

Statistik

Wie “erweitert” man sein Wissen?

Bayes und Kalman



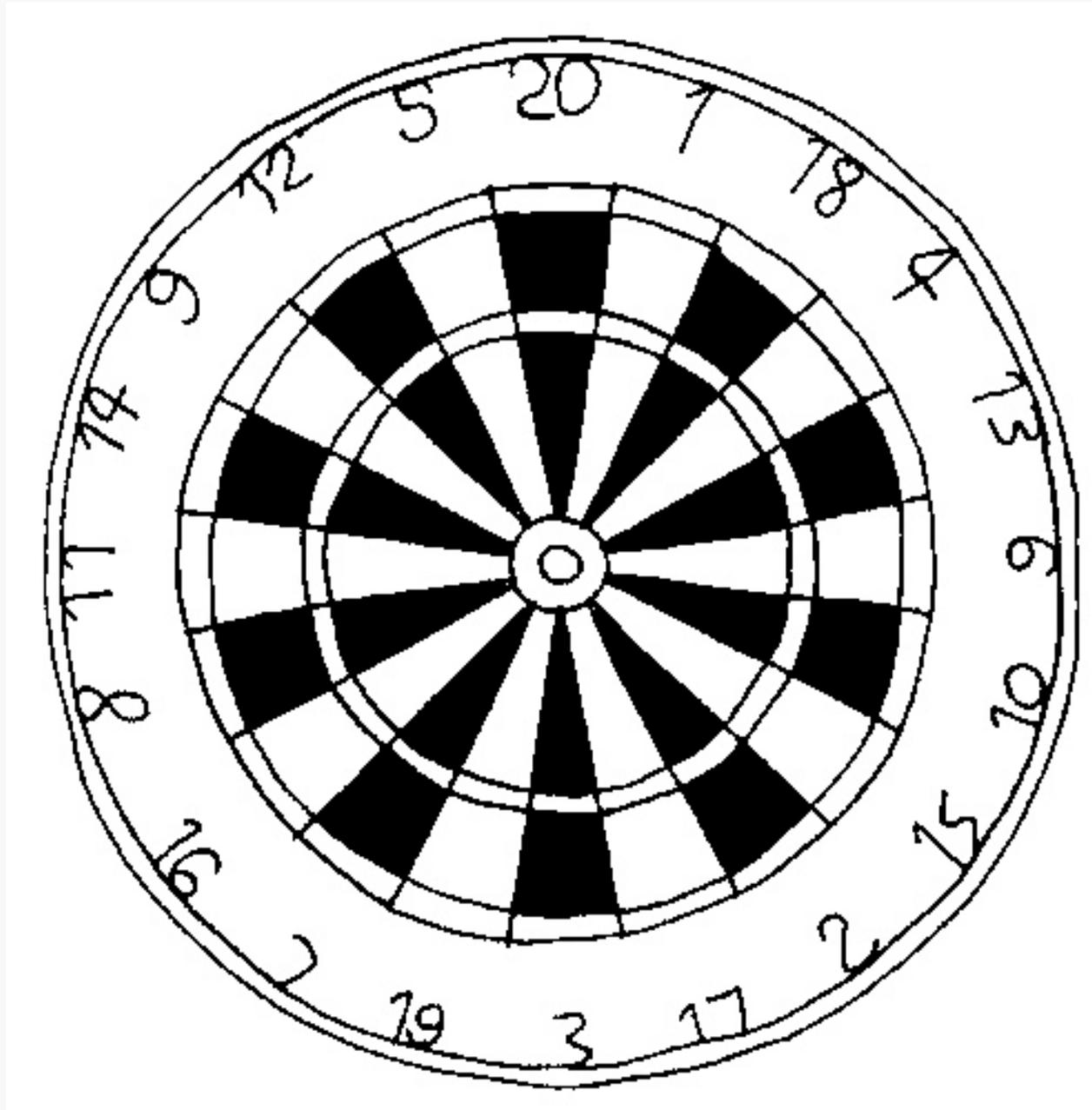
REV. T. BAYES

1701-1761



Rudolf Emil Kálmán
Geb. 1930

Bayes als Dartspiel



Genauer gesagt als *Blind Darts*

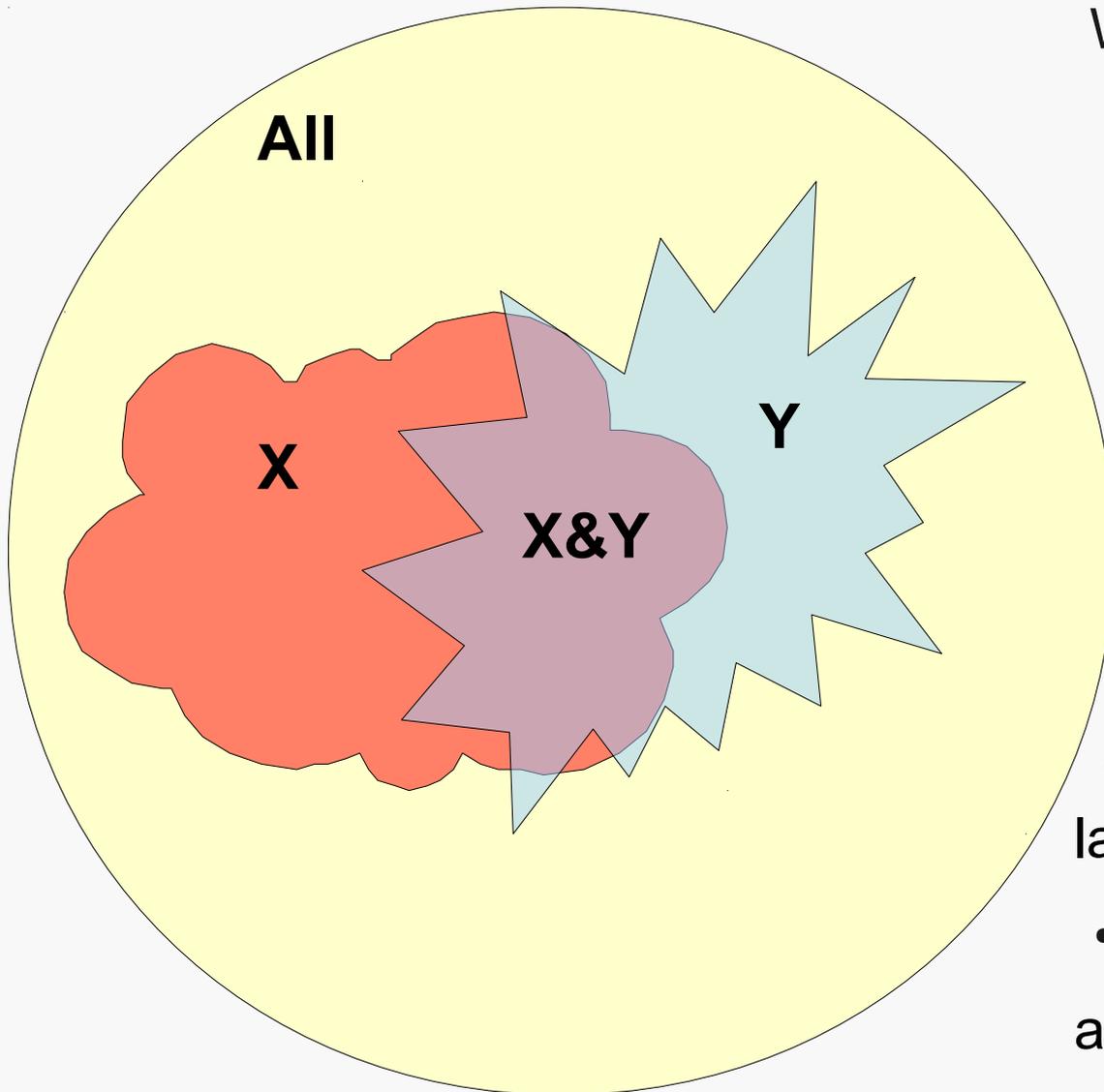
© Original Artist

Reproduction rights obtainable from
www.CartoonStock.com



- Wir werfen blind auf die Dartscheibe und fragen nach der Wahrscheinlichkeit einen Bereich X zu treffen

Bayes als Dartspiel



Wahrscheinlichkeit **X** zu treffen

- $p(x) = A(\mathbf{X}) / \text{All}$

Wahrscheinlichkeit **Y** zu treffen

- $p(y) = A(\mathbf{Y}) / \text{All}$

Wahrscheinlichkeit **X** und **Y**
zu treffen

- $p(x,y) = A(\mathbf{X\&Y}) / \text{All}$

Wahrscheinlichkeit **X** zu treffen,
wenn der Pfeil innerhalb von **Y**
landet

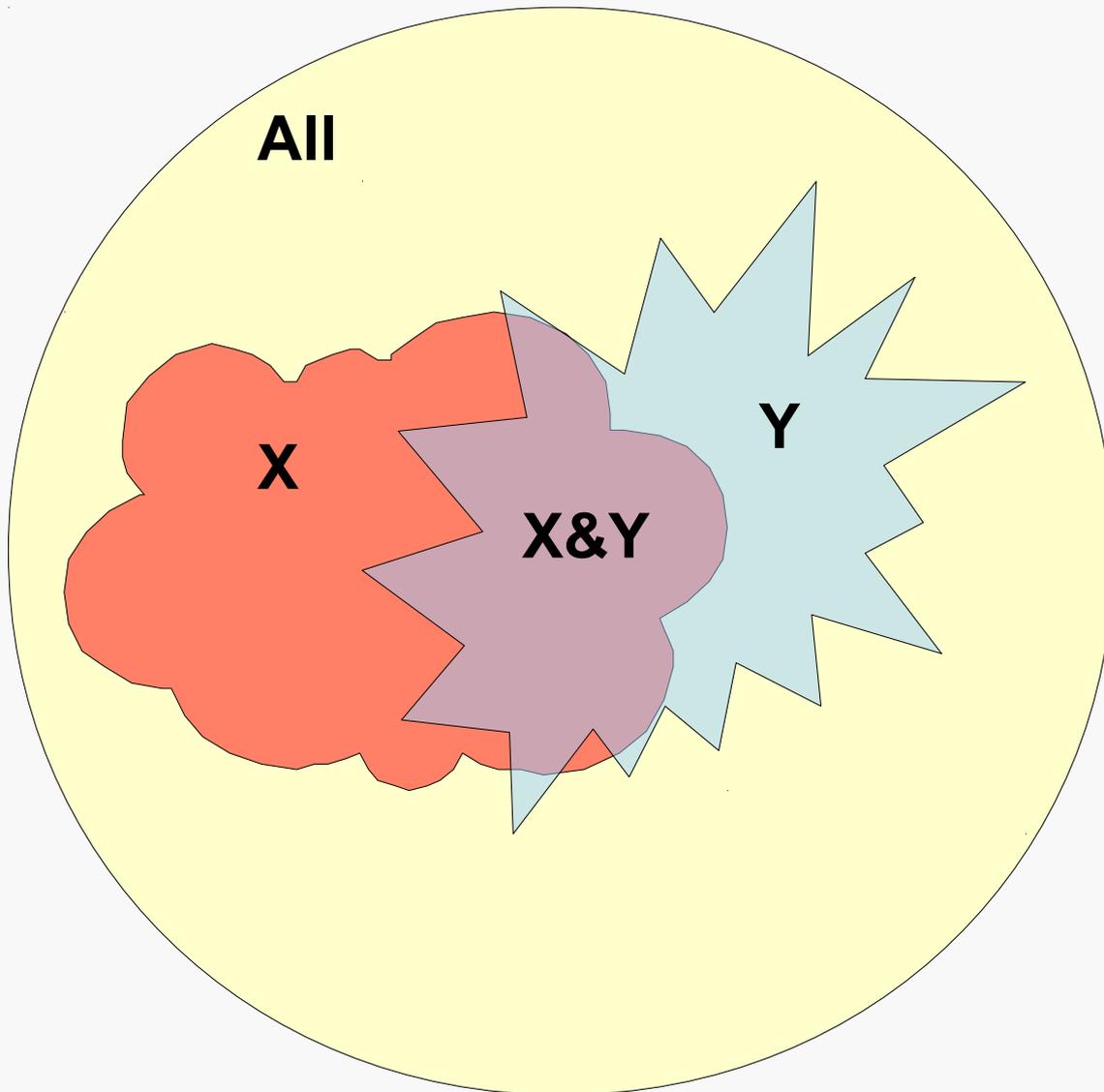
- $p(x|y) = A(\mathbf{X\&Y}) / A(\mathbf{Y})$

analog für **X**

- $p(y|x) = A(\mathbf{X\&Y}) / A(\mathbf{X})$

Fläche von **X** = $A(x)$, etc.

Bayes als Dartspiel



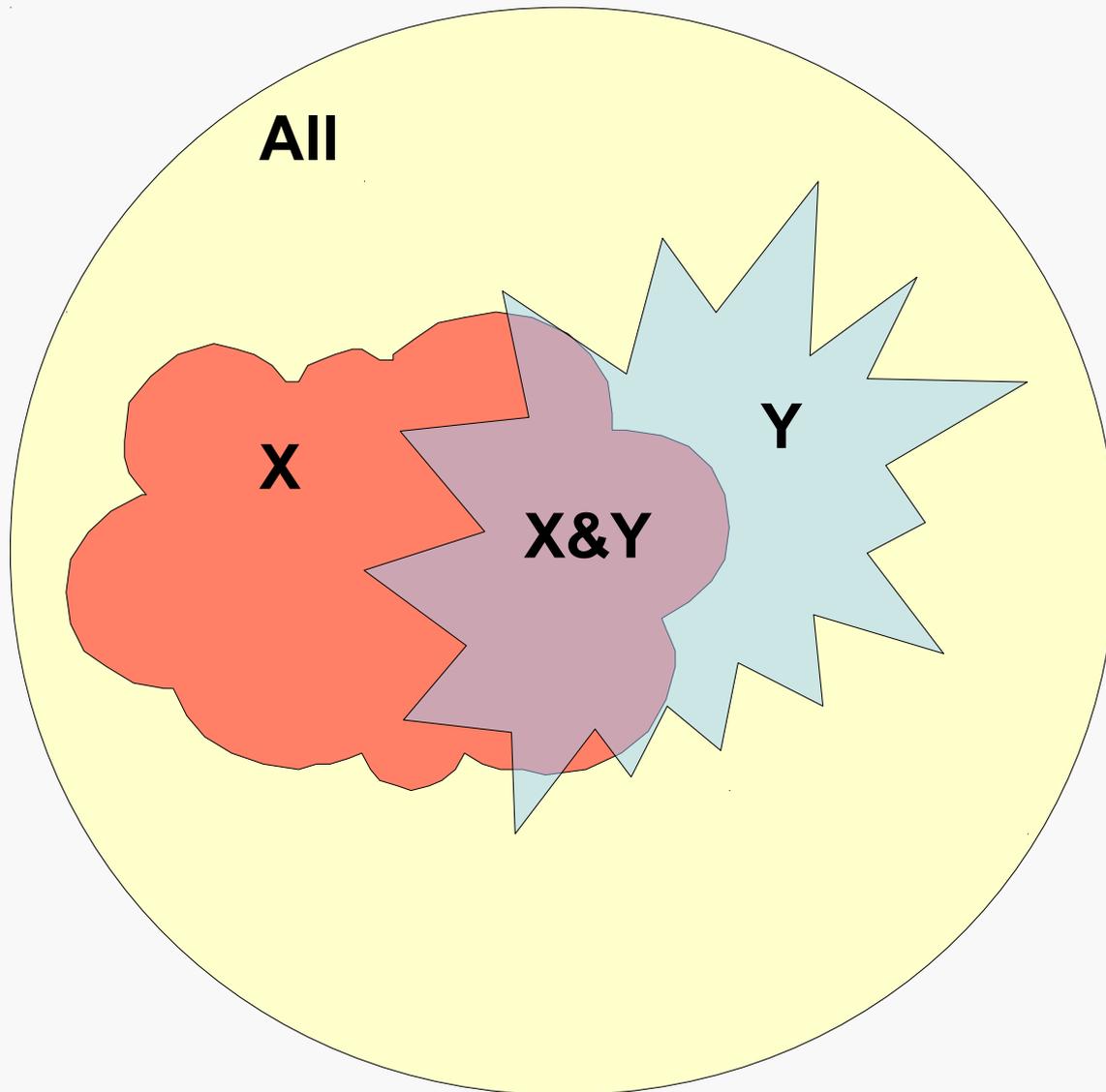
- $p(x) = A(X) / All$
 - $p(y) = A(Y) / All$
 - $p(x,y) = A(X\&Y) / All$
 $= p(y,x)$
 - $p(x|y) = A(X\&Y) / A(Y)$
 $= p(x,y) * All / p(y) * All$
 $= p(x,y) / p(y)$
- $p(x,y) = p(x|y) * p(y)$

- $p(y|x) = A(X\&Y) / A(X)$
 $= p(x,y) / p(x)$

→ $p(y,x) = p(y|x) * p(x)$

Fläche von X = A(x), etc.

Bayes als Dartspiel



Produktregel

- $p(x,y) = p(x|y) * p(y)$
- $p(x,y) = p(y|x) * p(x)$

Bayes Theorem

- $p(x|y) = p(y|x) * p(x) / p(y)$

Das ist natürlich keine Herleitung,
sondern eine Visualisierung.

Namensgebung

Posterior

Likelihood

Prior

$$p(x|y) = \frac{p(y|x) p(x)}{p(y)}$$

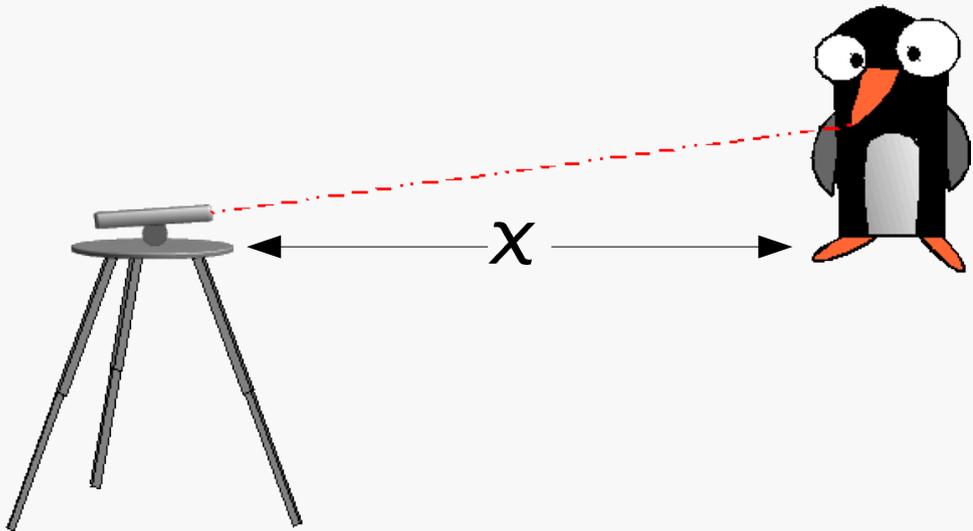
Prior: $p(\text{Hypothese})$
Likelihood: $p(\text{Daten} | \text{Hypothese})$
Posterior: $p(\text{Hypothese} | \text{Daten})$

$$\int p(y|x) p(x) dx$$

Marginalisierung

$$p(x) = \int p(x, y) dy$$

Data Assimilation



Messung
1. $m_1 \pm e_1$
2. $m_2 \pm e_2$
etc.

- Gesucht wird eine Größe x . Wir haben bereits eine Messung m_1 mit Gaußschem Fehler e_1 .
- Nun kommt eine neue Messung $m_2 \pm e_2$ hinzu.

Wie faßt man die alte und die neue Messung optimal zusammen?

Data Assimilation

Wir starten mit der Hypothese = 1. Messung

$$p(x) = g(x|\mu = m_1, \sigma = e_1) = \frac{1}{\sqrt{2\pi} e_1} e^{-\frac{1}{2} \frac{(x-m_1)^2}{e_1^2}}$$

Für ein gegebenes x ist die Wahrscheinlichkeit (*likelihood*) der 2. Messung

$$p(m_2|x) = g(m_2|\mu = x, \sigma = e_2) = \frac{1}{\sqrt{2\pi} e_2} e^{-\frac{1}{2} \frac{(m_2-x)^2}{e_2^2}}$$

Damit ist die Wahrscheinlichkeit für einen Wert x bei Berücksichtigung der 2. Messung (Bayes)

$$p(x|m_2) = p(m_2|x)p(x) / p(m_2)$$

Data Assimilation

Wir wählen als optimalen Wert für x den Wert \hat{x} ,
der die Wahrscheinlichkeit maximiert

$$\partial_x p(x|m_2) |_{x=\hat{x}} = 0$$

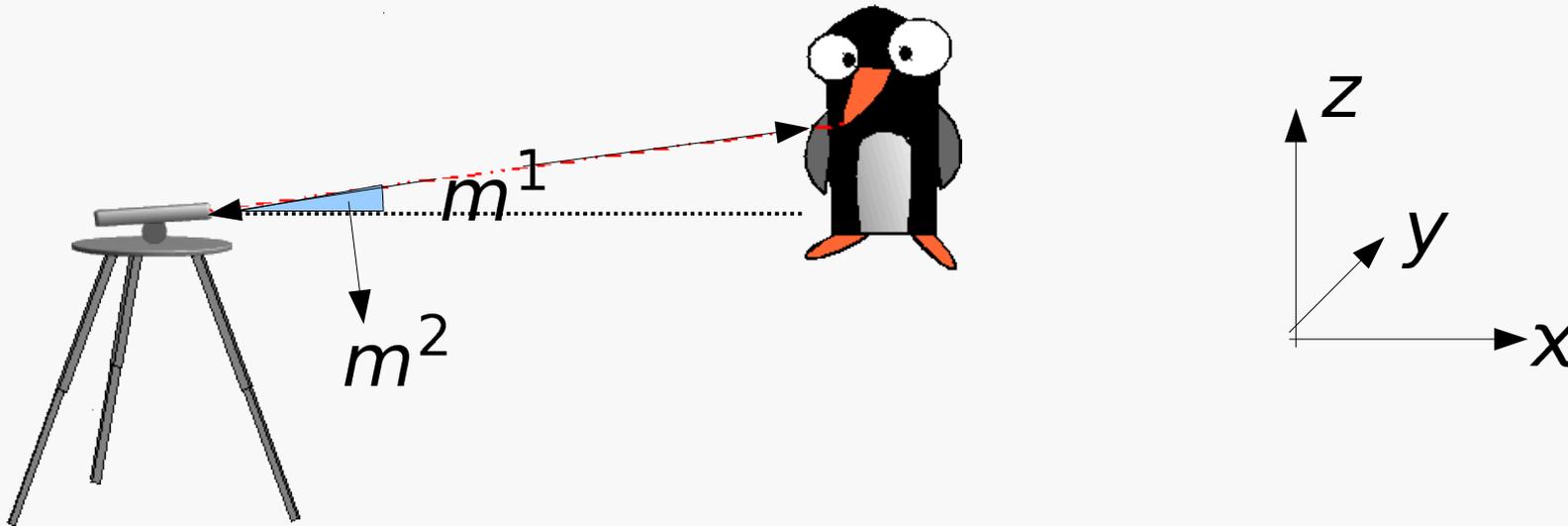
$$\sim \partial_x (p(m_2|x) p(x))$$

$$= p(x) \partial_x p(m_2|x) p(x) + p(m_2|x) \partial_x p(x)$$

$$= p(m_2|x) p(x) \left(\frac{m_2 - x}{e_2^2} - \frac{x - m_1}{e_1^2} \right)$$

$$\Rightarrow \hat{x} = \frac{\frac{m_1}{e_1^2} + \frac{m_2}{e_2^2}}{\frac{1}{e_1^2} + \frac{1}{e_2^2}} \quad \text{gewichteter Mittelwert}$$

Kalman Filter



Das einfache Beispiel wird jetzt verallgemeinert

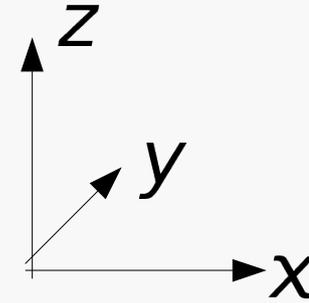
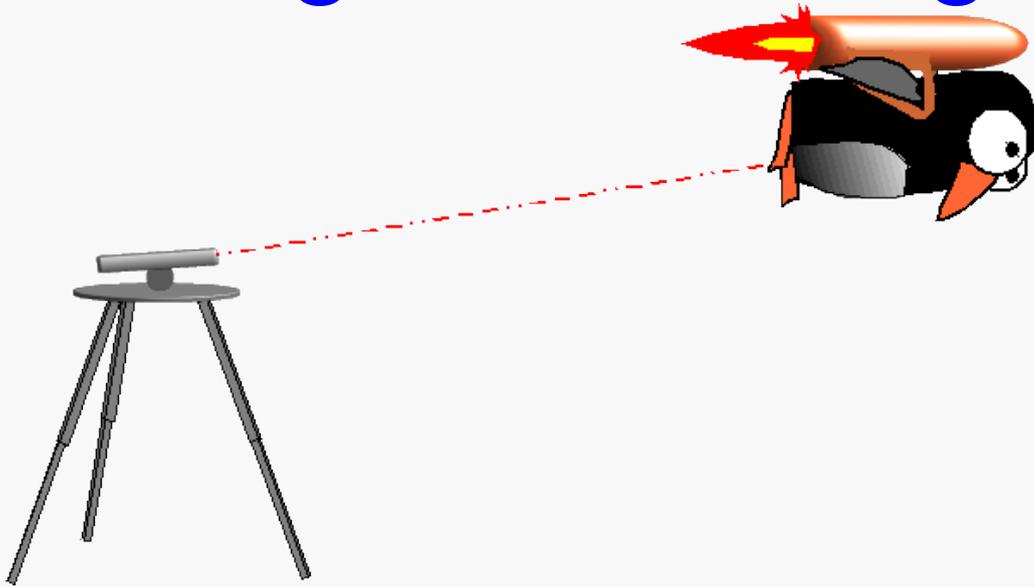
x wird n -dim \vec{x} z.B. $\vec{x} = (x, y, z, p_x, p_y, p_z)$

die Messung m -dim \vec{m} z.B. $\vec{m} = (m^1, m^2)$

$$m \leq n$$

Darüber hinaus ...

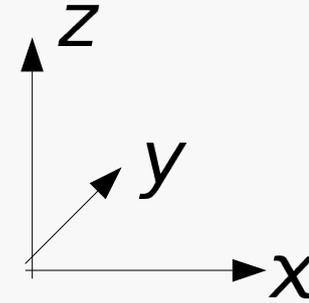
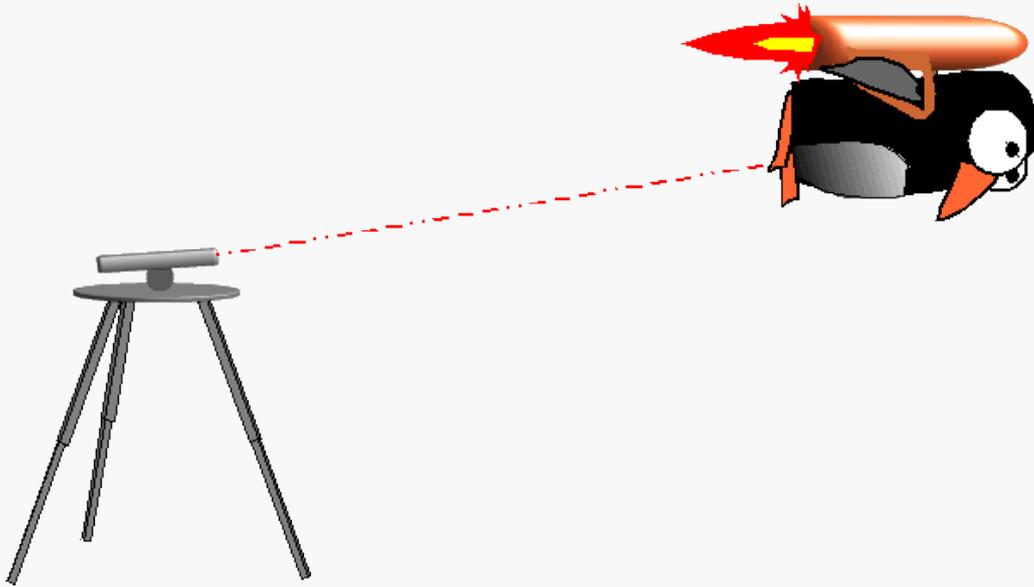
Penguin Tracking = Kalman Filter



... hält unser Messobjekt nicht still!

Dynamisches System

Kalman Filter



Messung

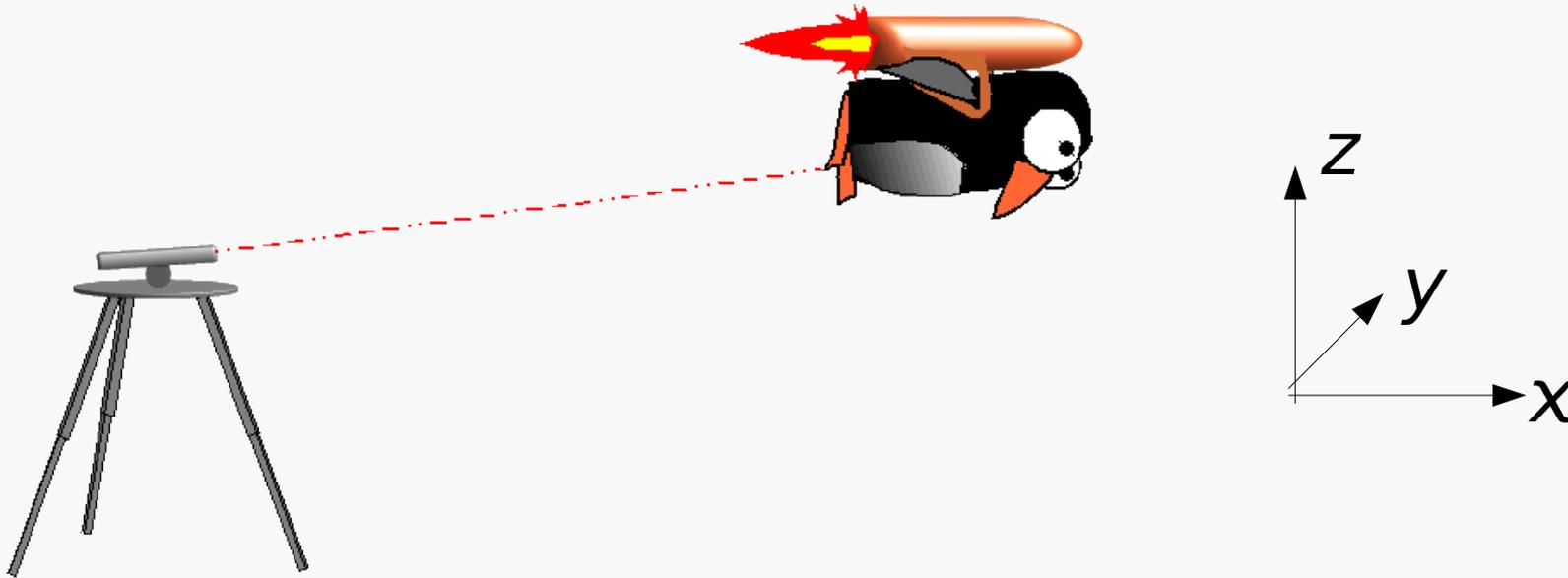
m_1

Zustand (state)

x_1



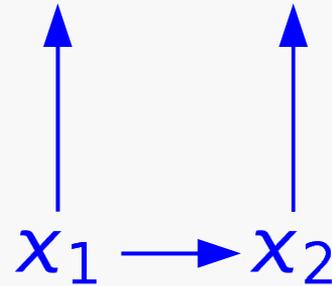
Kalman Filter



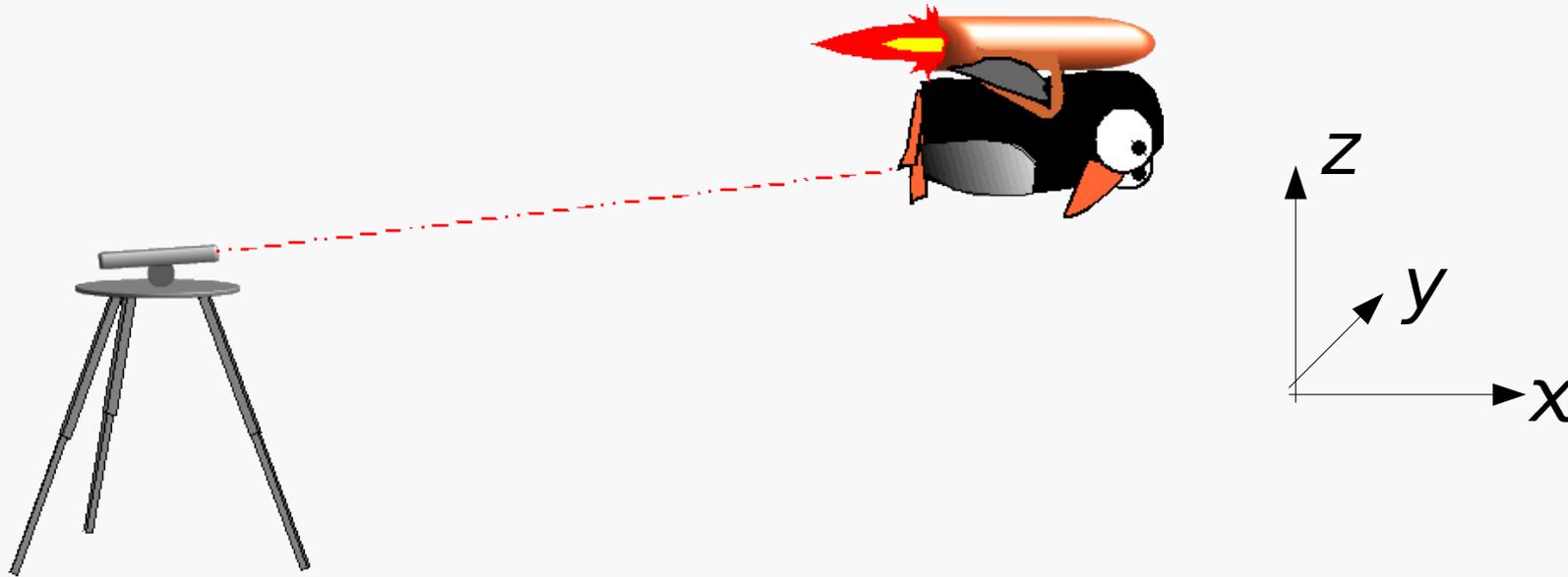
Messung

m_1 m_2

Zustand (state)



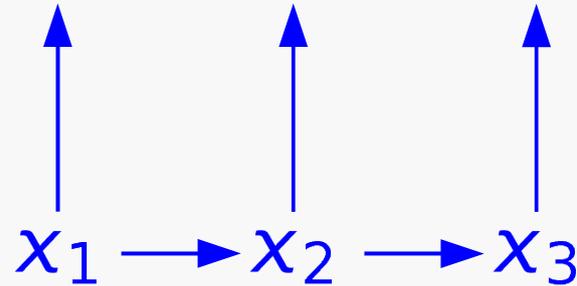
Kalman Filter



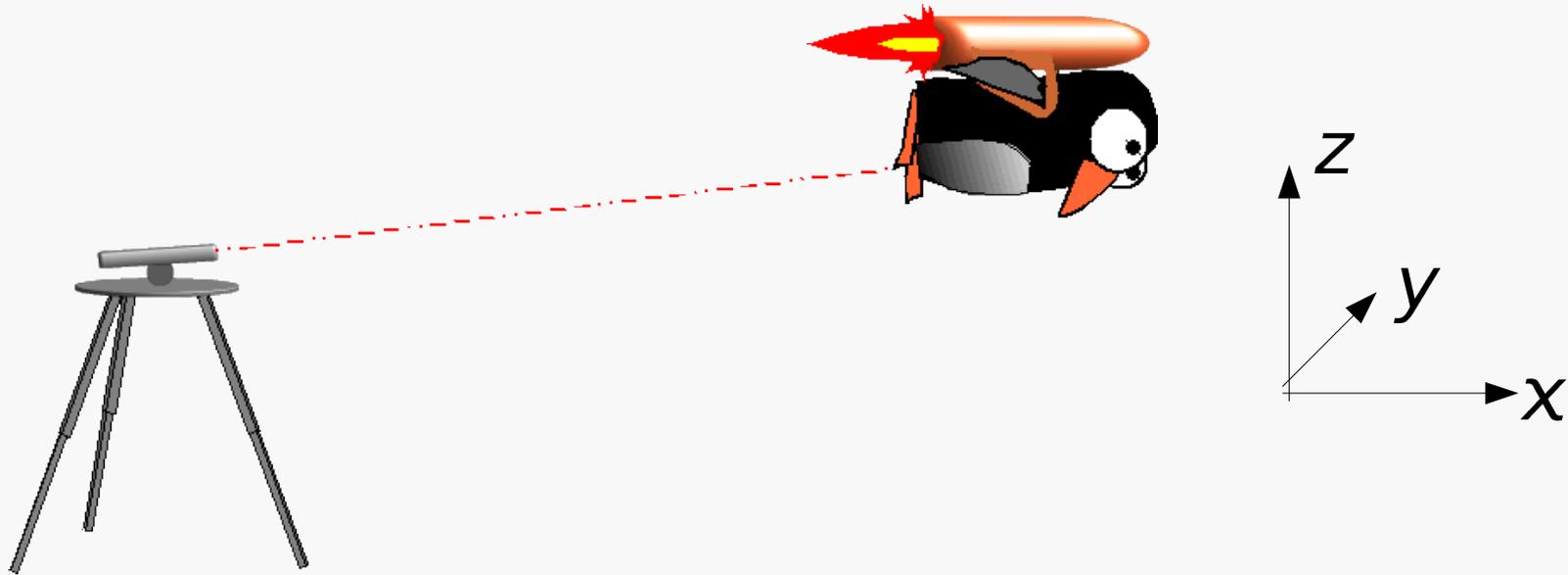
Messung

m_1 m_2 m_3

Zustand (state)



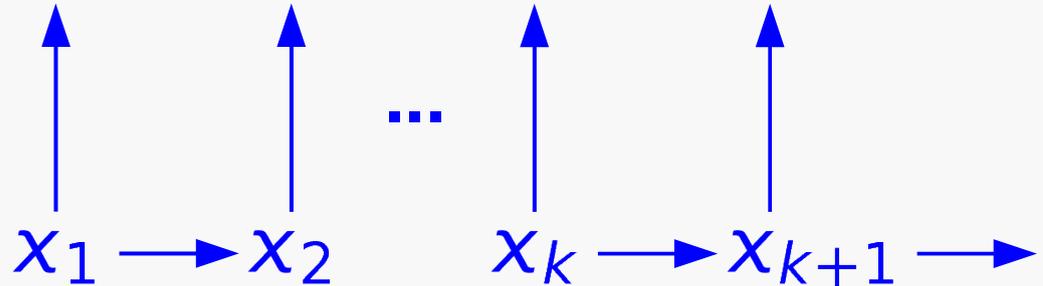
Kalman Filter



Messung

m_1 m_2 ... m_k m_{k+1}

Zustand (state)



Exkurs deterministische lineare Systeme in der klassischen Mechanik

Beispiel gedämpfter Oszillator

$$m\ddot{x} + \rho\dot{x} + kx = 0$$

Masse m , Reibung ρ und Federkonstante k .

DGL 2. Ordnung \rightarrow 2 Gleichungen 1. Ordnung

$$\frac{d}{dt} \begin{pmatrix} x \\ \dot{x} \end{pmatrix} = \begin{bmatrix} 0 & 1 \\ -k/m & -\rho/m \end{bmatrix} \begin{pmatrix} x \\ \dot{x} \end{pmatrix}$$

Phasenraum Bahn

Diskretisierung $\Delta t, t_{k+1} = t_k + \Delta t$

$$\vec{x}_{k+1} = \vec{x}_k + \Delta t \frac{d}{dt} \vec{x}_k = \begin{bmatrix} 1 & \Delta t \\ -\frac{k\Delta t}{m} & 1 - \frac{\rho\Delta t}{m} \end{bmatrix} \vec{x}_k$$



Stochastische lineare Systeme



Die Bewegung erhält eine **Zufallskomponente** - Streuung

System matrix

Zustand / state $x_k = Ax_{k-1} + \eta_k$ Process noise

Messung $m_k = Hx_k + \epsilon_k$ Measurement noise

Measurement matrix

The diagram shows the state transition equation $x_k = Ax_{k-1} + \eta_k$ and the measurement equation $m_k = Hx_k + \epsilon_k$. Arrows point from the labels 'System matrix', 'Process noise', 'Measurement matrix', and 'Measurement noise' to their respective terms in the equations.

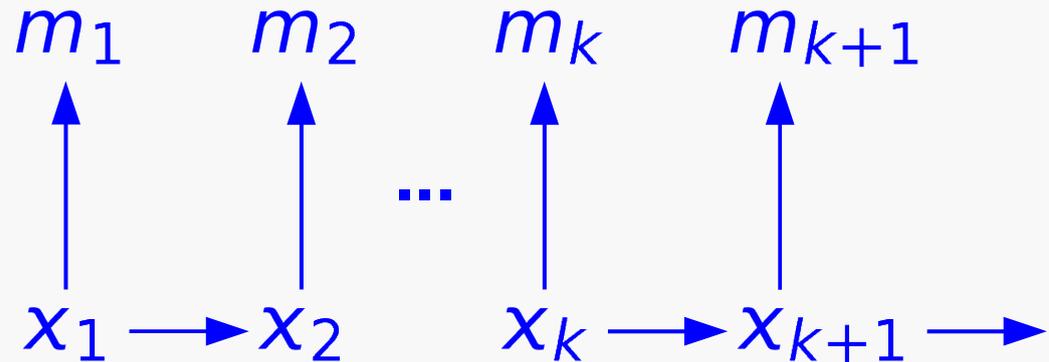
Stochastische lineare Systeme



Die Bewegung erhält eine **Zufallskomponente** - Streuung

Messung

Zustand (state)



Hidden - der Zustand ist nur Indirekt (über die Messung) zugänglich.

Hidden Markov model

Markov – das System hat kein Gedächtnis

Kalman Filter

Spezialisierung:

- Lineares System mit gaußschem Rauschen
- Wir wollen wissen, wo sich das System gerade befindet

➔ **Kalman Filter**

System- und Messungsmodell

$$x_k = Ax_{k-1} + \eta_k$$

$$m_k = Hx_k + \epsilon_k$$

O.B.d.A.

$$\langle \eta_k \rangle = 0, \langle \epsilon_k \rangle = 0$$

Kalman Filter

System- und Messungsmodell

$$x_k = Ax_{k-1} + \eta_k$$

$$m_k = Hx_k + \epsilon_k$$

Korrelationen in den Noise-Termen werden durch die entsprechenden Kovariansmatrizen dargestellt:

$$\langle \eta_k \rangle = 0, \quad \text{cov}(\eta_k) =: Q_k$$

$$\langle \epsilon_k \rangle = 0, \quad \text{cov}(\epsilon_k) =: R_k$$

$$\text{cov}(x) \equiv \langle (x - \langle x \rangle)(x - \langle x \rangle)^T \rangle$$

Kalman Filter

Ohne die neue Messung (a priori)

$$x_k^- := \langle x_k \rangle = A \langle x_{k-1} \rangle + \underbrace{\langle \eta_k \rangle}_{=0}$$

Die Ungewissheit im Ausgangszustand

$$C_{k-1} := \text{cov}(x_{k-1})$$

Die Kovarianz des neuen Zustandes vor der Messung

$$C_k^- := \text{cov}(x_k^-) = A C_{k-1} A^T + Q_k$$

Kalman Filter

Die neue Messung wird mit Bayes assimiliert

$$p(x_k|m_k) = \frac{p(m_k|x_k)p(x_k)}{p(m_k)}$$

Optimaler Schätzwert als Maximum der Wahrscheinlichkeit

$$\begin{aligned} \partial_{x_k} p(x_k|m_k) &= 0 \\ \sim \partial_{x_k} [p(m_k|x_k)p(x_k)] \end{aligned}$$

n-dim Gleichungssystem

Kalman Filter

n -dimensionale Normalverteilung

A priori Wahrscheinlichkeit für den Zustand:

$$\begin{aligned} p(x_k) &= g(x_k | \mu = x_k^-, \Sigma = C_k^-) \\ &= \frac{1}{\sqrt[n]{2\pi} \sqrt{|\Sigma|}} e^{-\frac{1}{2}(x_k - x_k^-)^T (C_k^-)^{-1} (x_k - x_k^-)} \end{aligned}$$

Likelihood der Messung:

$$\begin{aligned} p(m_k | x_k) &= g(m_k | \mu = Hx_k, \Sigma = R_k) \\ &= \frac{1}{\sqrt[m]{2\pi} \sqrt{|\Sigma|}} e^{-\frac{1}{2}(m_k - Hx_k)^T (R_k)^{-1} (m_k - Hx_k)} \end{aligned}$$

$$\partial_x (x - \mu)^T A (x - \mu) \stackrel{A=A^T}{=} 2A(x - \mu)$$

$$\partial_x (\mu - Hx)^T A (\mu - Hx) \stackrel{A=A^T}{=} 2H^T A (Hx - \mu)$$

Kalman Filter

$$\begin{aligned}
 0 &= \partial_{x_k} [p(m_k|x_k)p(x_k)] \\
 &= p(m_k|x_k)p(x_k) \\
 &\quad \cdot \left[(C_k^-)^{-1}(x_k - x_k^-) + H^T R^{-1}(Hx_k - m_k) \right]
 \end{aligned}$$

Auflösen nach x_k

$$x_k = \left\{ (C_k^-)^{-1} + H^T R^{-1} H \right\}^{-1} \left\{ (C_k^-)^{-1} x_k^- + H^T R^{-1} m_k \right\}$$

$$x_k = x_k^- + K(m_k - Hx_k^-)$$

$$K = C_k^- H^T (H C_k^- H^T + R)^{-1} \quad \text{Gain-Matrix}$$

$$\underbrace{\left((C_k^-)^{-1} + H^T R^{-1} H \right)^{-1} H^T R^{-1}}_{n \times n} = C_k^- H^T \underbrace{\left(H C_k^- H^T + R \right)^{-1}}_{m \times m}$$

Numerisch einfacher für $m < n$

Matrix Cookbook

http://www.imm.dtu.dk/pubdb/views/edoc_download.php/3274/pdf/imm3274.pdf



Kalman Filter Zusammenfassung

Prediction

$$x_k^- = Ax_{k-1}$$

$$C_k^- = AC_{k-1}A^T + Q_k$$

Update

$$x_k = x_k^- + K(m_k - Hx_k^-)$$

$$C_k = (1 - KH)C_k^-$$

$$K = C_k^- H^T (H C_k^- H^T + R)^{-1} \quad \text{Gain-Matrix}$$

Smoothing: back propagation $m_k \rightarrow x_{k-1}$

Anwendungen

- Radarverfolgung von Flugzeugen etc.
- Luftabwehrraketen
- CD-Player
- Bewegungsverfolgung in Filmaufnahmen, in bildgebenden Verfahren in der Medizin
- Wettervorhersage ...
- Teilchendetektoren – CMS
 - Track reconstruction
 - Track Fit
 - Vertex Fit
 - Alignment

1. Verallgemeinerung

Extended Kalman Filter - EKF

- Nichtlineares Zustandsmodell und/oder Modell der Messung

$$x_k = f(x_{k-1}) + \eta_k$$

$$m_k = h(x_k) + \epsilon_k$$

Prediction: $x_k^- = f(x_{k-1})$

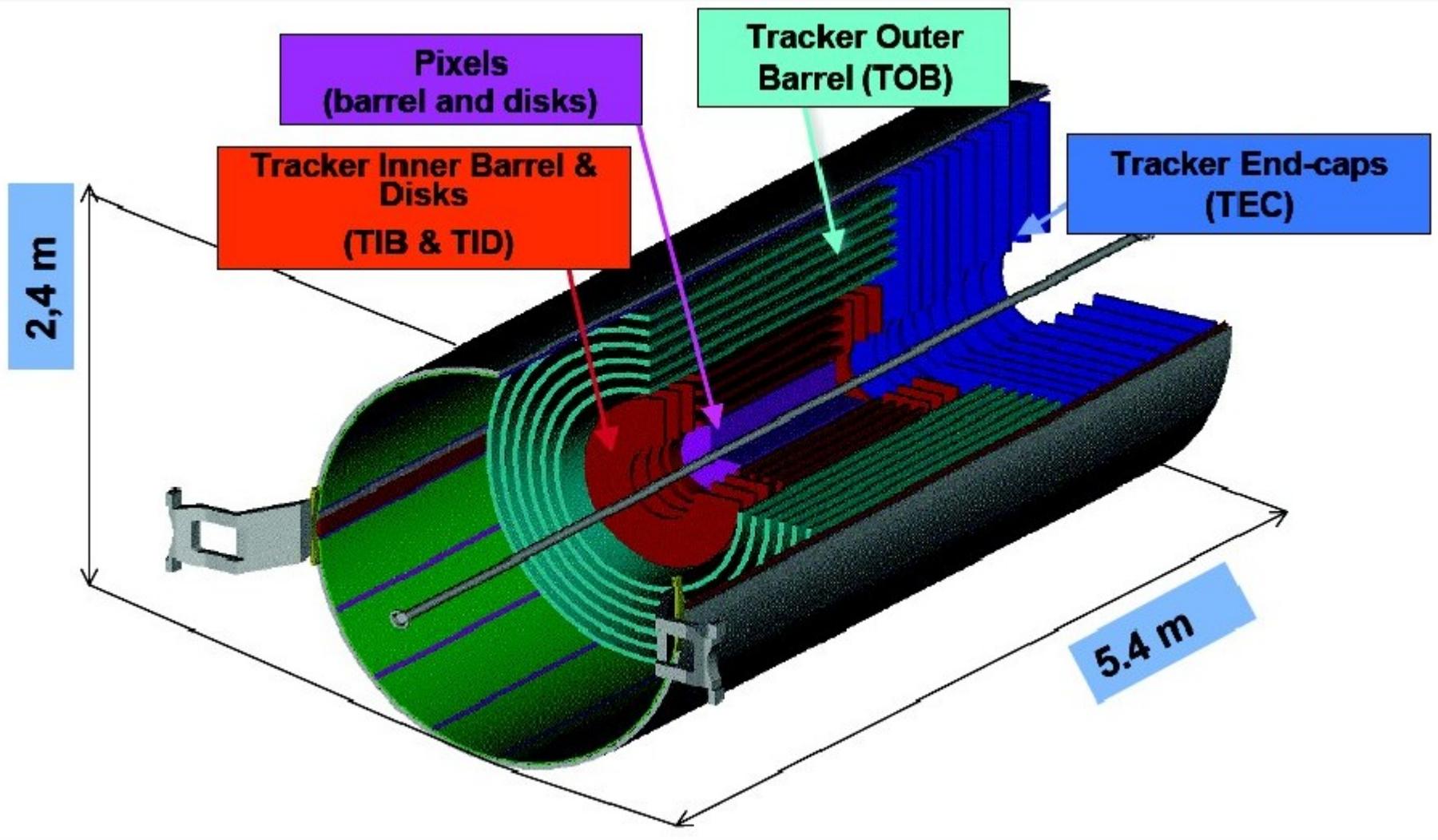
$$C_k^- = A_k C_{k-1} A_k^T + Q_k$$

Update: $x_k = x_k^- + K_k (m_k - h(x_k^-))$

Gain: $K_k = C_k^- H_k^T (H_k C_k^- H_k^T + R_k)$

$$A_{k-1} = \left. \frac{\partial f}{\partial x} \right|_{x_{k-1}}, \quad H_k = \left. \frac{\partial h}{\partial x} \right|_{x_k^-}$$

CMS Tracker

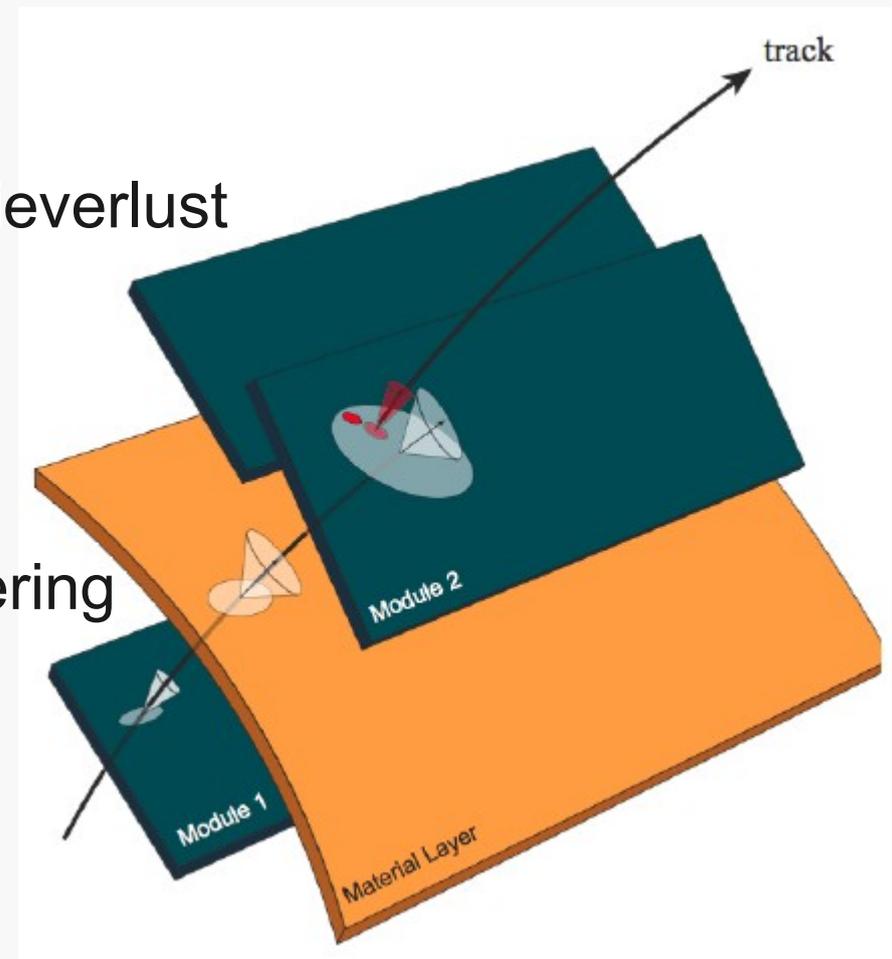


Kalman Filter zur Trackrekonstruktion bei CMS

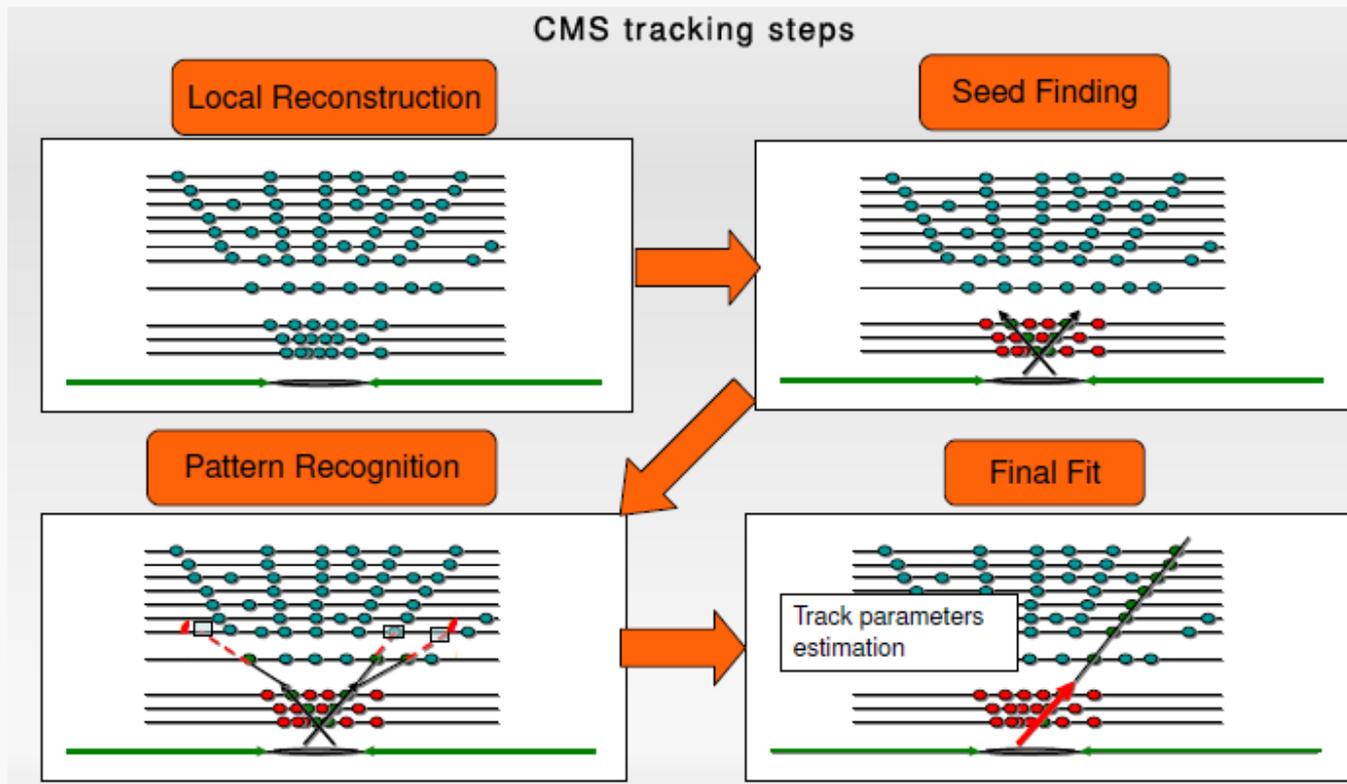
$$x_k = f(x_{k-1})$$

Nichtlinear wegen: B-Feld, Energieverlust

Process noise η : Multiple scattering



Anwendungen bei CMS



Seeding, um Startwert für den Filter zu finden

Der Prediction Schritt des Kalman Filter hilft bei der Rekonstruktion d.h die Hits zu finden, die zu einem Track gehören

Combinatorial Track Finder (CTF)

- Seeding: done using the innermost layers of the tracker (pixels and first layers of TIB).
- Pattern recognition: Kalman Filter based propagation and update of the trajectory parameters on more external layers. Search for hits in narrower windows as the track approaches external layers.
- Final fit: done running two Kalman filter in the opposite senses (inside-out and outside-in).

2. Verallgemeinerung Gaussian Sum Filter

- In manchen Fällen ist die Annahme von weißem Prozessrauschen ungenügend
- Insbesondere bei Elektronen, wo Bremsstrahlung einen stärkeren Einfluss hat
 - Non-gaussian tails
- Der GSF berücksichtigt das durch eine erweitertes Modell des process noise durch Überlagerung mehrerer Gaußverteilungen
- Nachteil langsamer (~ 200)

Gaussian Sum Filter

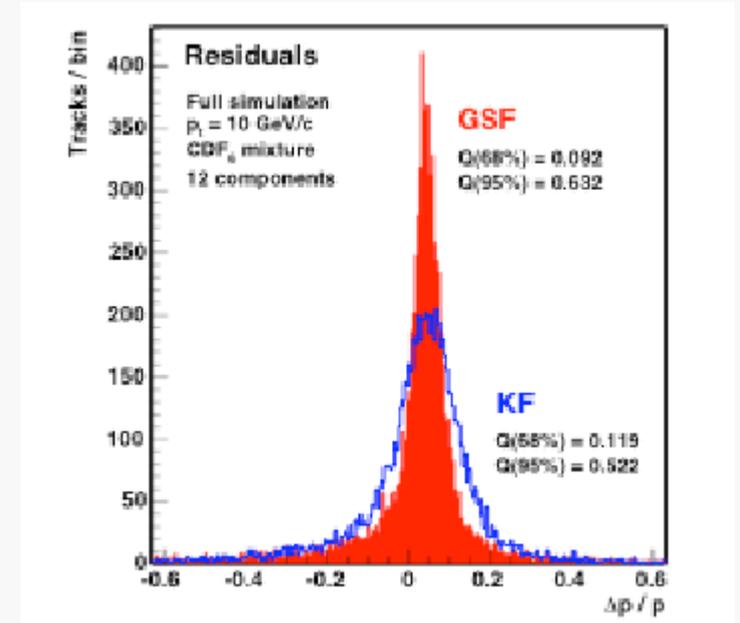
A priori Wahrscheinlichkeit
für den Zustand:

$$p(\mathbf{x}_k) = \sum_j^{N_{k-1}} g(\mathbf{x}_k | \mu = \mathbf{x}_{j,k}^-, \Sigma = C_{j,k}^-)$$

Likelihood der Messung:

$$C_{j,k}^- = AC_{k-1}A^T + Q_{j,k}$$

$$p(m_k | \mathbf{x}_k) = \sum_j^{M_k} g(m_k | \mu = H\mathbf{x}_k, \Sigma = R_{j,k})$$



Siehe: R. Frühwirth, Comp. Phys. Comm. 100 (1997) 1,

Literatur



Wenn man ernsthaft einsteigen will

- D.S.Sivia – Data Analysis – A Bayesian Tutorial
- http://en.wikipedia.org/wiki/Bayesian_probability

Für den „philosophisch“ Interessierten:

- E.T.Jaynes: Probability Theory – The Logic of Science
- R.T.Cox, Am.J.Phys.14(1946)1

Kalman Filter

- http://en.wikipedia.org/wiki/Kalman_filter
- <http://www.cs.unc.edu/~welch/kalman/media/pdf/Kalman1960.pdf>
- Tracking: R. Frühwirth, Nucl. Instrum. and Methods A 262(1987) 444
- GSF: R. Frühwirth, Comp. Phys. Comm. 100 (1997) 1, Comp. Phys. Comm. 110 (1998) 80