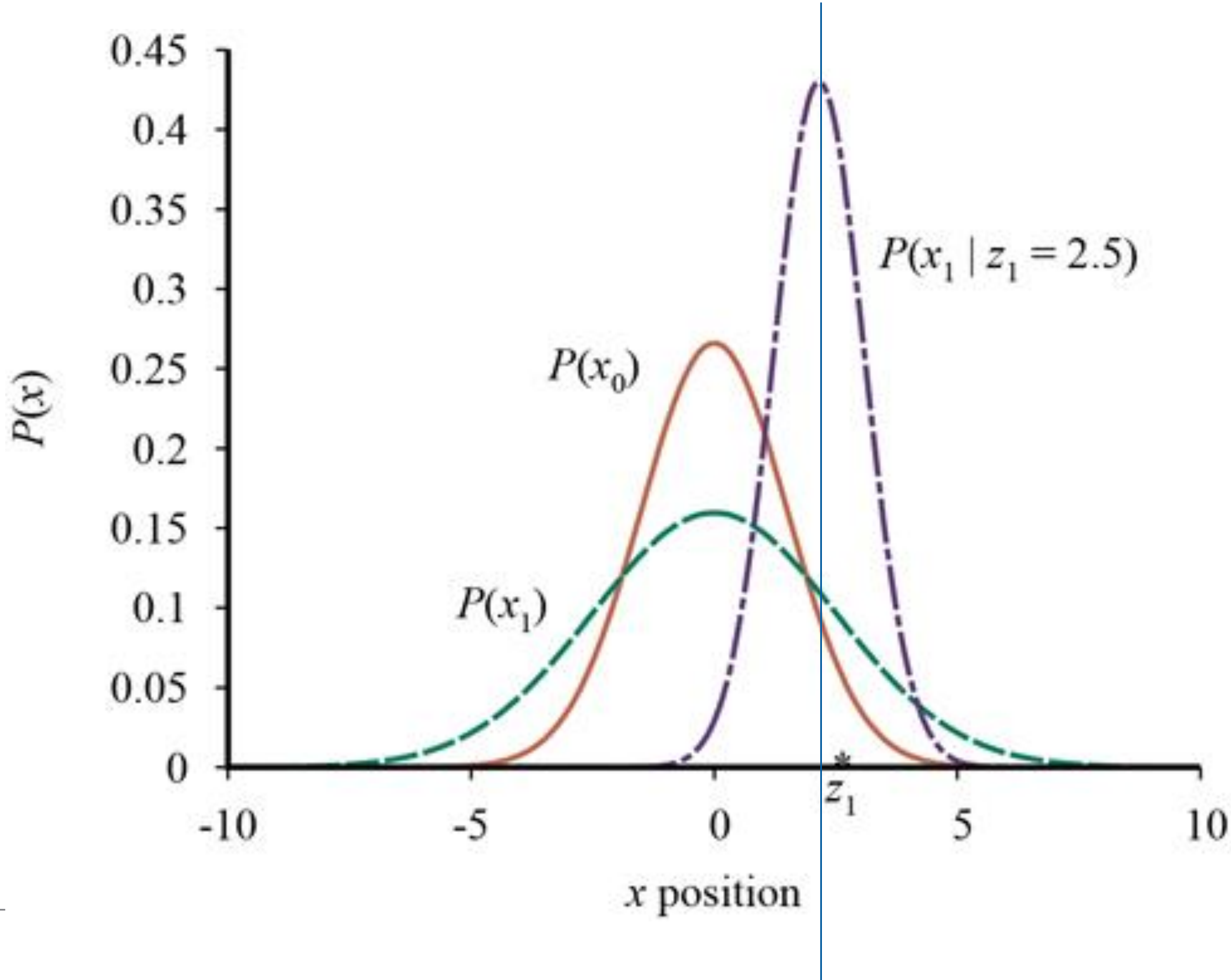
$$= \alpha e^{-\frac{1}{2}\frac{\left(x_1-\frac{(\sigma_0^2+\sigma_x^2)z_1+\sigma_z^2\mu_0}{\sigma_0^2+\sigma_x^2+\sigma_z^2}\right)^2}{(\sigma_0^2+\sigma_x^2)\sigma_z^2/(\sigma_0^2+\sigma_x^2+\sigma_z^2)}}$$

We see that the new mean and standard deviation can be calculated from the old mean and standard deviation:

$$\mu_{t+1} = \frac{(\sigma_t^2+\sigma_x^2)z_{t+1}+\sigma_z^2\mu_t}{\sigma_t^2+\sigma_x^2+\sigma_z^2} \qquad \text{and} \qquad \sigma_{t+1}^2 = \frac{(\sigma_t^2+\sigma_x^2)\sigma_z^2}{\sigma_t^2+\sigma_x^2+\sigma_z^2}$$

The exponent is a **quadratic** form which is the key property to help filtering preserves the Gaussian nature of the state distribution.

A One-dimensional Example



General case

$$
\begin{aligned}
P(\mathbf{x}_{t+1}|\mathbf{x}_t) &= N(\mathbf{x}_{t+1}; \mathbf{F}\mathbf{x}_t, \Sigma_x) \\
P(\mathbf{z}_t|\mathbf{x}_t) &= N(\mathbf{z}_t; \mathbf{H}\mathbf{x}_t, \Sigma_z),
\end{aligned}
$$

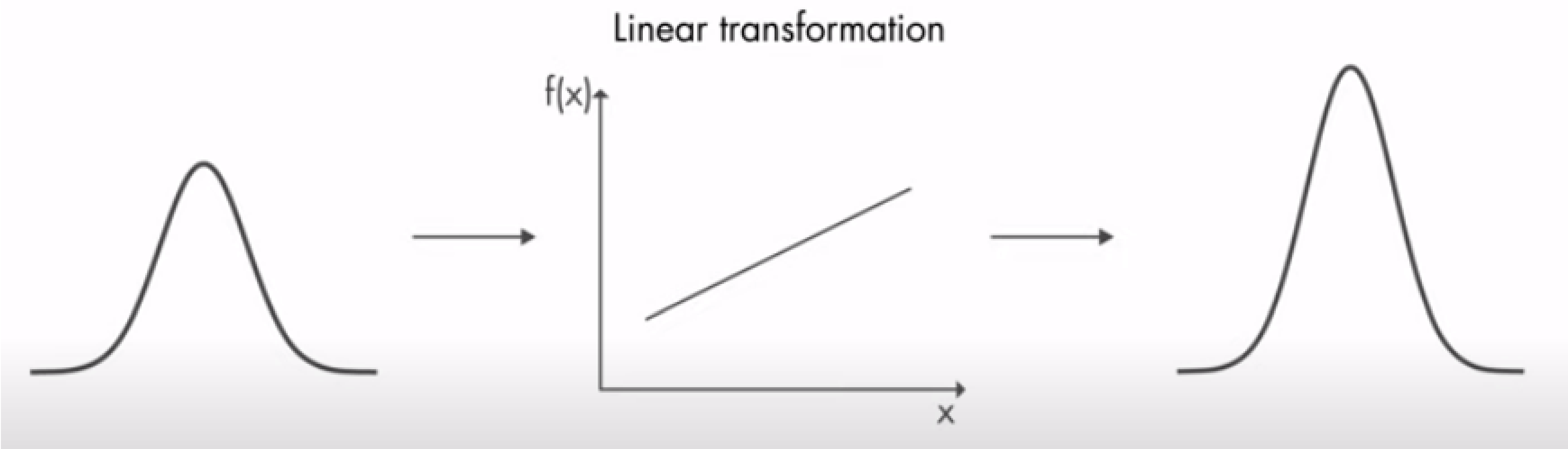
$\mathbf{F}$ and Σ_x are matrices describing the linear transition model and transition noise covariance, and $\mathbf{H}$ and Σ_z are the corresponding matrices for the sensor model.

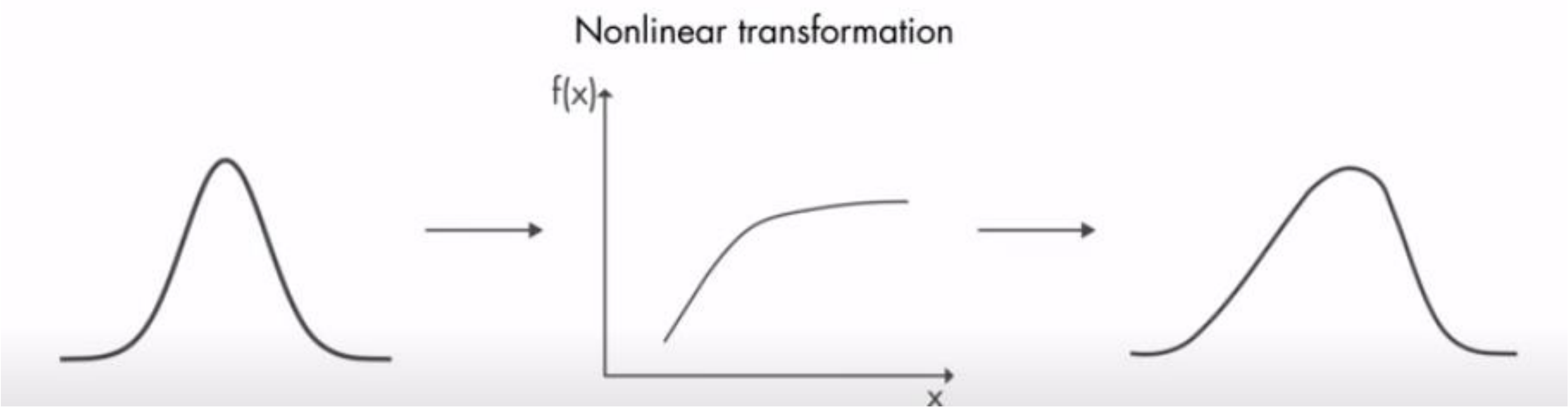
$$
\begin{aligned}
\mu_{t+1} &= \mathbf{F}\mu_t + \mathbf{K}_{t+1}(\mathbf{z}_{t+1} - \mathbf{H}\mathbf{F}\mu_t) \\
\Sigma_{t+1} &= (\mathbf{I} - \mathbf{K}_{t+1}\mathbf{H})(\mathbf{F}\Sigma_t\mathbf{F}^\top + \Sigma_x)
\end{aligned}
$$

Kalman gain matrix

$$
\mathbf{K}_{t+1} = (\mathbf{F}\Sigma_t\mathbf{F}^\top + \Sigma_x)\mathbf{H}^\top(\mathbf{H}(\mathbf{F}\Sigma_t\mathbf{F}^\top + \Sigma_x)\mathbf{H}^\top + \Sigma_z)^{-1}
$$

Kalman Filters





Kalman Filters

- Problem: The assumptions made—a linear Gaussian transition and sensor models—are very strong.
- Extended Kalman filter (EKF):

  -- modelling the system as locally linear in X_t in a region of $X_t = \mu_t$
- Switching Kalman filter:

  -- multiple Kalman filters run in parallel, each using a different model of the system