Probabilistic Reasoning Over Time 3: Kalman Filters

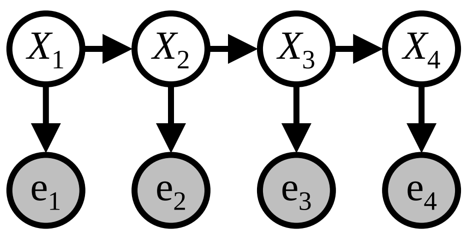
3007/7059 Artificial Intelligence

Wei Zhang

School of Computer Science
University of Adelaide

Filtering for Continuous Random Variables

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



Discrete

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

- If the random variables are **continuous,** rather than discrete as in HMM, the number of states become **infinite**.
- 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…

Bayesian Network 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

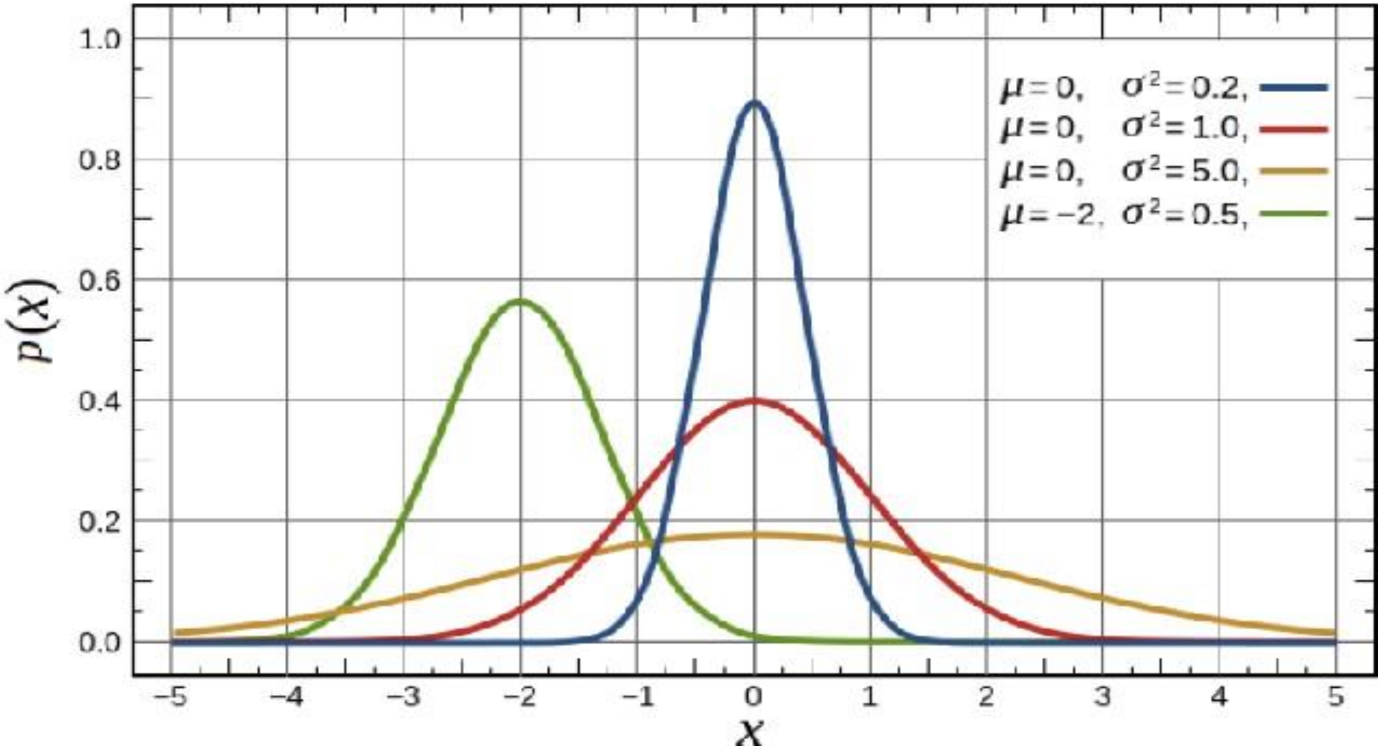
σ is standard deviation

σ^2 is variance

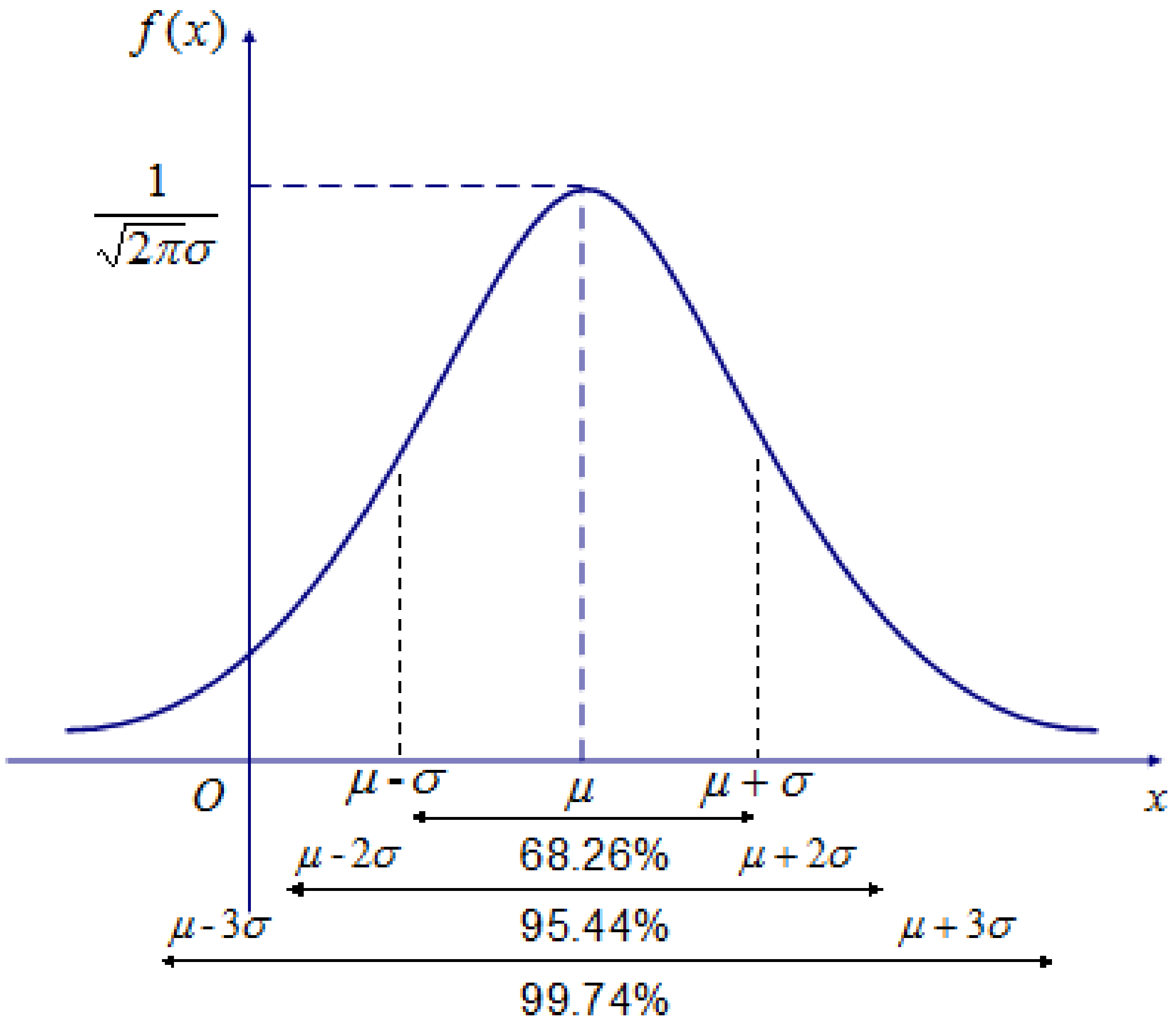
Gaussian Distribution

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

Example:



Gaussian Distribution



Bivariate Gaussian Distribution

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

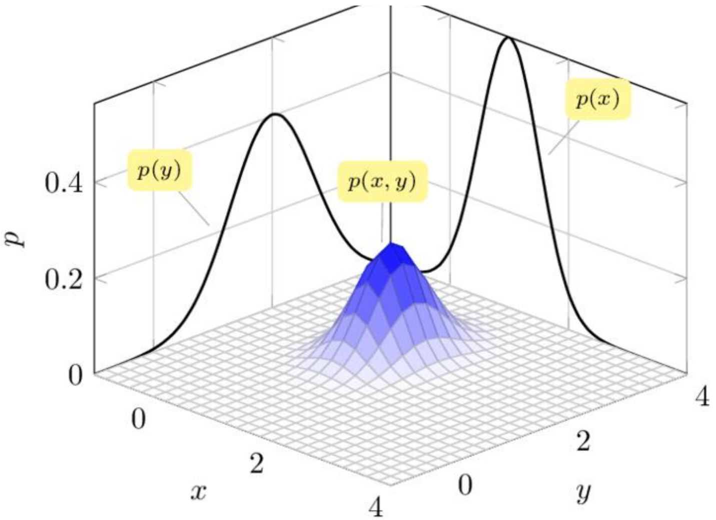
- Bivariate Gaussian Distribution:
  - PDF:

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

ρ is the correlation of X and Y.

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

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



Conditional Distribution of Bivariate Gaussian Distribution

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

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

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

Marginal Distribution of Bivariate Gaussian Distribution

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

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

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**

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

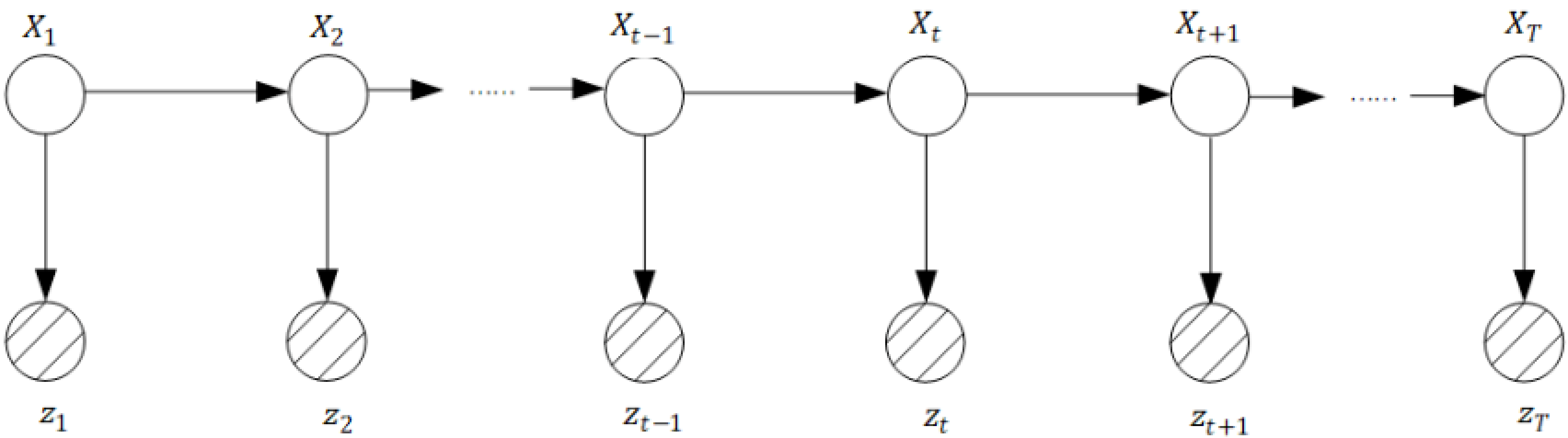
$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

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

Kalman Filters



X_t: continuous states, Z_t: observations.

- The linear assumption

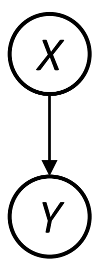
$$X_{t+1} = AX_t + B + \epsilon$$
$$Z_{t+1} = HX_{t+1} + C + \delta$$

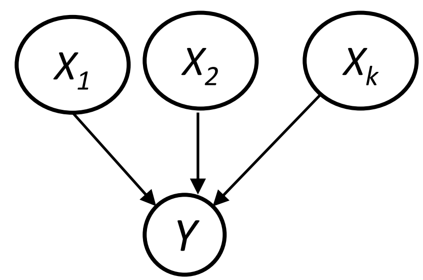
- The Gaussian noise Assumption

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

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.





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

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

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

Kalman Filters

$$X_{t+1} = AX_t + B + \epsilon$$
$$Z_{t+1} = HX_{t+1} + C + \delta$$

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

- Transition model

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

Sensor/emission model

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

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

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

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

  - 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 the 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 1-dimensional Example

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

Transition model:

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

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

When A = I and B = 0, Q reduce to X's variance.

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

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

A 1-dimensional Example

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

Sensor model:

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

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

When H= I and C= 0, R reduce to Z's variance.

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

A 1-dimensional Example

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

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

Prior: $\mathbf{P}(\mathbf{X}_0) \sim N(\mu_0, \Sigma_0)$. $f(x_0) = \alpha\, e^{-\frac{1}{2}\left(\frac{(x_0-\mu_0)^2}{\sigma_0^2}\right)}$

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

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

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

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

A 1-dimensional Example

$$f\,(x_1|z_1) = \alpha\, f(z_1|x_1)\, f\,(x_1)$$

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

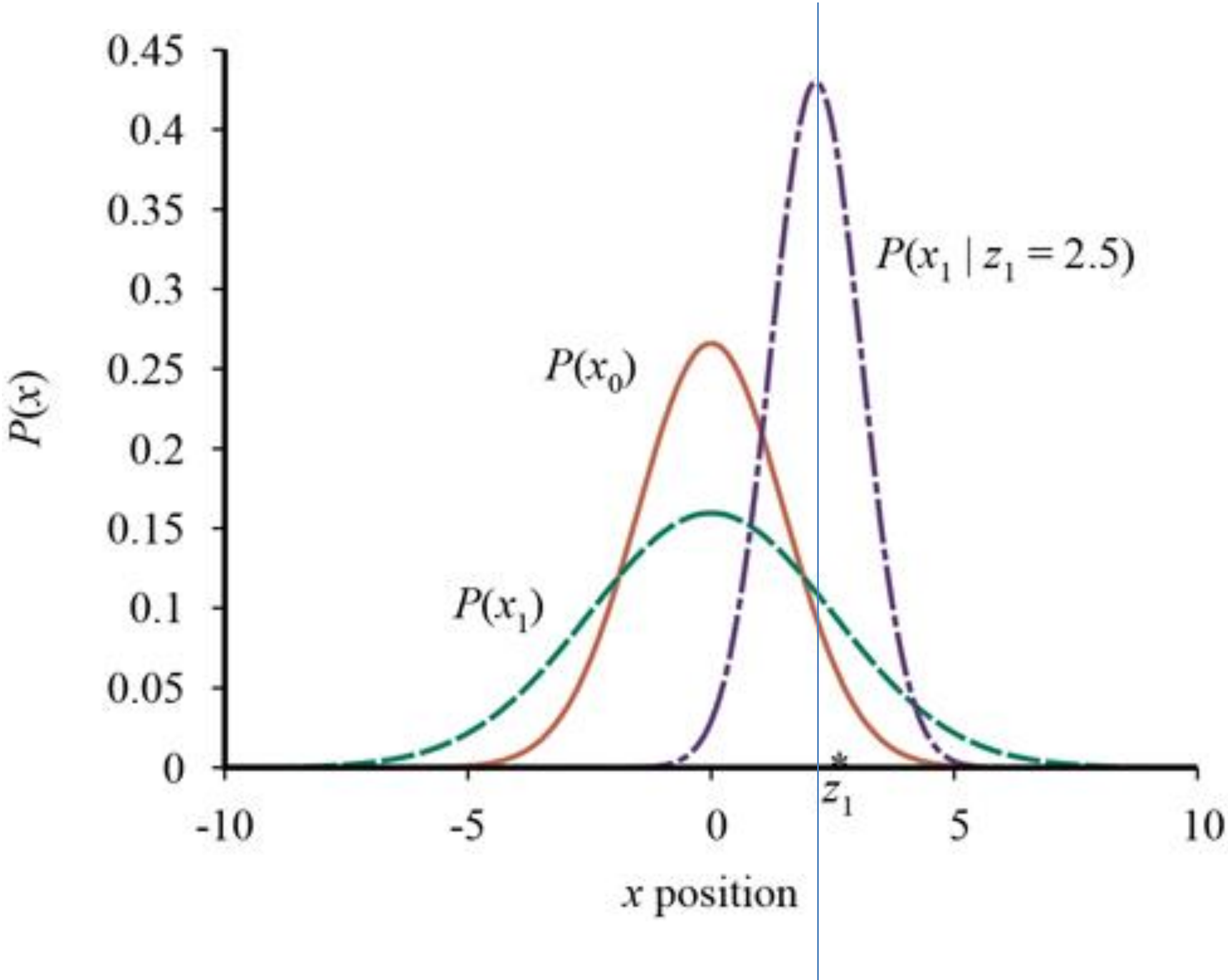
$$= \alpha e^{-\frac{1}{2}\frac{\left(x_1-\frac{(\sigma_0^2+\sigma_x^2)z_1+\sigma_z^2\mu_0}{\sigma_0^2+\sigma_x^2+\sigma_z^2}\right)^2}{(\sigma_0^2+\sigma_x^2)\sigma_z^2/(\sigma_0^2+\sigma_x^2+\sigma_z^2)}}$$

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 1-dimensional Example



General case

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

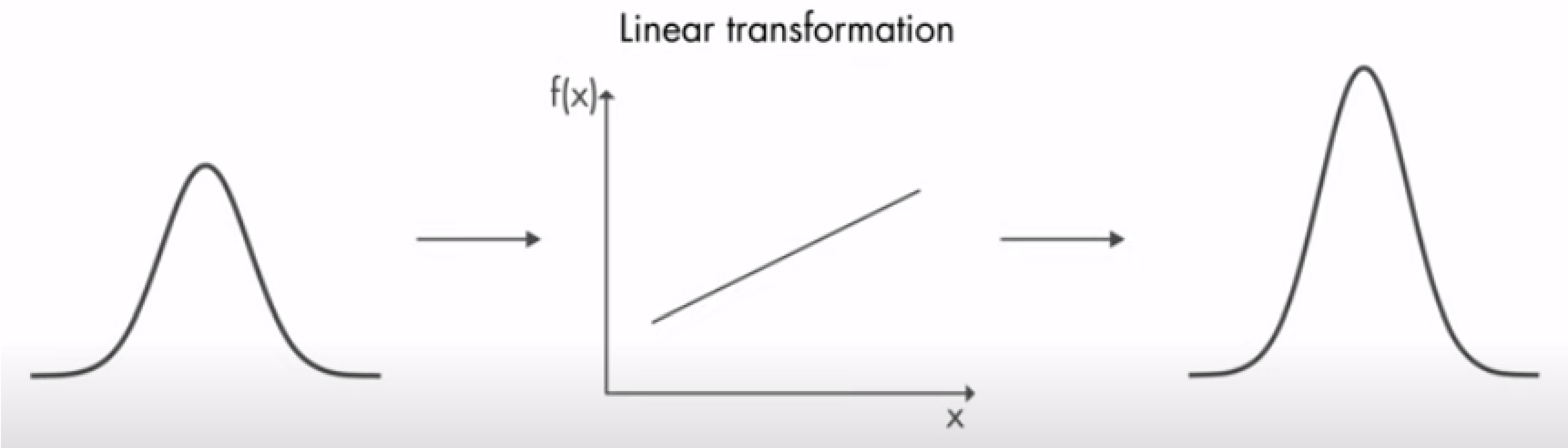
$\mathbf{F}$ and Σ_x are matrices describing the linear transition model and transition noise covariance, and $\mathbf{H}$ and Σ_z are the corresponding matrices for the sensor model.

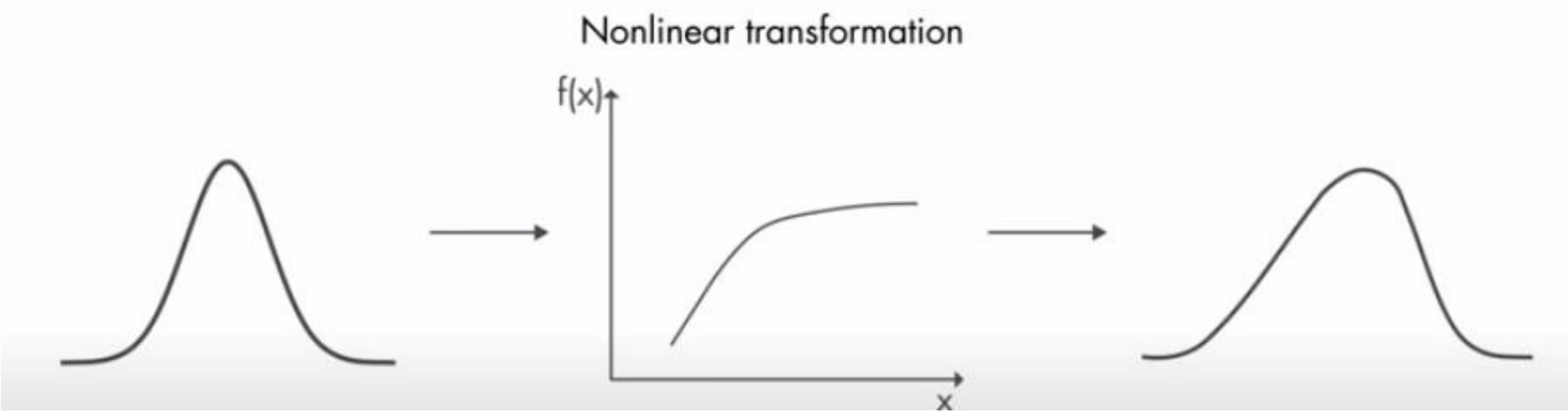
$$
\begin{aligned}
\mu_{t+1} &= \mathbf{F}\mu_t + \mathbf{K}_{t+1}(\mathbf{z}_{t+1} - \mathbf{H}\mathbf{F}\mu_t) \\
\Sigma_{t+1} &= (\mathbf{I} - \mathbf{K}_{t+1}\mathbf{H})(\mathbf{F}\Sigma_t\mathbf{F}^\top + \Sigma_x)
\end{aligned}
$$

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