Juergen Gall

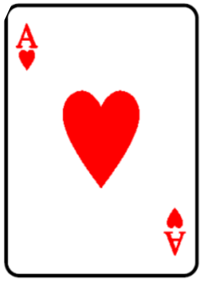
# Background Subtraction and Tracking
## MA-INF 2201 - Computer Vision
## WS24/25

# Grab Cut

Source: K. Grauman

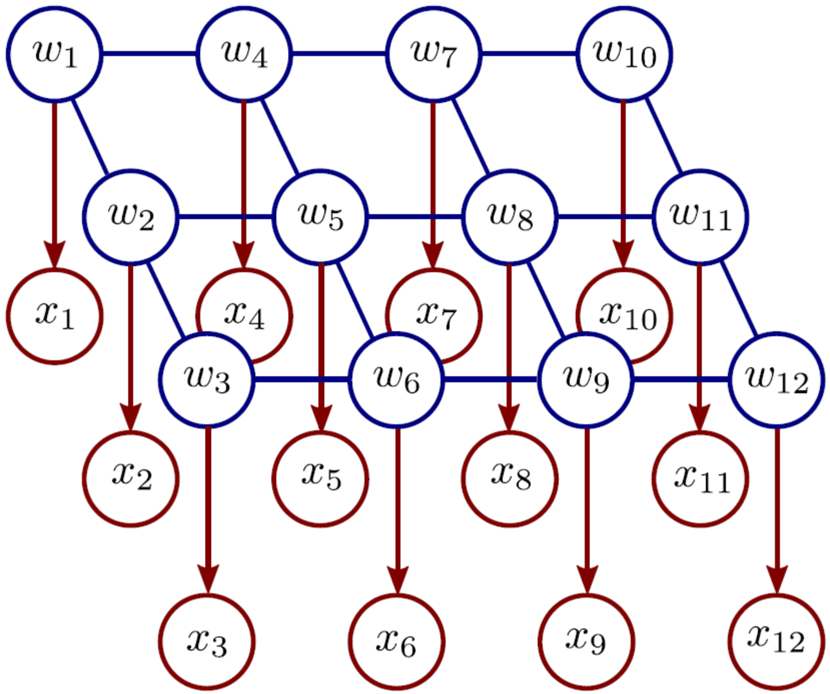
# Denoising with MRFs





Original image, **w**

Observed image, **x**




MRF Prior (pairwise cliques)

$$Pr(w_{1\ldots N}) = \frac{1}{Z}\exp\left[-\sum_{(m,n)\in\mathcal{C}}\psi[w_m, w_n, \boldsymbol{\theta}]\right]$$

Likelihoods

$$\begin{aligned} Pr(x_n|w_n=0) &= \text{Bern}_{x_n}[\rho] \\ Pr(x_n|w_n=1) &= \text{Bern}_{x_n}[1-\rho] \end{aligned}$$

Inference :

$$Pr(w_{1\ldots N}|x_{1\ldots N}) = \frac{\prod_{n=1}^{N} Pr(x_n|w_n)Pr(w_{1\ldots N})}{Pr(x_{1\ldots N})}$$

# Grab Cut

UNIVERSITÄT BONN

- Loosely specify foreground region
- Iterated graph cut


User initialization

**Learning colour distributions**

$$Pr(\mathbf{x}_n|w = j)$$

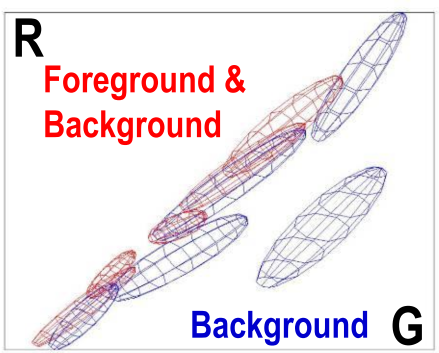
**Graph cuts to infer the segmentation**

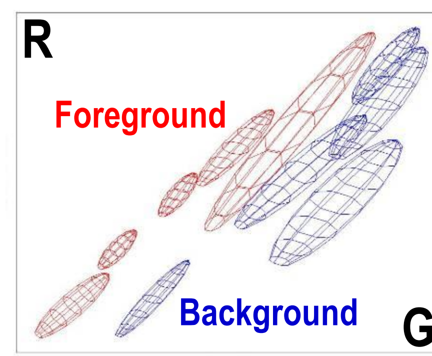
C. Rother et al. **GrabCut - Interactive Foreground Extraction using Iterated Graph Cuts.** SIGGRAPH 2004

Source: K. Grauman

# Grab Cut

- Loosely specify foreground region
- Iterated graph cut



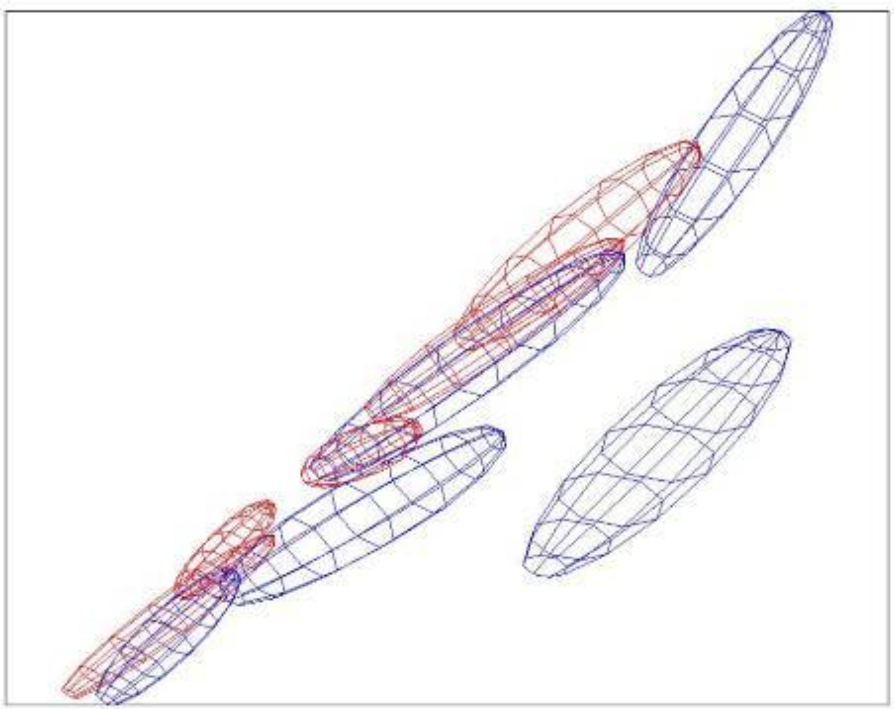

Gaussian Mixture Model (typically 5-8 components)

Source: K. Grauman

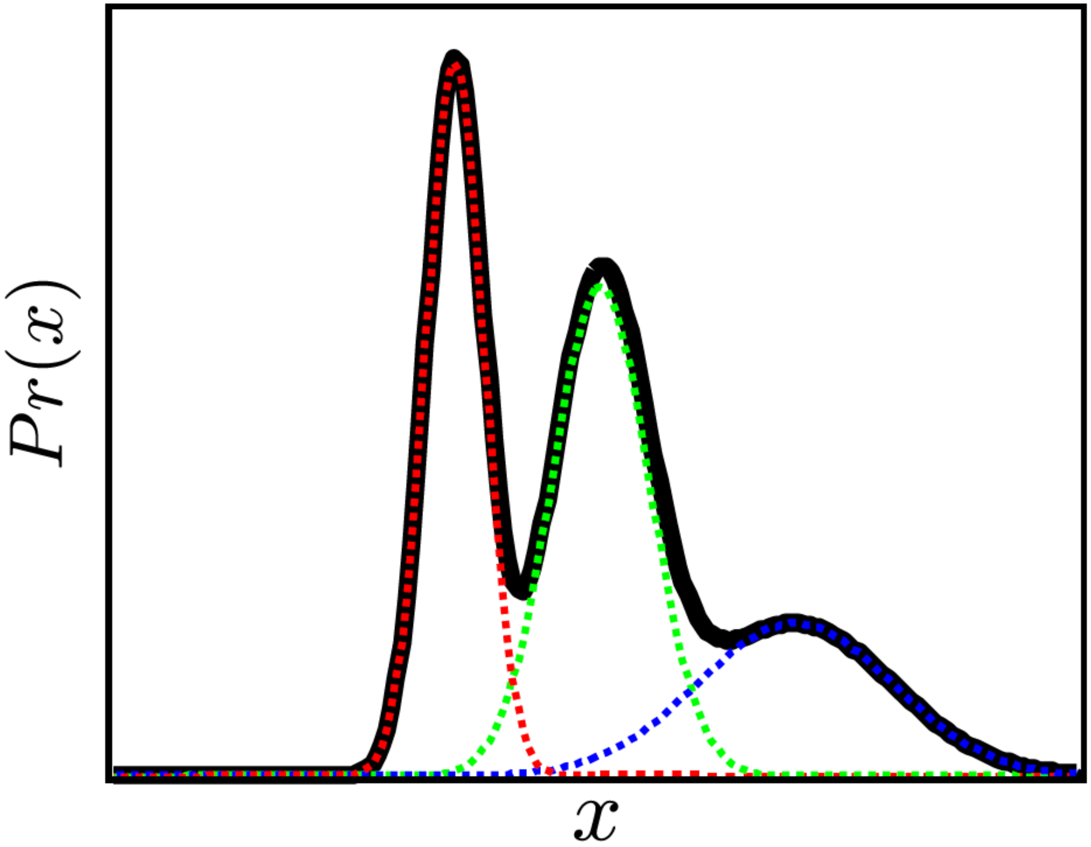
# How do learn Gaussian mixture model?

Expectation-maximization:

$$Pr(\mathbf{x}_n|w=j) = \sum_{k=1}^{K} \lambda_{jk} \text{Norm}_{\mathbf{x}_n}[\boldsymbol{\mu}_{jk}, \boldsymbol{\Sigma}_{jk}]$$



# Mixture of Gaussians (MoG)




$$Pr(\mathbf{x}|\boldsymbol{\theta}) = \sum_{k=1}^{K} \lambda_k \text{Norm}_{\mathbf{x}}[\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k]$$

Source: S. Prince

# MoG as a marginalization

Define a variable $h \in \{1 \ldots K\}$ and then write

$$
\begin{aligned}
Pr(\mathbf{x}|h,\boldsymbol{\theta}) &= \text{Norm}_{\mathbf{x}}[\boldsymbol{\mu}_h, \boldsymbol{\Sigma}_h] \\
Pr(h|\boldsymbol{\theta}) &= \lambda_k
\end{aligned}
$$

Then we can recover the density by marginalizing Pr(**x**,h)

$$
\begin{aligned}
Pr(\mathbf{x}|\boldsymbol{\theta}) &= \sum_{k=1}^{K} Pr(\mathbf{x}, h=k|\boldsymbol{\theta}) \\
&= \sum_{k=1}^{K} Pr(\mathbf{x}|h=k,\boldsymbol{\theta}) Pr(h=k|\boldsymbol{\theta}) \\
&= \sum_{k=1}^{K} \lambda_k \text{Norm}_{\mathbf{x}}[\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k].
\end{aligned}
$$

Source: S. Prince

# MoG as a marginalization

Define a variable $h \in \{1 \ldots K\}$ and then write

$$\begin{aligned} Pr(\mathbf{x}|h, \boldsymbol{\theta}) &= \text{Norm}_{\mathbf{x}}[\boldsymbol{\mu}_h, \boldsymbol{\Sigma}_h] \\ Pr(h|\boldsymbol{\theta}) &= \lambda_k \end{aligned}$$

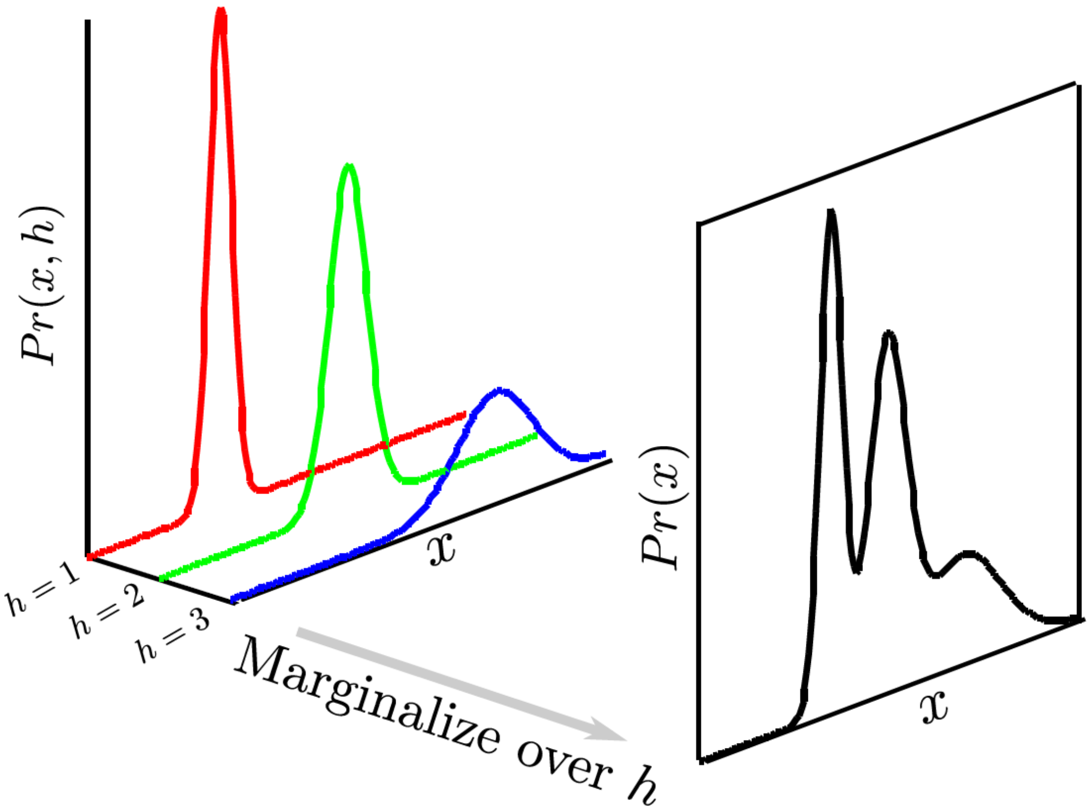
Note:

- This gives us a method to generate data from MoG
  First sample Pr(h), then sample Pr(**x**|h)
- The hidden variable h has a clear interpretation –
  it tells you which Gaussian created data point **x**

Source: S. Prince

# Expectation Maximization for MoG

How do we get $\boldsymbol{\theta} = \{\lambda_{1\ldots K}, \boldsymbol{\mu}_{1\ldots K}, \boldsymbol{\Sigma}_{1\ldots K}\}$ ?

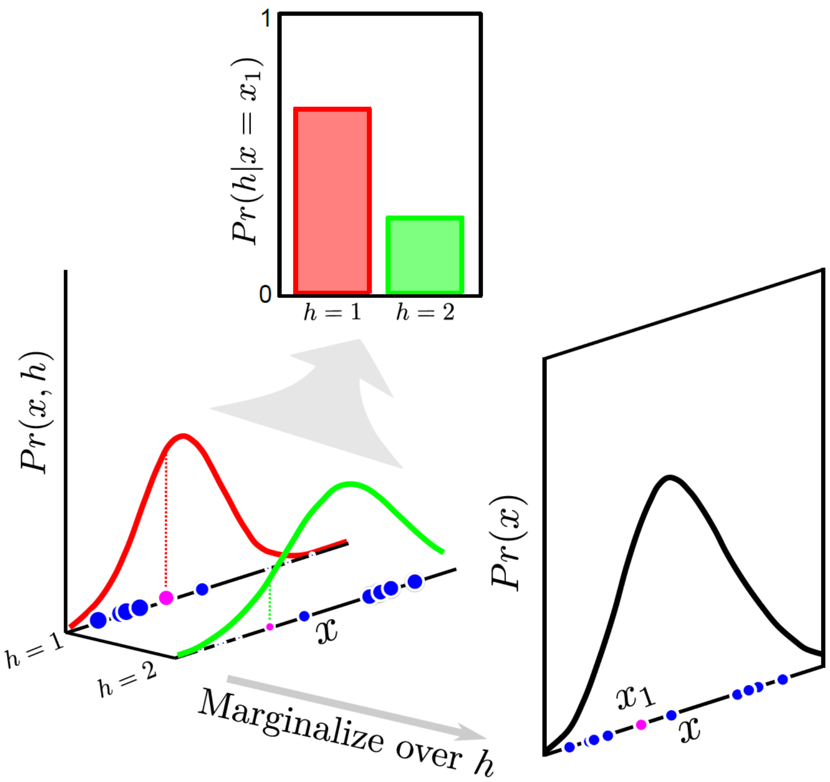



$$Pr(\mathbf{x}|\boldsymbol{\theta}) = \sum_{k=1}^{K} \lambda_k \text{Norm}_{\mathbf{x}}[\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k]$$

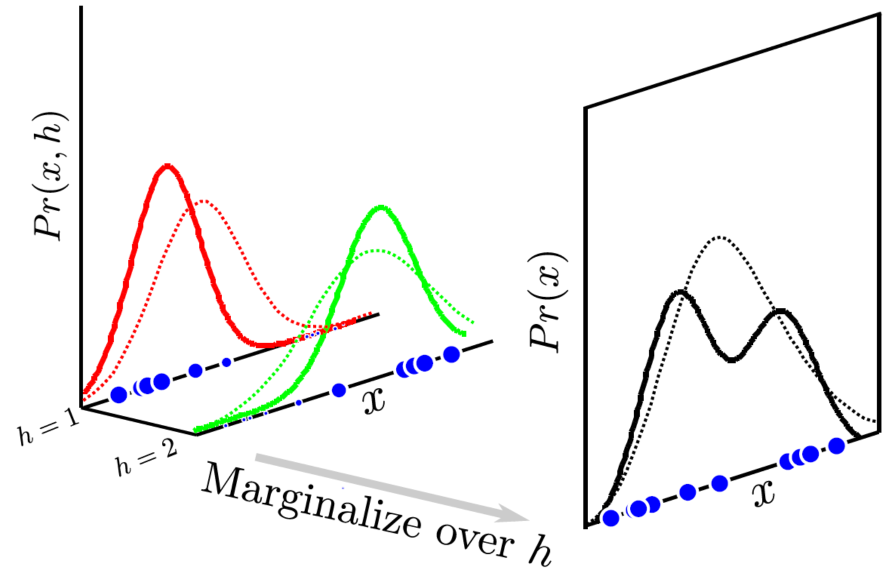
Source: S. Prince

# Expectation Maximization for MoG

- Iterate between Expectation (E) and Maximization (M)




E-Step




M-Step

Source: S. Prince

# Expectation Maximization for MoG

UNIVERSITÄT BONN

GOAL: to learn parameters $\boldsymbol{\theta} = \{\lambda_{1\ldots K}, \boldsymbol{\mu}_{1\ldots K}, \boldsymbol{\Sigma}_{1\ldots K}\}$ from training data $\mathbf{x}_{1\ldots I}$

E-Step – Maximize bound w.r.t. distributions q($\mathbf{h}_i$)

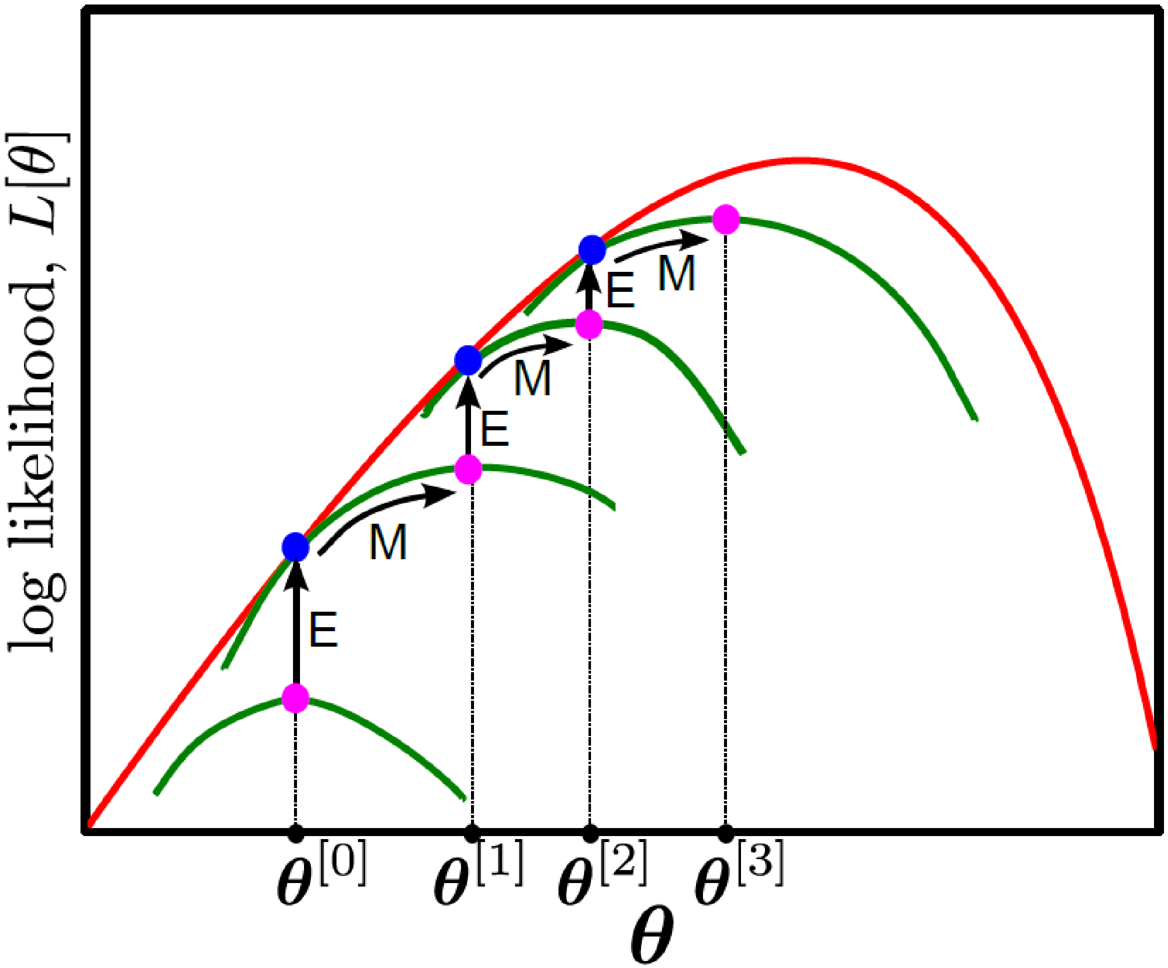
$$\hat{q}_i(\mathbf{h}_i) = Pr(\mathbf{h}_i|\mathbf{x}_i, \boldsymbol{\theta}^{[t]}) = \frac{Pr(\mathbf{x}_i|\mathbf{h}_i, \boldsymbol{\theta}^{[t]})Pr(\mathbf{h}_i|\boldsymbol{\theta}^{[t]})}{Pr(\mathbf{x}_i)}$$

M-Step – Maximize bound w.r.t. parameters θ

$$\hat{\boldsymbol{\theta}}^{[t+1]} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[ \sum_{i=1}^{I} \sum_{k=1}^{K} \hat{q}_i(\mathbf{h}_i = k) \log \left[ Pr(\mathbf{x}_i, \mathbf{h}_i = k|\boldsymbol{\theta}) \right] \right]$$

Source: S. Prince

# Expectation Maximization




Source: S. Prince

# Expectation Maximization for MoG

GOAL: to learn parameters $\boldsymbol{\theta} = \{\lambda_{1\ldots K}, \boldsymbol{\mu}_{1\ldots K}, \boldsymbol{\Sigma}_{1\ldots K}\}$ from training data $\mathbf{x}_{1\ldots I}$

E-Step – Maximize bound w.r.t. distributions q($\mathbf{h}_i$)

$$\hat{q}_i(\mathbf{h}_i) = Pr(\mathbf{h}_i|\mathbf{x}_i, \boldsymbol{\theta}^{[t]}) = \frac{Pr(\mathbf{x}_i|\mathbf{h}_i, \boldsymbol{\theta}^{[t]})Pr(\mathbf{h}_i|\boldsymbol{\theta}^{[t]})}{Pr(\mathbf{x}_i)}$$

Source: S. Prince

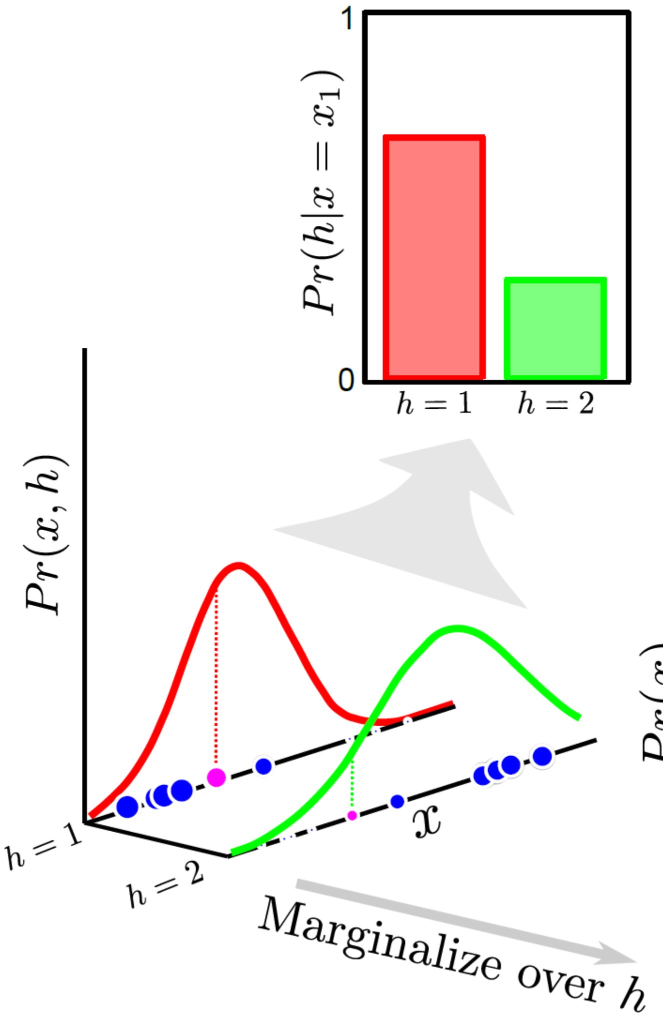
# E-Step

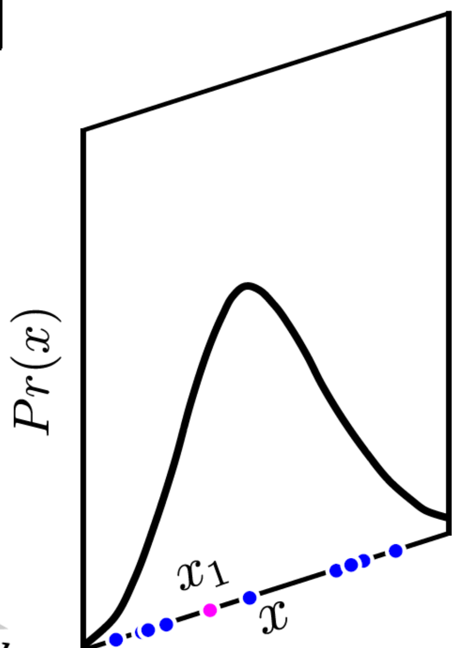
$$Pr(h_i = k|\mathbf{x}_i, \boldsymbol{\theta}^{[t]})$$

$$= \frac{Pr(\mathbf{x}_i|h_i = k, \boldsymbol{\theta}^{[t]})Pr(h_i = k|\boldsymbol{\theta}^{[t]})}{\sum_{j=1}^{K} Pr(\mathbf{x}_i|h_i = j, \boldsymbol{\theta}^{[t]})Pr(h_i = j|\boldsymbol{\theta}^{[t]})}$$

$$= \frac{\lambda_k \text{Norm}_{\mathbf{x}_i}[\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k]}{\sum_{j=1}^{K} \lambda_j \text{Norm}_{\mathbf{x}_i}[\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j]}$$

$$= r_{ik}$$




We'll call this the responsibility of the $k^{\text{th}}$ Gaussian for the $i^{\text{th}}$ data point

Repeat this procedure for every datapoint!

Source: S. Prince

# Expectation Maximization for MoG



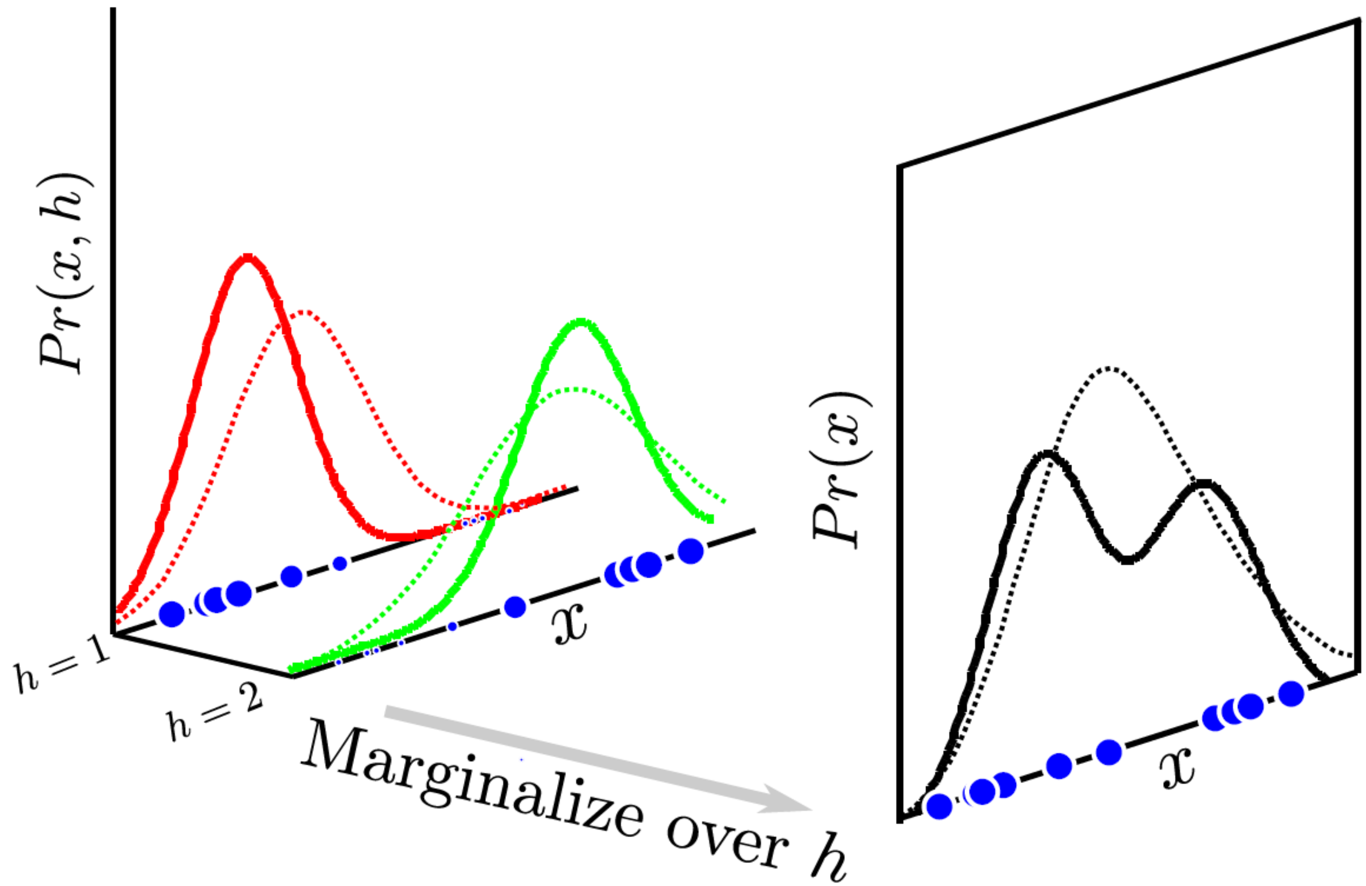
GOAL: to learn parameters $\boldsymbol{\theta} = \{\lambda_{1\ldots K}, \boldsymbol{\mu}_{1\ldots K}, \boldsymbol{\Sigma}_{1\ldots K}\}$ from training data $\mathbf{x}_{1\ldots I}$

M-Step – Maximize bound w.r.t. parameters θ

$$\hat{\boldsymbol{\theta}}^{[t+1]} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[ \sum_{i=1}^{I} \sum_{k=1}^{K} \hat{q}_i(\mathbf{h}_i = k) \log \left[ Pr(\mathbf{x}_i, \mathbf{h}_i = k | \boldsymbol{\theta}) \right] \right]$$

Source: S. Prince

# M-Step




Update means, covariances and weights according to responsibilities of datapoints

Source: S. Prince

# M-Step

$$
\begin{aligned}
\hat{\boldsymbol{\theta}}^{[t+1]} &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[ \sum_{i=1}^{I} \sum_{k=1}^{K} \hat{q}_i(h_i = k) \log \left[ Pr(\mathbf{x}_i, h_i = k | \boldsymbol{\theta}) \right] \right] \\
&= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[ \sum_{i=1}^{I} \sum_{k=1}^{K} r_{ik} \log \left[ \lambda_k \mathrm{Norm}_{\mathbf{x}_i}[\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k] \right] \right].
\end{aligned}
$$

Take derivative, equate to zero and solve (Lagrange multipliers for $\lambda$)

$$
\begin{aligned}
\lambda_k^{[t+1]} &= \frac{\sum_{i=1}^{I} r_{ik}}{\sum_{j=1}^{K} \sum_{i=1}^{I} r_{ij}} \\
\boldsymbol{\mu}_k^{[t+1]} &= \frac{\sum_{i=1}^{I} r_{ik} \mathbf{x}_i}{\sum_{i=1}^{I} r_{ik}} \\
\boldsymbol{\Sigma}_k^{[t+1]} &= \frac{\sum_{i=1}^{I} r_{ik} (\mathbf{x}_i - \boldsymbol{\mu}_k^{[t+1]})(\mathbf{x}_i - \boldsymbol{\mu}_k^{[t+1]})^T}{\sum_{i=1}^{I} r_{ik}}
\end{aligned}
$$

Source: S. Prince

# Derivatives

Scalar x, vector **x**, matrix **X**:

$$y = f(x) \quad \frac{\partial f}{\partial x}$$

$$y = f(\mathbf{x}) \quad \frac{\partial f}{\partial \mathbf{x}} = \left( \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_N} \right)$$

$$y = f(\mathbf{X}) \quad \frac{\partial f}{\partial \mathbf{X}} = \begin{pmatrix} \frac{\partial f}{\partial x_{11}} & \cdots & \frac{\partial f}{\partial x_{1N}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial x_{M1}} & \cdots & \frac{\partial f}{\partial x_{MN}} \end{pmatrix}$$

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

# Derivatives

$$\frac{\partial \mathbf{x}^T \mathbf{a}}{\partial \mathbf{x}} = \mathbf{a}$$

$$\frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$$

$$\frac{\partial \mathbf{a}^T \mathbf{X} \mathbf{b}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T$$

$$\frac{\partial \mathbf{a}^T \mathbf{X}^T \mathbf{b}}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^T$$

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

$$\frac{\partial (\mathbf{B}\mathbf{x} + \mathbf{b})^T \mathbf{C} (\mathbf{D}\mathbf{x} + \mathbf{d})}{\partial \mathbf{x}} = \mathbf{B}^T \mathbf{C} (\mathbf{D}\mathbf{x} + \mathbf{d}) + \mathbf{D}^T \mathbf{C}^T (\mathbf{B}\mathbf{x} + \mathbf{b})$$

$$\frac{\partial \mathbf{x}^T \mathbf{B} \mathbf{x}}{\partial \mathbf{x}} = (\mathbf{B} + \mathbf{B}^T)\mathbf{x}$$

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

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

# M-Step

$$\hat{\boldsymbol{\theta}}^{[t+1]} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[ \sum_{i=1}^{I} \sum_{k=1}^{K} \hat{q}_i(h_i = k) \log \left[ Pr(\mathbf{x}_i, h_i = k | \boldsymbol{\theta}) \right] \right]$$

$$= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[ \sum_{i=1}^{I} \sum_{k=1}^{K} r_{ik} \log \left[ \lambda_k \text{Norm}_{\mathbf{x}_i} [\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k] \right] \right].$$

Take derivative, equate to zero and solve (Lagrange multipliers for $\lambda$)

$$\lambda_k^{[t+1]} = \frac{\sum_{i=1}^{I} r_{ik}}{\sum_{j=1}^{K} \sum_{i=1}^{I} r_{ij}}$$

$$\boldsymbol{\mu}_k^{[t+1]} = \frac{\sum_{i=1}^{I} r_{ik} \mathbf{x}_i}{\sum_{i=1}^{I} r_{ik}}$$

$$\boldsymbol{\Sigma}_k^{[t+1]} = \frac{\sum_{i=1}^{I} r_{ik} (\mathbf{x}_i - \boldsymbol{\mu}_k^{[t+1]})(\mathbf{x}_i - \boldsymbol{\mu}_k^{[t+1]})^T}{\sum_{i=1}^{I} r_{ik}}$$

Source: S. Prince

# Lagrange multiplier

- Optimize $\lambda_1, \ldots, \lambda_K$

$$\underset{\lambda_1,\ldots,\lambda_K}{\operatorname{argmax}} \left[ \sum_{i=1}^{I} \sum_{k=1}^{K} r_{ik} \log \left[ \lambda_k \text{Norm}_{\mathbf{x}_i} [\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k] \right] \right]$$

- Subject to constraint:

$$\sum_{k=1}^{K} \lambda_k = 1$$

# Lagrange multiplier

- Maximize $\lambda_1, \ldots, \lambda_K$

$$f(\lambda_1, \ldots, \lambda_K) = \sum_{i=1}^{I} \sum_{k=1}^{K} r_{ik} \log \left[\lambda_k \text{Norm}_{\mathbf{x}_i}[\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k]\right]$$

- Subject to constraint:

$$\sum_{k=1}^{K} \lambda_k - 1 = 0$$

# Lagrange multiplier

- Rewrite without constraints
  using Lagrange multiplier $\lambda$:

$$f(\lambda_1, \ldots, \lambda_K, \lambda) = \sum_i \sum_k r_{ik} \log\left(\lambda_k \mathcal{N}(x_i; \mu_k, \Sigma_k)\right) + \lambda \left( \sum_k \lambda_k - 1 \right)$$

- Compute gradients gives K+1 equations:

$$\frac{\partial}{\partial \lambda_k} \sum_i \sum_k r_{ik} \log\left(\lambda_k \mathcal{N}(x_i; \mu_k, \Sigma_k)\right) + \lambda \left( \sum_k \lambda_k - 1 \right) = 0 \qquad \forall k$$

$$\frac{\partial}{\partial \lambda} \sum_i \sum_k r_{ik} \log\left(\lambda_k \mathcal{N}(x_i; \mu_k, \Sigma_k)\right) + \lambda \left( \sum_k \lambda_k - 1 \right) = 0$$

# Lagrange multiplier

- Compute gradients

$$\frac{\partial}{\partial \lambda_k} \sum_i \sum_k r_{ik} \log\left(\lambda_k \mathcal{N}(x_i;\mu_k,\Sigma_k)\right) + \lambda \left(\sum_k \lambda_k - 1\right) = 0 \qquad \forall k$$

$$\frac{\partial}{\partial \lambda} \sum_i \sum_k r_{ik} \log\left(\lambda_k \mathcal{N}(x_i;\mu_k,\Sigma_k)\right) + \lambda \left(\sum_k \lambda_k - 1\right) = 0$$

- Gives K+1 equations:

$$\sum_i \frac{r_{ik}}{\lambda_k} + \lambda = 0 \qquad \forall k$$

$$\sum_{k=1}^{K} \lambda_k - 1 = 0$$

# Lagrange multiplier

- Gives K+1 equations

$$\sum_i \frac{r_{ik}}{\lambda_k} + \lambda = 0 \qquad \forall k \qquad\qquad \sum_{k=1}^{K} \lambda_k - 1 = 0$$

- We therefore get by summing over all K equations:

$$\sum_k \sum_i r_{ik} = -\lambda \sum_k \lambda_k = -\lambda$$

$$\lambda_k = \frac{\sum_i r_{ik}}{\sum_k \sum_i r_{ik}}$$

# Proof

$$\frac{\partial}{\partial \mu_k}\left\{\sum_i \sum_k r_{ik} \log\left(\frac{\lambda_k}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x_i - \mu_k)^T \Sigma_k^{-1}(x_i - \mu_k)\right)\right)\right\} = 0$$

$$\frac{\partial}{\partial \mu_k}\left\{\sum_i r_{ik}\left(-\frac{1}{2}(x_i - \mu_k)^T \Sigma_k^{-1}(x_i - \mu_k)\right)\right\} = 0$$

# Derivatives

$$\frac{\partial \mathbf{x}^T\mathbf{a}}{\partial \mathbf{x}} = \mathbf{a}$$

$$\frac{\partial \mathbf{a}^T\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}\mathbf{b}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T$$

$$\frac{\partial \mathbf{a}^T\mathbf{X}^T\mathbf{b}}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^T$$

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

$$\frac{\partial (\mathbf{B}\mathbf{x}+\mathbf{b})^T\mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{d})}{\partial \mathbf{x}} = \mathbf{B}^T\mathbf{C}(\mathbf{D}\mathbf{x}+\mathbf{d}) + \mathbf{D}^T\mathbf{C}^T(\mathbf{B}\mathbf{x}+\mathbf{b})$$

$$\frac{\partial \mathbf{x}^T\mathbf{B}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{B}+\mathbf{B}^T)\mathbf{x}$$

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

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

# Proof

$$\frac{\partial}{\partial \mu_k} \left\{ \sum_i \sum_k r_{ik} \log \left( \frac{\lambda_k}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} \exp \left( -\frac{1}{2} (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) \right) \right) \right\} = 0$$

$$\frac{\partial}{\partial \mu_k} \left\{ \sum_i r_{ik} \left( -\frac{1}{2} (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) \right) \right\} = 0$$

$$\sum_i r_{ik} \Sigma_k^{-1} (x_i - \mu_k) = 0$$

$$\frac{\partial (\mathbf{B}\mathbf{x} + \mathbf{b})^T \mathbf{C} (\mathbf{D}\mathbf{x} + \mathbf{d})}{\partial \mathbf{x}} = \mathbf{B}^T \mathbf{C} (\mathbf{D}\mathbf{x} + \mathbf{d}) + \mathbf{D}^T \mathbf{C}^T (\mathbf{B}\mathbf{x} + \mathbf{b})$$

# Proof

$$\frac{\partial}{\partial \mu_k}\left\{\sum_i \sum_k r_{ik} \log\left(\frac{\lambda_k}{(2\pi)^{d/2}|\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2}(x_i-\mu_k)^T \Sigma_k^{-1}(x_i-\mu_k)\right)\right)\right\} = 0$$

$$\frac{\partial}{\partial \mu_k}\left\{\sum_i r_{ik}\left(-\frac{1}{2}(x_i-\mu_k)^T \Sigma_k^{-1}(x_i-\mu_k)\right)\right\} = 0$$

$$\sum_i r_{ik}\Sigma_k^{-1}(x_i-\mu_k) = 0$$

$$\Sigma_k \Sigma_k^{-1} \sum_i r_{ik} x_i = \Sigma_k \Sigma_k^{-1} \mu_k \sum_i r_{ik}$$

$$\mu_k = \frac{\sum_i r_{ik} x_i}{\sum_i r_{ik}}$$

# Proof

$$\frac{\partial}{\partial \Sigma_k^{-1}} \left\{ \sum_i \sum_k r_{ik} \log \left( \frac{\lambda_k}{(2\pi)^{d/2} |\Sigma_k|^{1/2}} \exp \left( -\frac{1}{2} (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) \right) \right) \right\} = 0$$

$$\frac{\partial}{\partial \Sigma_k^{-1}} \left\{ \sum_i r_{ik} \frac{1}{2} \log \left( \frac{1}{|\Sigma_k|} \right) + r_{ik} \left( -\frac{1}{2} (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) \right) \right\} = 0$$

Using:

$$x^T A x = \mathrm{Tr}(A x x^T)$$

$$|A^{-1}| = \frac{1}{|A|}$$

We get:

$$\frac{\partial}{\partial \Sigma_k^{-1}} \left\{ \sum_i r_{ik} \log \left( |\Sigma_k^{-1}| \right) - r_{ik} \,\mathrm{Tr} \left( \Sigma_k^{-1} B_{ik} \right) \right\} = 0 \quad B_{ik} = (x_i - \mu_k)(x_i - \mu_k)^T$$

# Proof

$$\frac{\partial}{\partial \Sigma_k^{-1}} \left\{ \sum_i r_{ik} \log\left(|\Sigma_k^{-1}|\right) - r_{ik} \operatorname{Tr}\left(\Sigma_k^{-1} B_{ik}\right) \right\} = 0 \quad B_{ik} = (x_i - \mu_k)(x_i - \mu_k)^T$$

Using:

$$\frac{\partial}{\partial A} \operatorname{Tr}(AB) = B^T$$

$$\frac{\partial}{\partial A} \log(|A|) = (A^{-1})^T$$

We get:

$$\sum_i r_{ik} (\Sigma_k^T - B_{ik}^T) = 0$$

$$\sum_i r_{ik} (\Sigma_k - B_{ik}) = 0$$

$$\Sigma_k = \frac{\sum_i r_{ik} (x_i - \mu_k)(x_i - \mu_k)^T}{\sum_i r_{ik}}$$

# M-Step

$$\hat{\boldsymbol{\theta}}^{[t+1]} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[ \sum_{i=1}^{I} \sum_{k=1}^{K} \hat{q}_i(h_i = k) \log \left[ Pr(\mathbf{x}_i, h_i = k | \boldsymbol{\theta}) \right] \right]$$

$$= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[ \sum_{i=1}^{I} \sum_{k=1}^{K} r_{ik} \log \left[ \lambda_k \mathrm{Norm}_{\mathbf{x}_i}[\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k] \right] \right].$$

Take derivative, equate to zero and solve (Lagrange multipliers for λ)

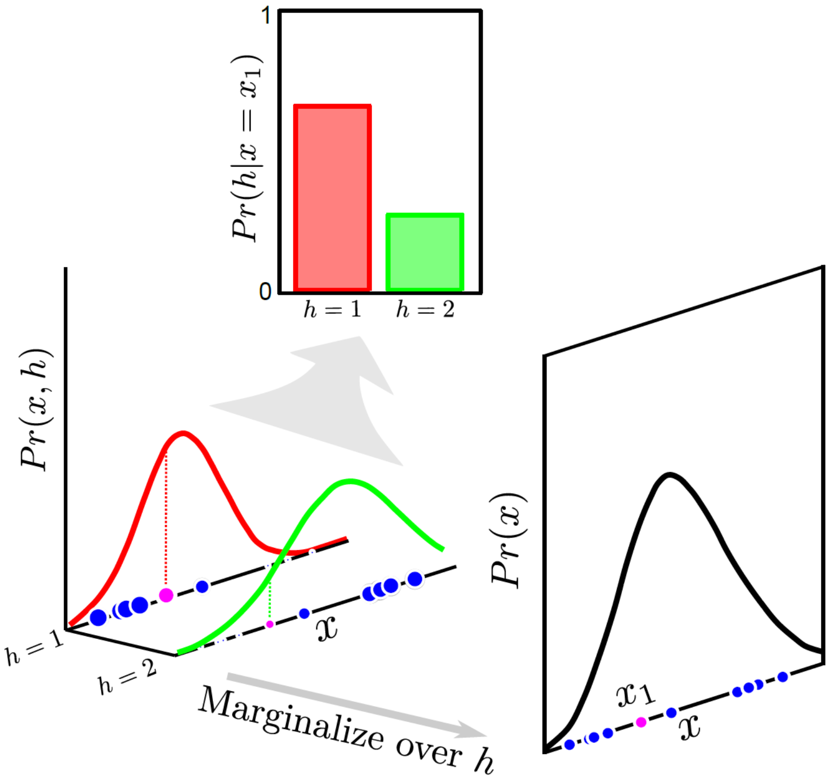
$$\lambda_k^{[t+1]} = \frac{\sum_{i=1}^{I} r_{ik}}{\sum_{j=1}^{K} \sum_{i=1}^{I} r_{ij}}$$

$$\boldsymbol{\mu}_k^{[t+1]} = \frac{\sum_{i=1}^{I} r_{ik} \mathbf{x}_i}{\sum_{i=1}^{I} r_{ik}}$$

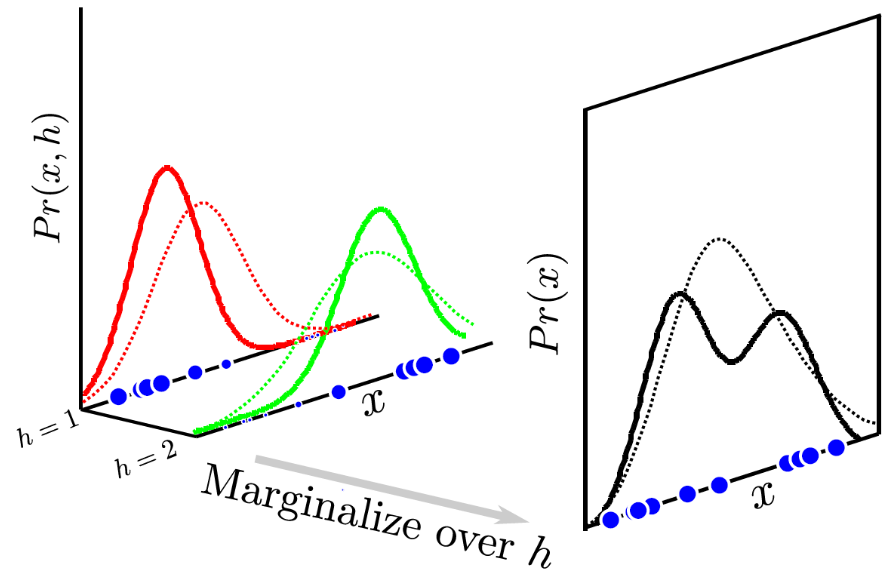
$$\boldsymbol{\Sigma}_k^{[t+1]} = \frac{\sum_{i=1}^{I} r_{ik} (\mathbf{x}_i - \boldsymbol{\mu}_k^{[t+1]})(\mathbf{x}_i - \boldsymbol{\mu}_k^{[t+1]})^T}{\sum_{i=1}^{I} r_{ik}}$$

Source: S. Prince

# Iterate until no further improvement

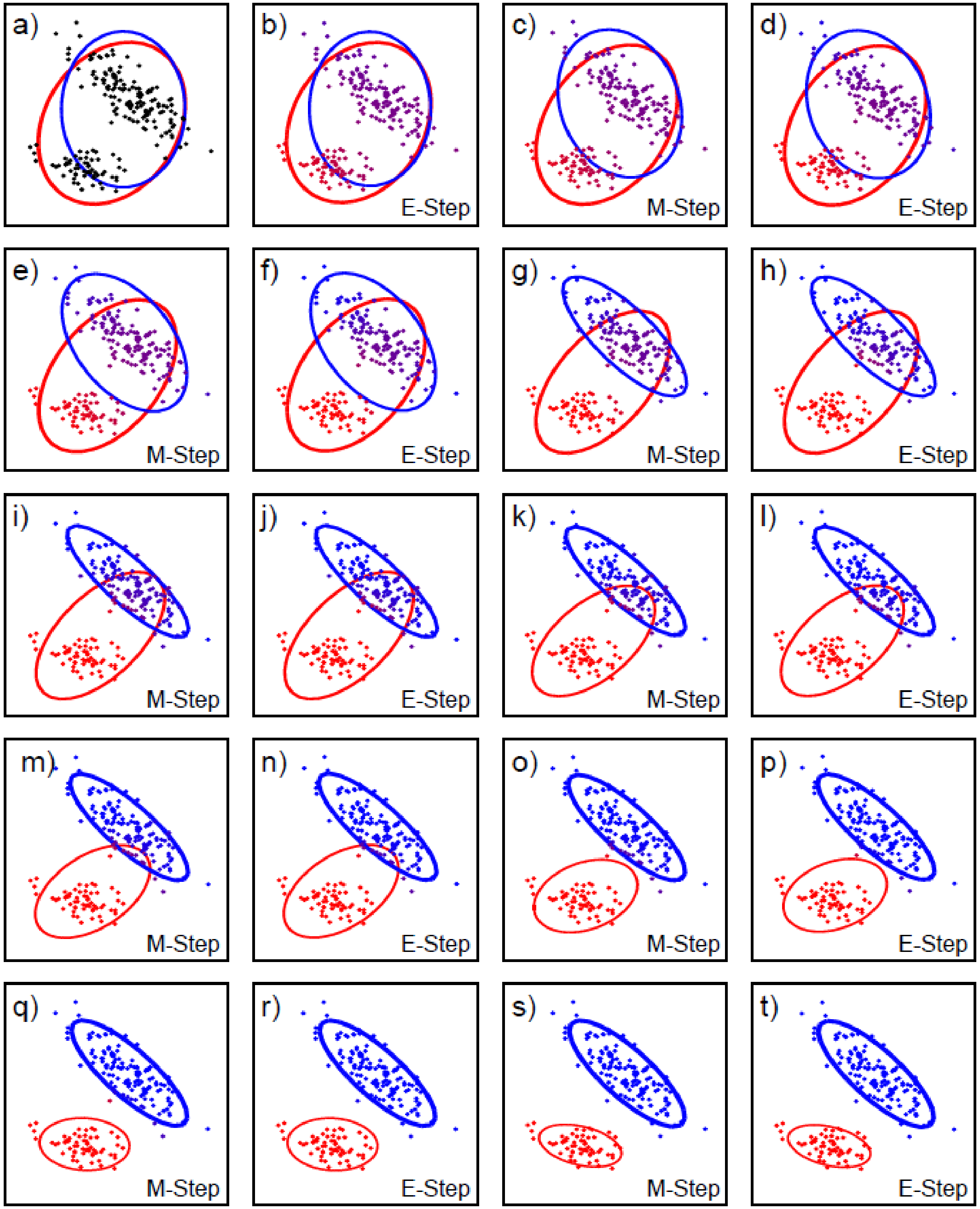



E-Step



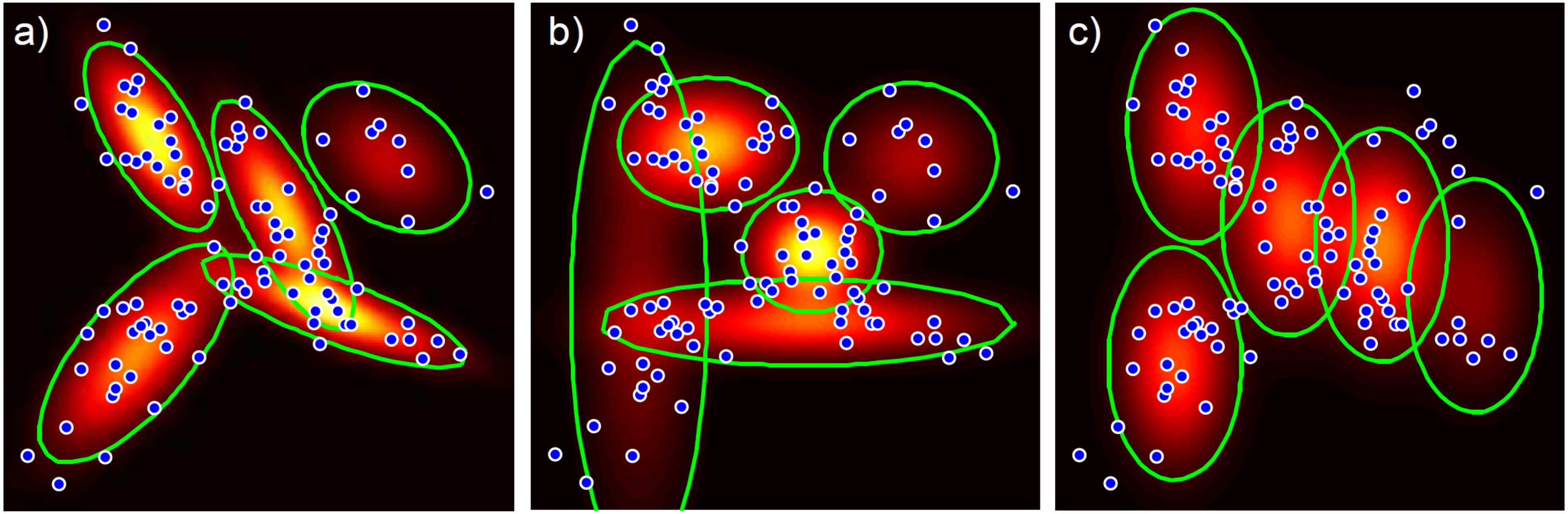

M-Step

Source: S. Prince




Source: S. Prince

# Different flavours...




Full covariance

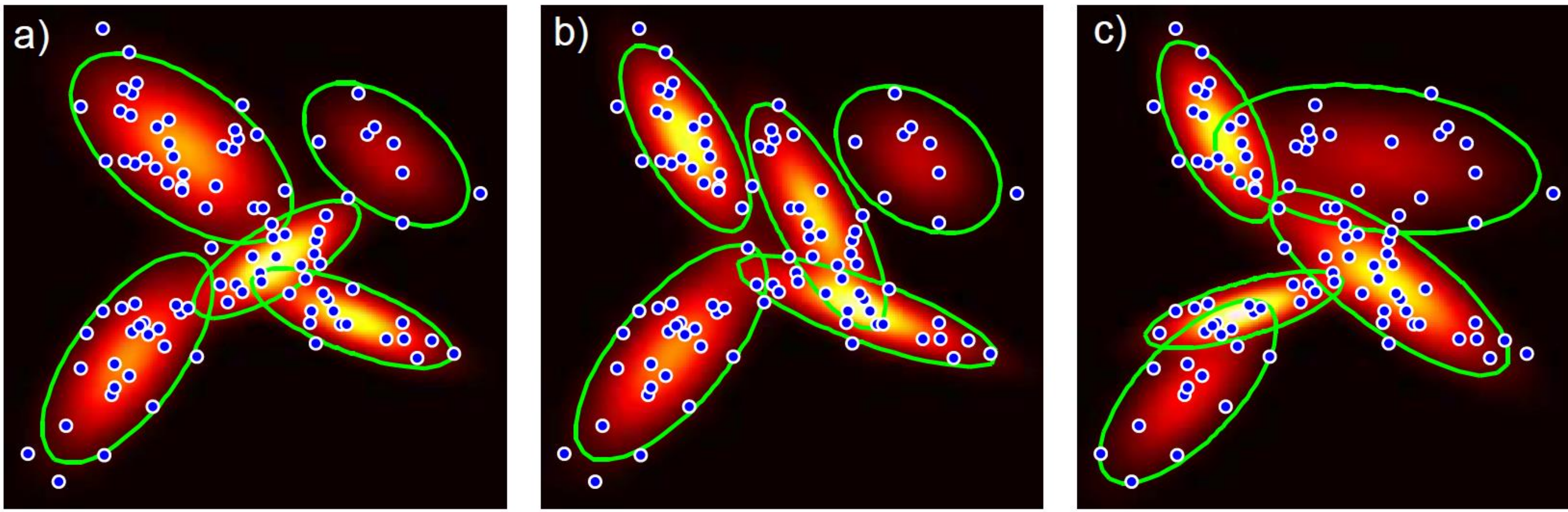
Diagonal covariance

Same covariance

Source: S. Prince

# Local Minima

Start from three random positions




L = 98.76 L = 96.97 L = 94.35

Source: S. Prince

# Expectation Maximization in General

UNIVERSITÄT BONN

Problem: Optimize cost functions of the form

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\arg\max} \sum_{i=1}^{I} \log \left[ \sum_{h} Pr(\mathbf{x}_i, h_i) \right]$$ ← Discrete case

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\arg\max} \sum_{i=1}^{I} \log \left[ \int Pr(\mathbf{x}_i, \mathbf{h}_i) d\mathbf{h}_i \right]$$ ← Continuous case

Solution: Expectation Maximization (EM) algorithm

(Dempster, Laird and Rubin 1977)

Key idea: Define lower bound on log-likelihood and increase at each iteration

Source: S. Prince

# E-Step & M-Step

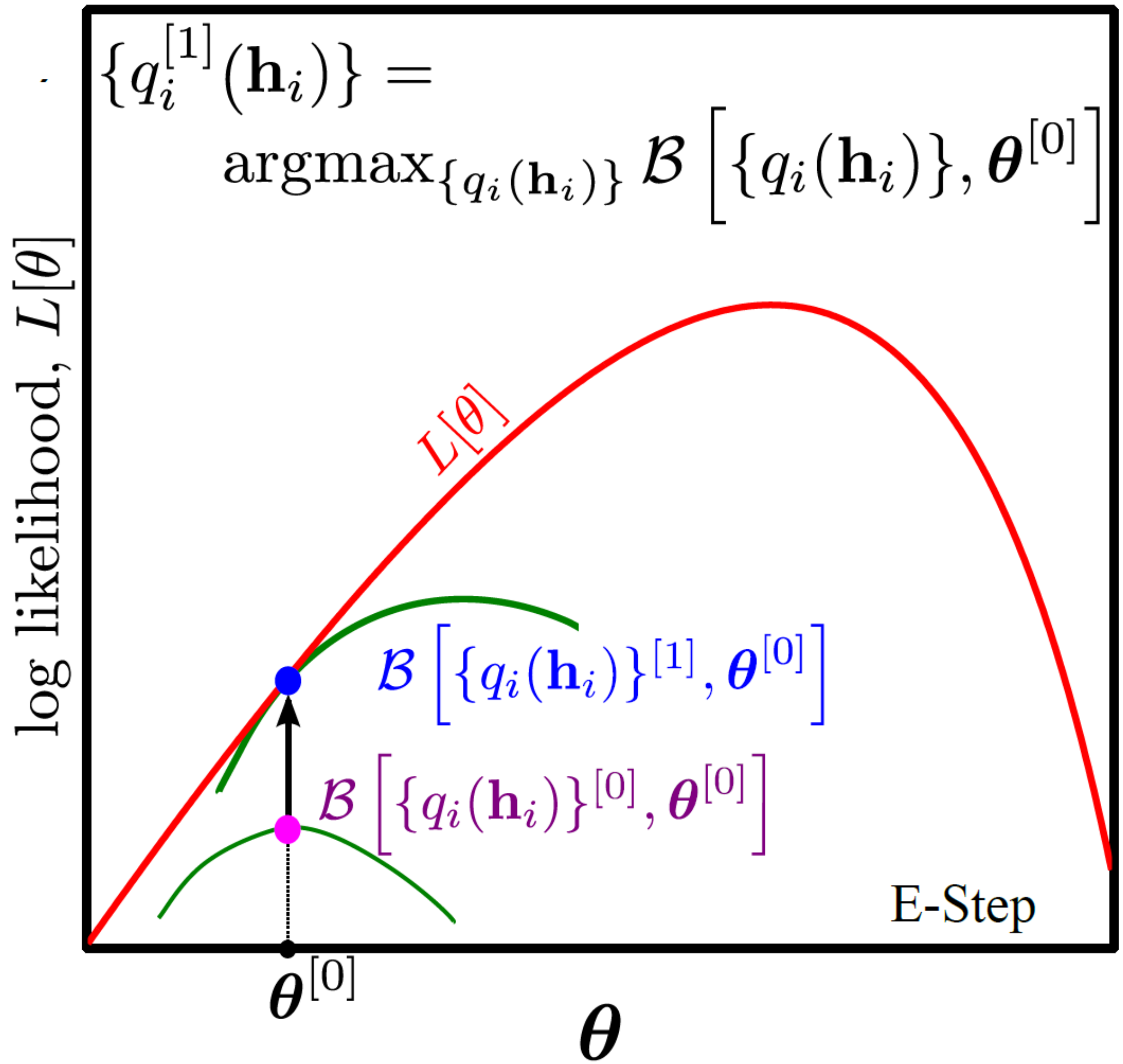
E-Step – Maximize bound w.r.t. distributions $\{q_i(\mathbf{h}_i)\}$

$$q_i^{[t]}[\mathbf{h}_i] = \underset{q_i[\mathbf{h}_i]}{\operatorname{argmax}} \left[ \mathcal{B}[\{q_i(\mathbf{h}_i)\}, \theta^{[t-1]}] \right]$$

M-Step – Maximize bound w.r.t. parameters $\theta$

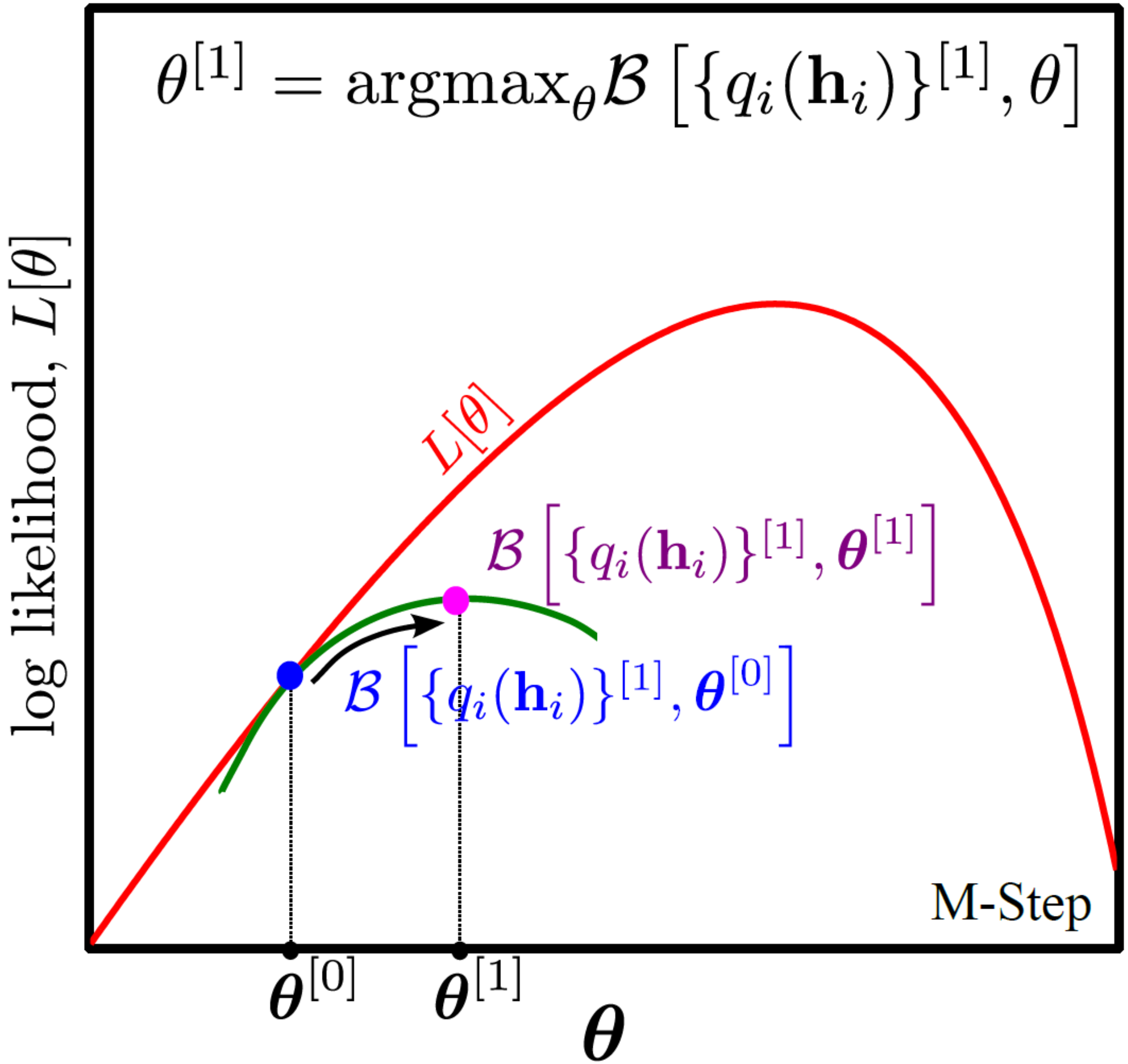
$$\boldsymbol{\theta}^{[t]} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[ \mathcal{B}[\{q_i^{[t]}(\mathbf{h}_i)\}, \boldsymbol{\theta}] \right]$$

Source: S. Prince




**E-Step:** Update $\{q_i[\mathbf{h}_i]\}$ so that bound equals log likelihood for this

Source: S. Prince




**M-Step:** Update θ to maximum

Source: S. Prince

# Expectation Maximization

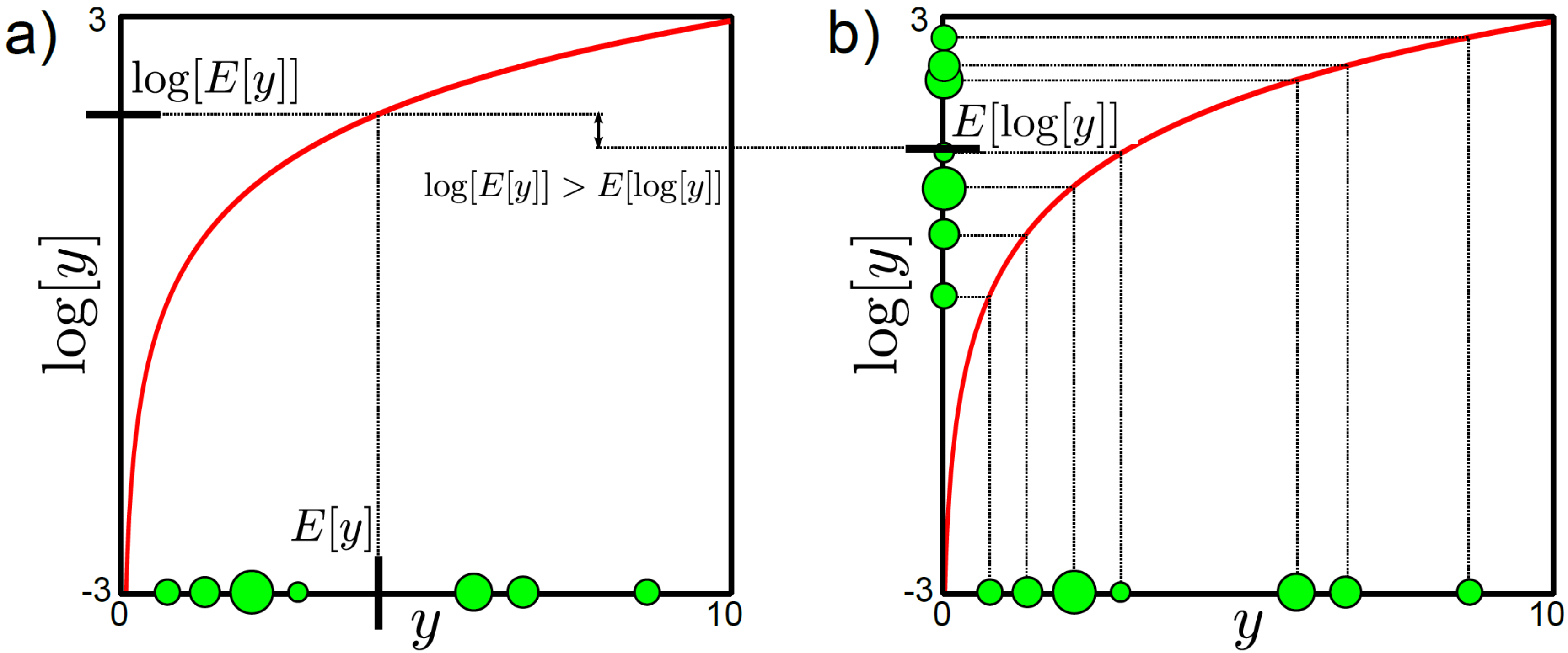
UNIVERSITÄT BONN

Defines a lower bound on log likelihood and increases bound iteratively

$$
\begin{aligned}
\mathcal{B}[\{q_i(\mathbf{h}_i)\}, \boldsymbol{\theta}] &= \sum_{i=1}^{I} \int q_i(\mathbf{h}_i) \log \left[ \frac{Pr(\mathbf{x}_i, \mathbf{h}_i | \boldsymbol{\theta})}{q_i(\mathbf{h}_i)} \right] d\mathbf{h}_i \\
&\leq \sum_{i=1}^{I} \log \left[ \int q_i(\mathbf{h}_i) \frac{Pr(\mathbf{x}_i, \mathbf{h}_i | \boldsymbol{\theta})}{q_i(\mathbf{h}_i)} \, d\mathbf{h}_i \right] \\
&= \sum_{i=1}^{I} \log \left[ \int Pr(\mathbf{x}_i, \mathbf{h}_i | \boldsymbol{\theta}) \, d\mathbf{h}_i \right],
\end{aligned}
$$

Source: S. Prince

# Jensen's Inequality



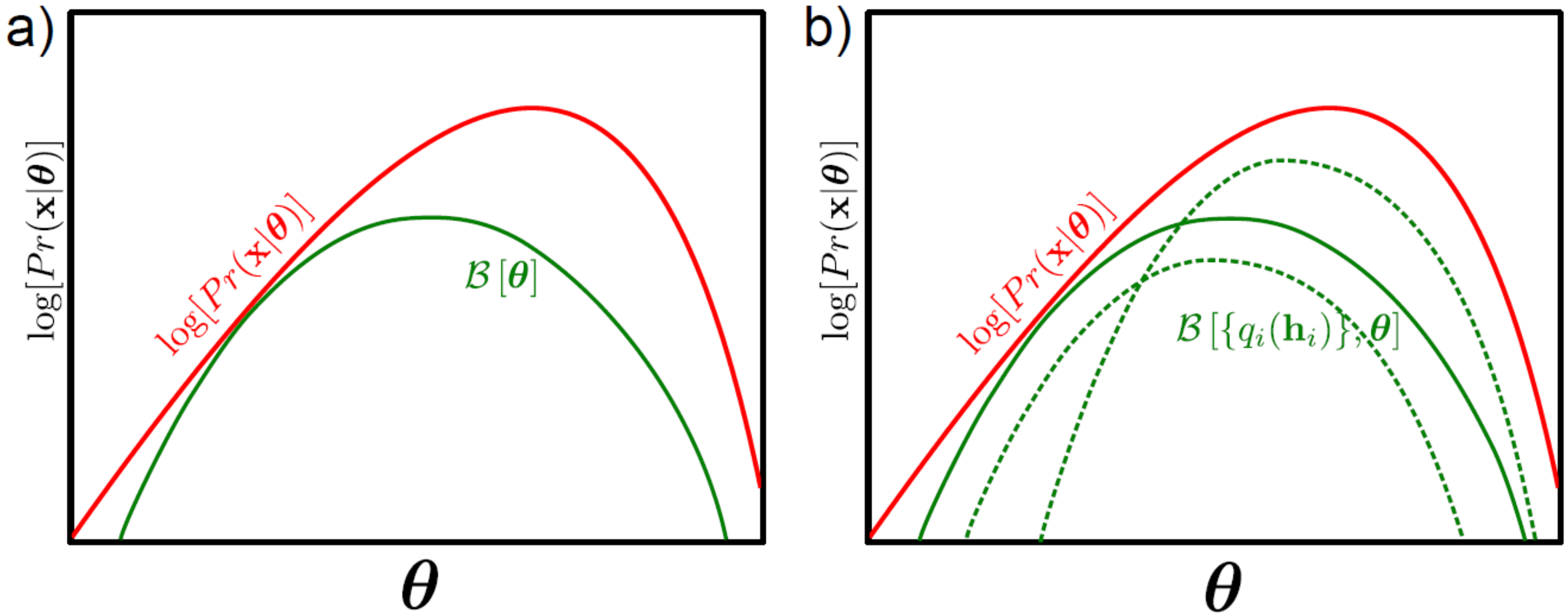

$$E\left[\log[y]\right] \leq \log\left[E[y]\right]$$ or…
$$\sum_y q(y)\log[y] \leq \log\left[\sum_y q(y)y\right]$$

similarly $$\int Pr(y)\log[y]dy \leq \log\left[\int yPr(y)dy\right]$$

Source: S. Prince

# Lower bound






$$\mathcal{B}[\{q_i(\mathbf{h}_i)\},\boldsymbol{\theta}] = \sum_{i=1}^{I}\int q_i(\mathbf{h}_i)\log\left[\frac{Pr(\mathbf{x}_i,\mathbf{h}_i|\boldsymbol{\theta})}{q_i(\mathbf{h}_i)}\right]d\mathbf{h}_i$$

Lower bound is a *function* of parameters θ and a *set* of probability distributions $\{q_i(\mathbf{h}_i)\}$

Source: S. Prince

# E-Step & M-Step

E-Step – Maximize bound w.r.t. distributions $\{q_i(\mathbf{h}_i)\}$

$$q_i^{[t]}[\mathbf{h}_i] = \underset{q_i[\mathbf{h}_i]}{\operatorname{argmax}} \left[ \mathcal{B}[\{q_i(\mathbf{h}_i)\}, \theta^{[t-1]}] \right]$$

M-Step – Maximize bound w.r.t. parameters $\theta$

$$\boldsymbol{\theta}^{[t]} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[ \mathcal{B}[\{q_i^{[t]}(\mathbf{h}_i)\}, \boldsymbol{\theta}] \right]$$

Source: S. Prince

# E-Step & M-Step

E-Step – Maximize bound w.r.t. distributions $q_i(\mathbf{h}_i)$

$$\hat{q}_i(\mathbf{h}_i) = Pr(\mathbf{h}_i|\mathbf{x}_i, \boldsymbol{\theta}^{[t]}) = \frac{Pr(\mathbf{x}_i|\mathbf{h}_i, \boldsymbol{\theta}^{[t]})Pr(\mathbf{h}_i|\boldsymbol{\theta}^{[t]})}{Pr(\mathbf{x}_i)}$$

M-Step – Maximize bound w.r.t. parameters θ

$$\hat{\boldsymbol{\theta}}^{[t+1]} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[ \sum_{i=1}^{I} \int \hat{q}_i(\mathbf{h}_i) \log \left[ Pr(\mathbf{x}_i, \mathbf{h}_i|\boldsymbol{\theta}) \right] \, d\mathbf{h}_i \right]$$

Source: S. Prince

# E-Step

E-Step – Maximize bound w.r.t. distributions $\{q_i(\mathbf{h}_i)\}$

$$q_i^{[t]}[\mathbf{h}_i] = \underset{q_i[\mathbf{h}_i]}{\operatorname{argmax}} \left[\mathcal{B}[\{q_i(\mathbf{h}_i)\}, \theta^{[t-1]}]\right]$$

Analytical solution:

$$\hat{q}_i(\mathbf{h}_i) = Pr(\mathbf{h}_i|\mathbf{x}_i, \boldsymbol{\theta}^{[t]}) = \frac{Pr(\mathbf{x}_i|\mathbf{h}_i, \boldsymbol{\theta}^{[t]})Pr(\mathbf{h}_i|\boldsymbol{\theta}^{[t]})}{Pr(\mathbf{x}_i)}$$

Source: S. Prince

# E-Step – Optimize bound w.r.t $\{q_i(h_i)\}$

$$\mathcal{B}[\{q_i(\mathbf{h}_i)\},\boldsymbol{\theta}] = \sum_{i=1}^{I}\int q_i(\mathbf{h}_i)\log\left[\frac{Pr(\mathbf{x}_i,\mathbf{h}_i|\boldsymbol{\theta})}{q_i(\mathbf{h}_i)}\right]\,d\mathbf{h}_i$$

$$= \sum_{i=1}^{I}\int q_i(\mathbf{h}_i)\log\left[\frac{Pr(\mathbf{h}_i|\mathbf{x}_i,\boldsymbol{\theta})Pr(\mathbf{x}_i|\boldsymbol{\theta})}{q_i(\mathbf{h}_i)}\right]\,d\mathbf{h}_i$$

$$= \sum_{i=1}^{I}\int q_i(\mathbf{h}_i)\log\left[Pr(\mathbf{x}_i|\boldsymbol{\theta})\right]\,d\mathbf{h}_i - \sum_{i=1}^{I}\int q_i(\mathbf{h}_i)\log\left[\frac{q_i(\mathbf{h}_i)}{Pr(\mathbf{h}_i|\mathbf{x}_i,\boldsymbol{\theta})}\right]\,d\mathbf{h}_i$$

$$= \sum_{i=1}^{I}\log\left[Pr(\mathbf{x}_i|\boldsymbol{\theta})\right] - \sum_{i=1}^{I}\int q_i(\mathbf{h}_i)\log\left[\frac{q_i(\mathbf{h}_i)}{Pr(\mathbf{h}_i|\mathbf{x}_i,\boldsymbol{\theta})}\right]\,d\mathbf{h}_i$$

Constant w.r.t. q(h)

Only this term matters

Source: S. Prince

# E-Step

UNIVERSITÄT BONN

$$\hat{q}_i(\mathbf{h}_i) \quad = \quad \underset{q_i(\mathbf{h}_i)}{\operatorname{argmax}} \left[ - \int q_i(\mathbf{h}_i) \log \left[ \frac{q_i(\mathbf{h}_i)}{Pr(\mathbf{h}_i | \mathbf{x}_i, \boldsymbol{\theta})} \right] d\mathbf{h}_i \right]$$

Kullback Leibler divergence – distance between probability distributions. We are maximizing the negative distance (i.e. Minimizing distance)

$$\hat{q}_i(\mathbf{h}_i) = Pr(\mathbf{h}_i | \mathbf{x}_i, \boldsymbol{\theta})$$

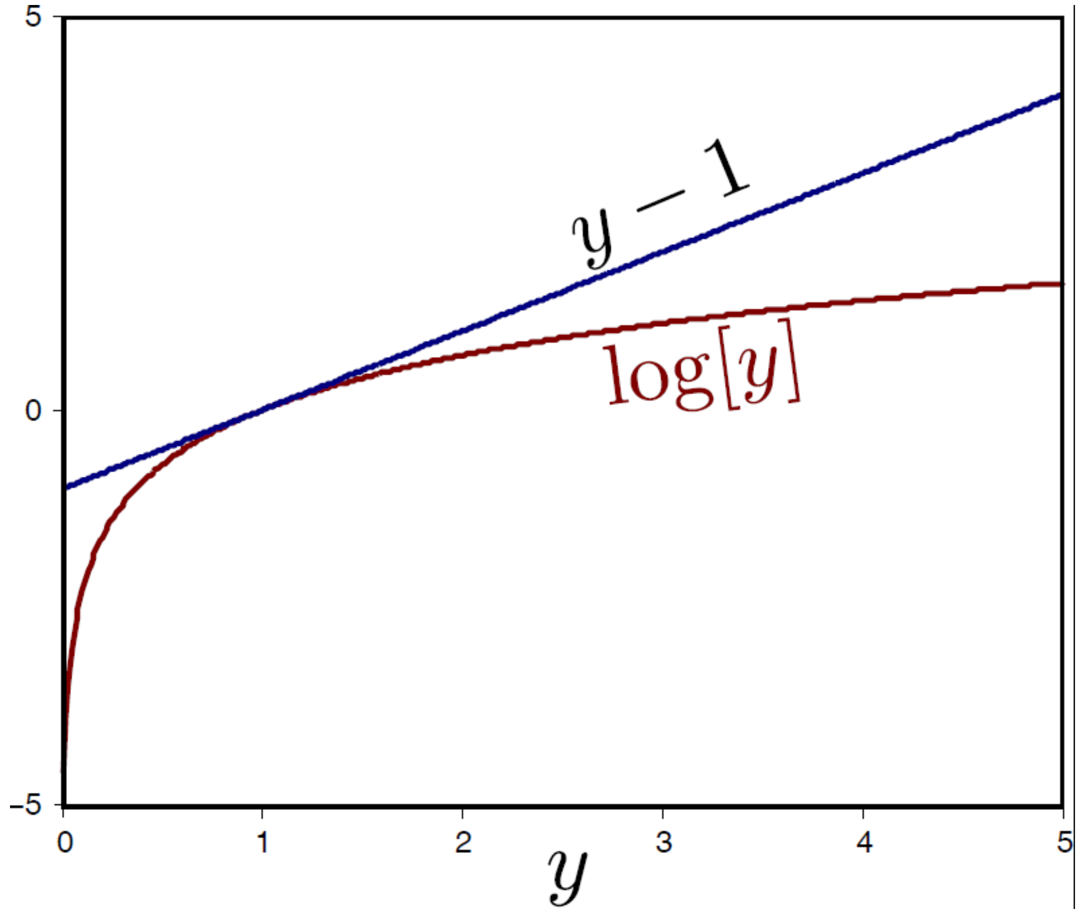
Source: S. Prince

# E-Step

$$
\begin{aligned}
\hat{q}_i(\mathbf{h}_i) &= \underset{q_i(\mathbf{h}_i)}{\operatorname{argmax}} \left[ -\int q_i(\mathbf{h}_i) \log \left[ \frac{q_i(\mathbf{h}_i)}{Pr(\mathbf{h}_i|\mathbf{x}_i, \boldsymbol{\theta})} \right] d\mathbf{h}_i \right] \\
&= \underset{q_i(\mathbf{h}_i)}{\operatorname{argmax}} \left[ \int q_i(\mathbf{h}_i) \log \left[ \frac{Pr(\mathbf{h}_i|\mathbf{x}_i, \boldsymbol{\theta})}{q_i(\mathbf{h}_i)} \right] d\mathbf{h}_i \right] \\
&= \underset{q_i(\mathbf{h}_i)}{\operatorname{argmin}} \left[ -\int q_i(\mathbf{h}_i) \log \left[ \frac{Pr(\mathbf{h}_i|\mathbf{x}_i, \boldsymbol{\theta})}{q_i(\mathbf{h}_i)} \right] d\mathbf{h}_i \right]
\end{aligned}
$$

Source: S. Prince

# Use this relation




$\log[y] \leq y-1$

Source: S. Prince

# Kullback Leibler Divergence

$$\begin{aligned}
\int q_i(\mathbf{h}_i) \log\left[\frac{Pr(\mathbf{h}_i|\mathbf{x}_i,\boldsymbol{\theta})}{q_i(\mathbf{h}_i)}\right] d\mathbf{h}_i &\leq \int q_i(\mathbf{h}_i)\left(\frac{Pr(\mathbf{h}_i|\mathbf{x}_i,\boldsymbol{\theta})}{q_i(\mathbf{h}_i)} - 1\right) d\mathbf{h}_i \\
&= \int Pr(\mathbf{h}_i|\mathbf{x}_i,\boldsymbol{\theta}) - q_i(\mathbf{h}_i)\, d\mathbf{h}_i \\
&= 1 - 1 = 0,
\end{aligned}$$

So the cost function must be positive

$$\hat{q}_i(\mathbf{h}_i) = \underset{q_i(\mathbf{h}_i)}{\operatorname{argmin}}\left[-\int q_i(\mathbf{h}_i) \log\left[\frac{Pr(\mathbf{h}_i|\mathbf{x}_i,\boldsymbol{\theta})}{q_i(\mathbf{h}_i)}\right] d\mathbf{h}_i\right]$$

In other words, the best we can do is choose $q_i(\mathbf{h}_i)$ so that this is zero

Source: S. Prince

# E-Step

So the cost function must be positive

$$\hat{q}_i(\mathbf{h}_i) = \underset{q_i(\mathbf{h}_i)}{\operatorname{argmin}} \left[ -\int q_i(\mathbf{h}_i) \log \left[ \frac{Pr(\mathbf{h}_i|\mathbf{x}_i, \boldsymbol{\theta})}{q_i(\mathbf{h}_i)} \right] d\mathbf{h}_i \right]$$

The best we can do is choose $q_i(\mathbf{h}_i)$ so that this is zero.

How can we do this? Easy – choose posterior Pr(**h**|x)

$$\int q_i(\mathbf{h}_i) \log \left[ \frac{Pr(\mathbf{h}_i|\mathbf{x}_i, \boldsymbol{\theta})}{q_i(\mathbf{h}_i)} \right] d\mathbf{h}_i = \int Pr(\mathbf{h}_i|\mathbf{x}_i, \boldsymbol{\theta}) \log \left[ \frac{Pr(\mathbf{h}_i|\mathbf{x}_i, \boldsymbol{\theta})}{Pr(\mathbf{h}_i|\mathbf{x}_i, \boldsymbol{\theta})} \right] d\mathbf{h}_i$$

$$= \int Pr(\mathbf{h}_i|\mathbf{x}_i, \boldsymbol{\theta}) \log [1] \; d\mathbf{h}_i = 0.$$

Source: S. Prince

# E-Step

UNIVERSITÄT BONN

E-Step – Maximize bound w.r.t. distributions $\{q_i(\mathbf{h}_i)\}$

$$q_i^{[t]}[\mathbf{h}_i] = \underset{q_i[\mathbf{h}_i]}{\operatorname{argmax}} \left[ \mathcal{B}[\{q_i(\mathbf{h}_i)\}, \theta^{[t-1]}] \right]$$

Analytical solution:

$$\hat{q}_i(\mathbf{h}_i) = Pr(\mathbf{h}_i | \mathbf{x}_i, \boldsymbol{\theta}^{[t]}) = \frac{Pr(\mathbf{x}_i | \mathbf{h}_i, \boldsymbol{\theta}^{[t]}) Pr(\mathbf{h}_i | \boldsymbol{\theta}^{[t]})}{Pr(\mathbf{x}_i)}$$

Source: S. Prince

# M-Step

M-Step – Maximize bound w.r.t. parameters θ

$$\boldsymbol{\theta}^{[t]} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[ \mathcal{B}[\{q_i^{[t]}(\mathbf{h}_i)\}, \boldsymbol{\theta}] \right]$$

Simplifies to:

$$\hat{\boldsymbol{\theta}}^{[t+1]} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[ \sum_{i=1}^{I} \int \hat{q}_i(\mathbf{h}_i) \log \left[ Pr(\mathbf{x}_i, \mathbf{h}_i | \boldsymbol{\theta}) \right] \, d\mathbf{h}_i \right]$$

Source: S. Prince

# M-Step – Optimize bound w.r.t. θ

$$
\begin{aligned}
\boldsymbol{\theta}^{[t]} &= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[\mathcal{B}[\{q_i^{[t]}(\mathbf{h}_i)\}, \boldsymbol{\theta}]\right] \\
&= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[\sum_{i=1}^{I} \int q_i^{[t]}(\mathbf{h}_i) \log\left[\frac{Pr(\mathbf{x}_i, \mathbf{h}_i|\boldsymbol{\theta})}{q_i^{[t]}(\mathbf{h}_i)}\right] d\mathbf{h}_i\right] \\
&= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[\sum_{i=1}^{I} \int q_i^{[t]}(\mathbf{h}_i) \log\left[Pr(\mathbf{x}_i, \mathbf{h}_i|\boldsymbol{\theta})\right] - q_i^{[t]}(\mathbf{h}_i) \log\left[q_i^{[t]}(\mathbf{h}_i)\right] d\mathbf{h}_i\right] \\
&= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[\sum_{i=1}^{I} \int q_i^{[t]}(\mathbf{h}_i) \log\left[Pr(\mathbf{x}_i, \mathbf{h}_i|\boldsymbol{\theta})\right] d\mathbf{h}_i\right]
\end{aligned}
$$

In the M-Step we optimize expected joint log likelihood with respect to parameters θ (Expectation w.r.t distribution from E-Step)

Source: S. Prince

# E-Step & M-Step

E-Step – Maximize bound w.r.t. distributions $q_i(\mathbf{h}_i)$

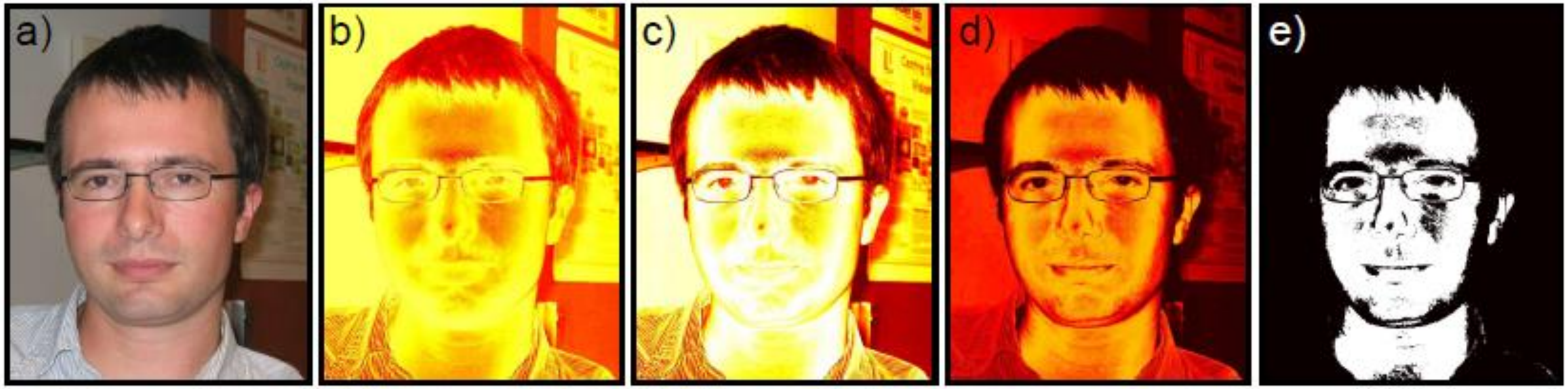
$$\hat{q}_i(\mathbf{h}_i) = Pr(\mathbf{h}_i|\mathbf{x}_i, \boldsymbol{\theta}^{[t]}) = \frac{Pr(\mathbf{x}_i|\mathbf{h}_i, \boldsymbol{\theta}^{[t]})Pr(\mathbf{h}_i)}{Pr(\mathbf{x}_i)}$$

M-Step – Maximize bound w.r.t. parameters $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}}^{[t+1]} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \left[ \sum_{i=1}^{I} \int \hat{q}_i(\mathbf{h}_i) \log \left[ Pr(\mathbf{x}_i, \mathbf{h}_i|\boldsymbol{\theta}) \right] \, d\mathbf{h}_i \right]$$

Source: S. Prince

# Skin detection



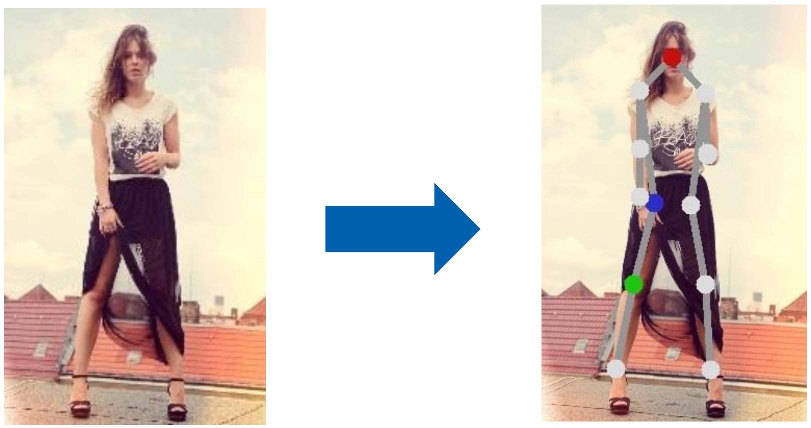

Log likelihood w=0 | Log likelihood w=1 | Posterior w=1 | Thresholded

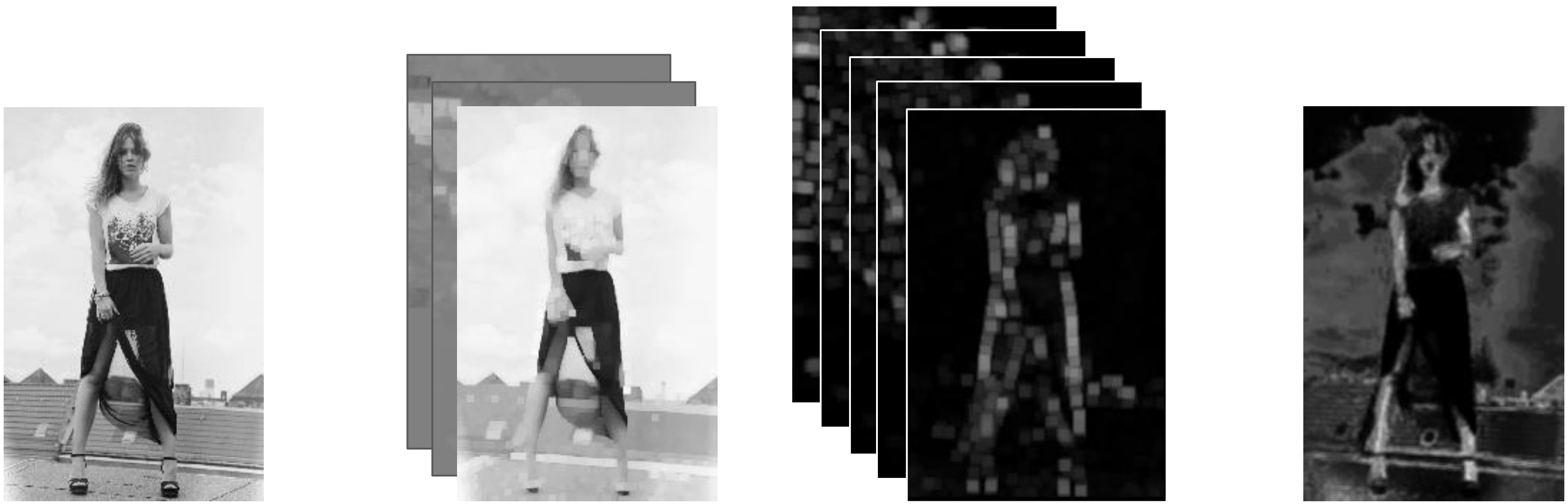
GMM for skin (w=1) and non-skin (w=0):

$$Pr(x|w=k) = \sum_i \lambda_{k,i} \mathrm{Norm}_x[\mu_{k,i}, \Sigma_{k,i}] \qquad Pr(w) = \mathrm{Bern}_w[\lambda]$$

$$Pr(w=1|\mathbf{x}) = \frac{Pr(\mathbf{x}|w=1)Pr(w=1)}{\sum_{k=0}^{1} Pr(\mathbf{x}|w=k)Pr(w=k)}$$

# Human pose estimation





Intensity LAB HOG Skin

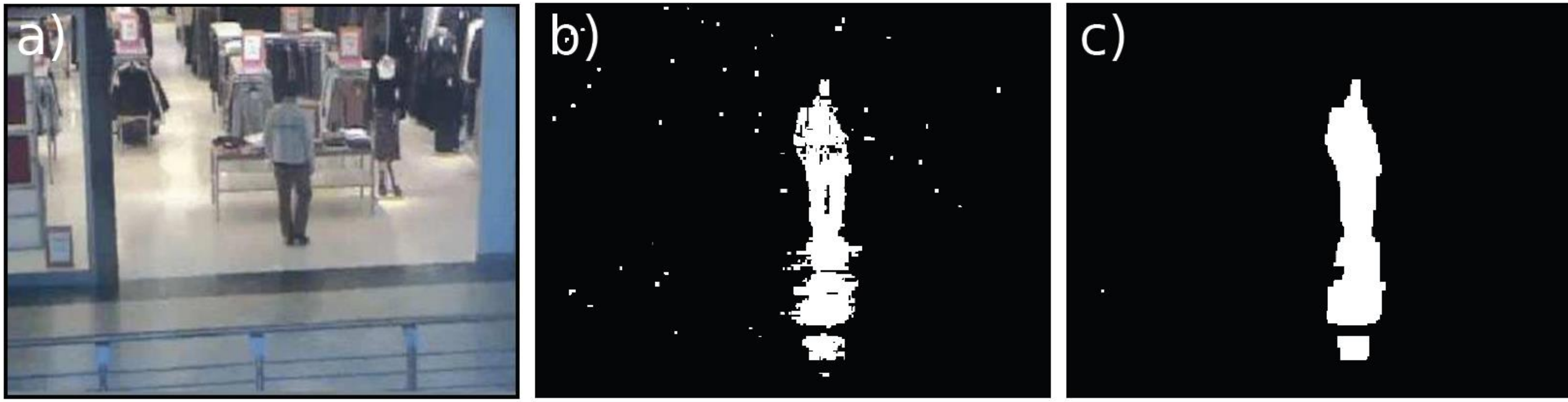
[ M. Dantone et al. **Body Parts Dependent Joint Regressors for Human Pose Estimation in Still Images.** PAMI 2014 ]

# Background subtraction


Model background distribution of each pixel by GMM:

$$Pr(x_n|w=0) = \sum_i \lambda_{n,i} \mathrm{Norm}_{x_n}[\mu_{n,i}, \Sigma_{n,i}]$$

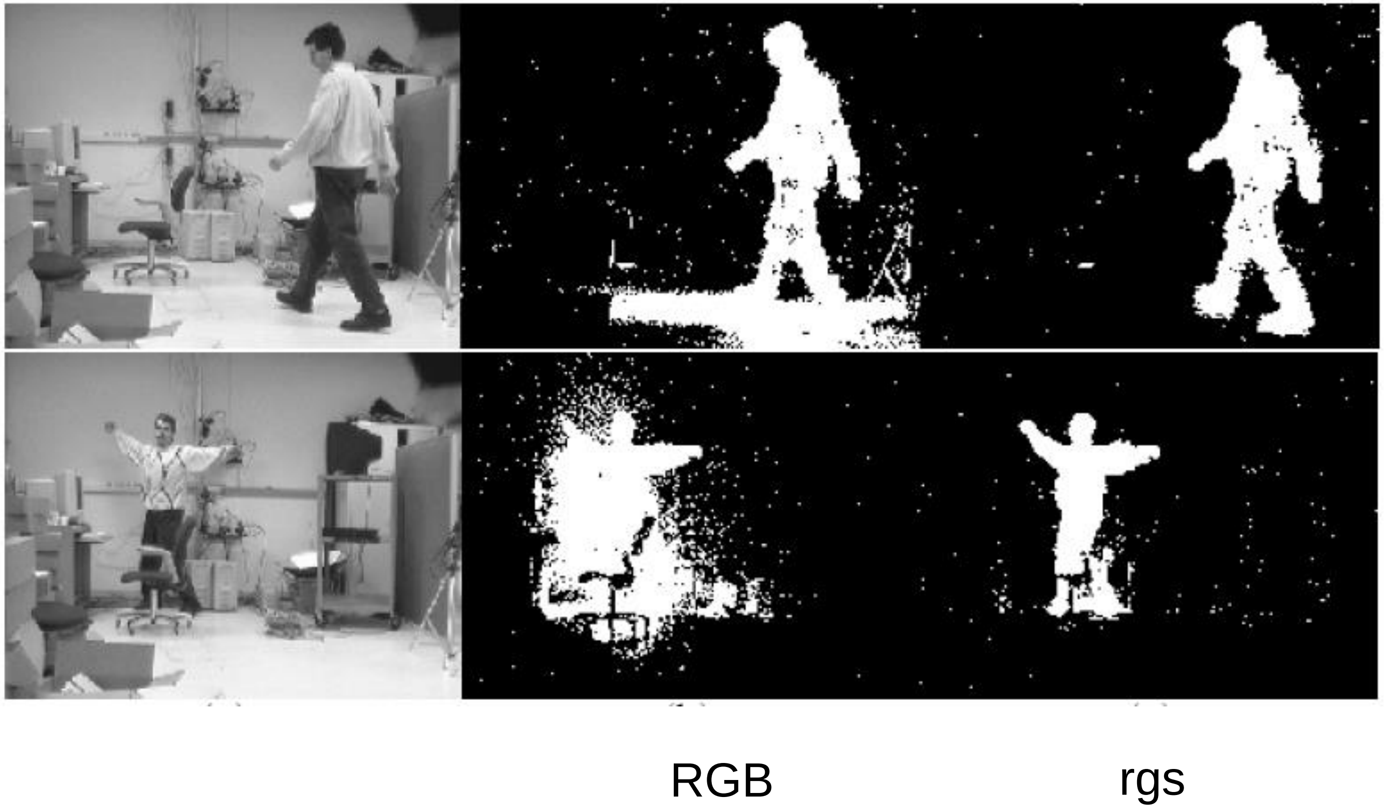
$$Pr(\mathbf{x}_n|w=1) \quad = \quad \kappa,$$




Test image | Pixel-wise classification | MRF

# Background subtraction

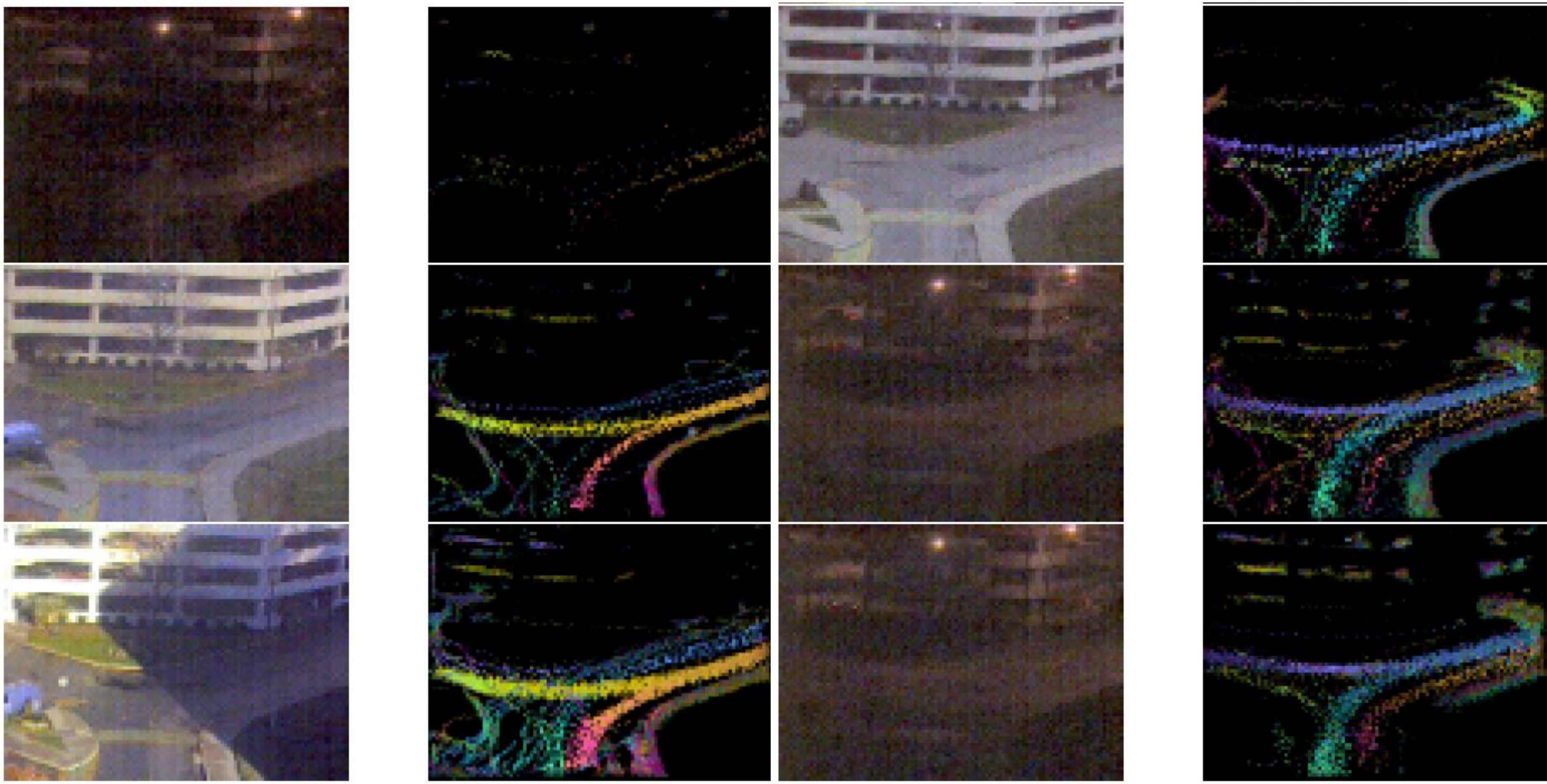
Shadows → Separate chromaticity from intensity (two thresholds)




Examples: HSV, LAB, rgs, … $r = \frac{R}{R+G+B}, g = \frac{G}{R+G+B}$ $s = R + G + B$

A. Elgammal. **Figure-ground segmentation - pixel-based.** Springer 2011

# Cars and pedestrians



1 day

# Background adaptation

MoG:

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t})$$

Update over time:

$$\omega_{k,t} = (1-\alpha)\omega_{k,t-1} + \alpha(M_{k,t}) \qquad M_{k,t} = \begin{cases} 1 & \|X_t - \mu_k\| < 2.5\sigma_k \\ 0 & \text{otherwise} \end{cases}$$

$$\mu_t = (1-\rho)\mu_{t-1} + \rho X_t$$

$$\rho = \alpha\eta(X_t|\mu_k, \sigma_k)$$

$$\sigma_t^2 = (1-\rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T(X_t - \mu_t)$$

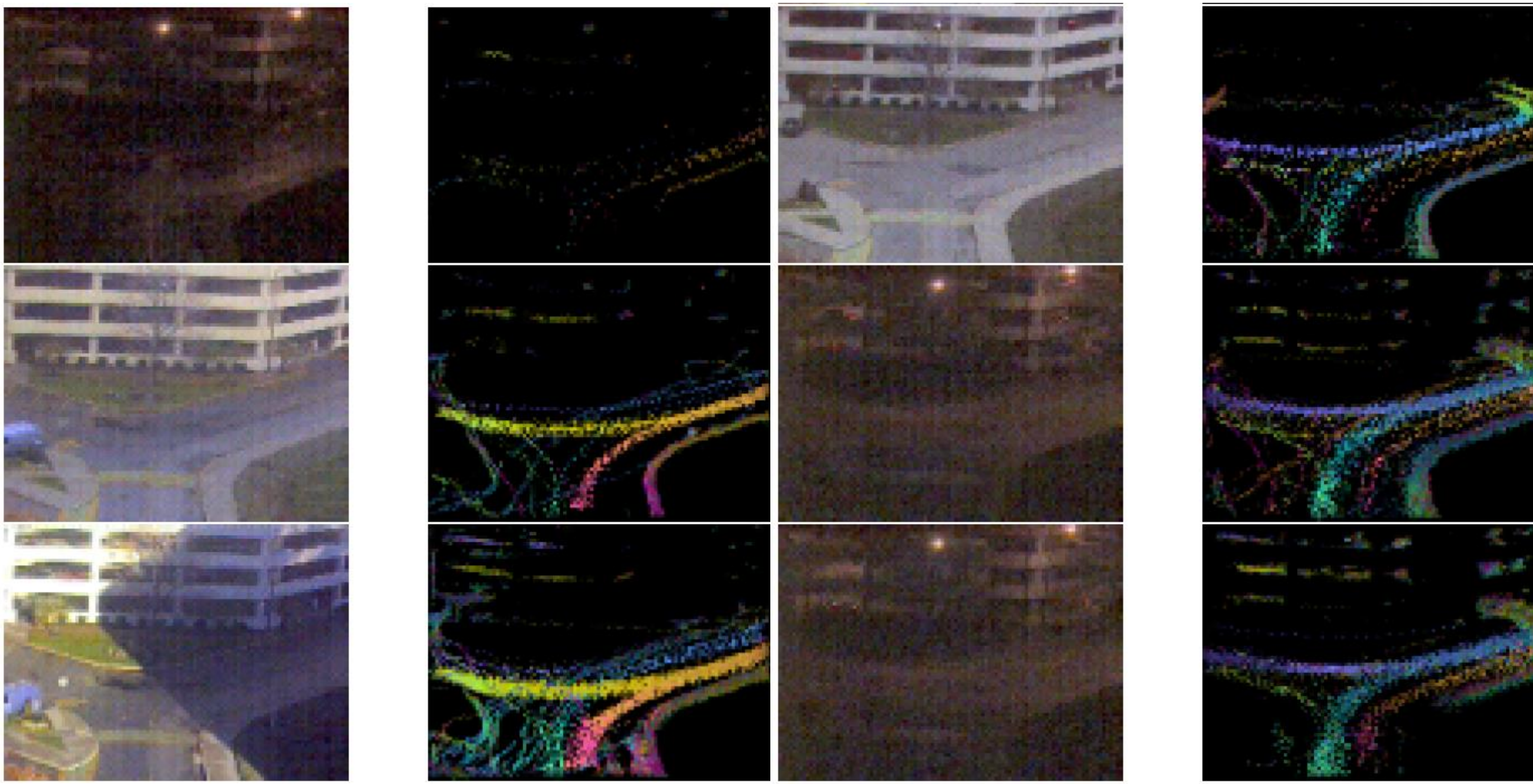
Distributions that are more likely and have less variance are usually part of the background.

Sort Gaussians decreasingly by $\omega/\sigma$ to obtain background model:

$$B = argmin_b \left( \sum_{k=1}^{b} \omega_k > T \right)$$

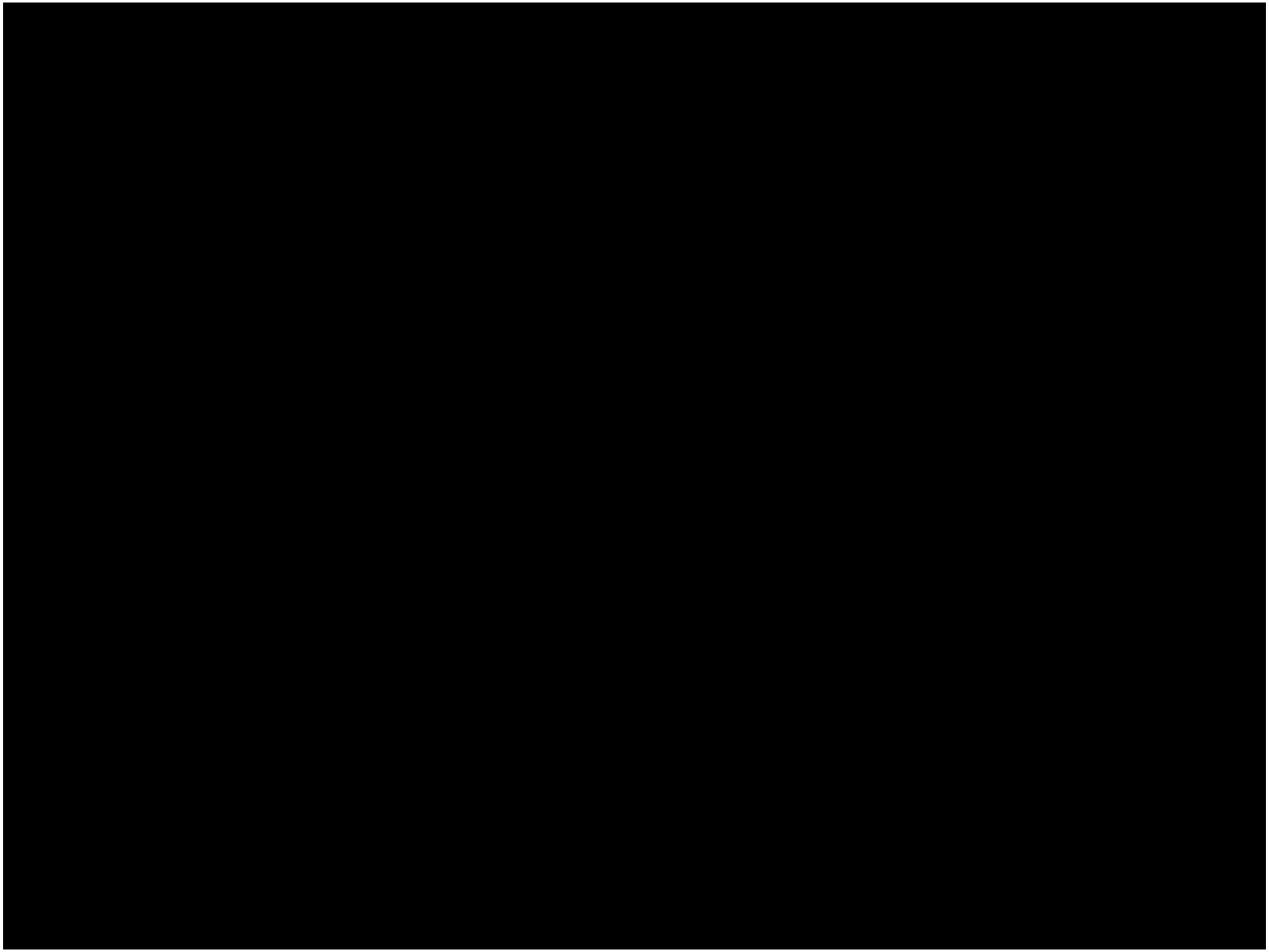
C. Stauffer and W. Grimson. **Adaptive background mixture models for real-time tracking.** CVPR 1999

# Cars and pedestrians



1 day

# Human pose tracking: 1997



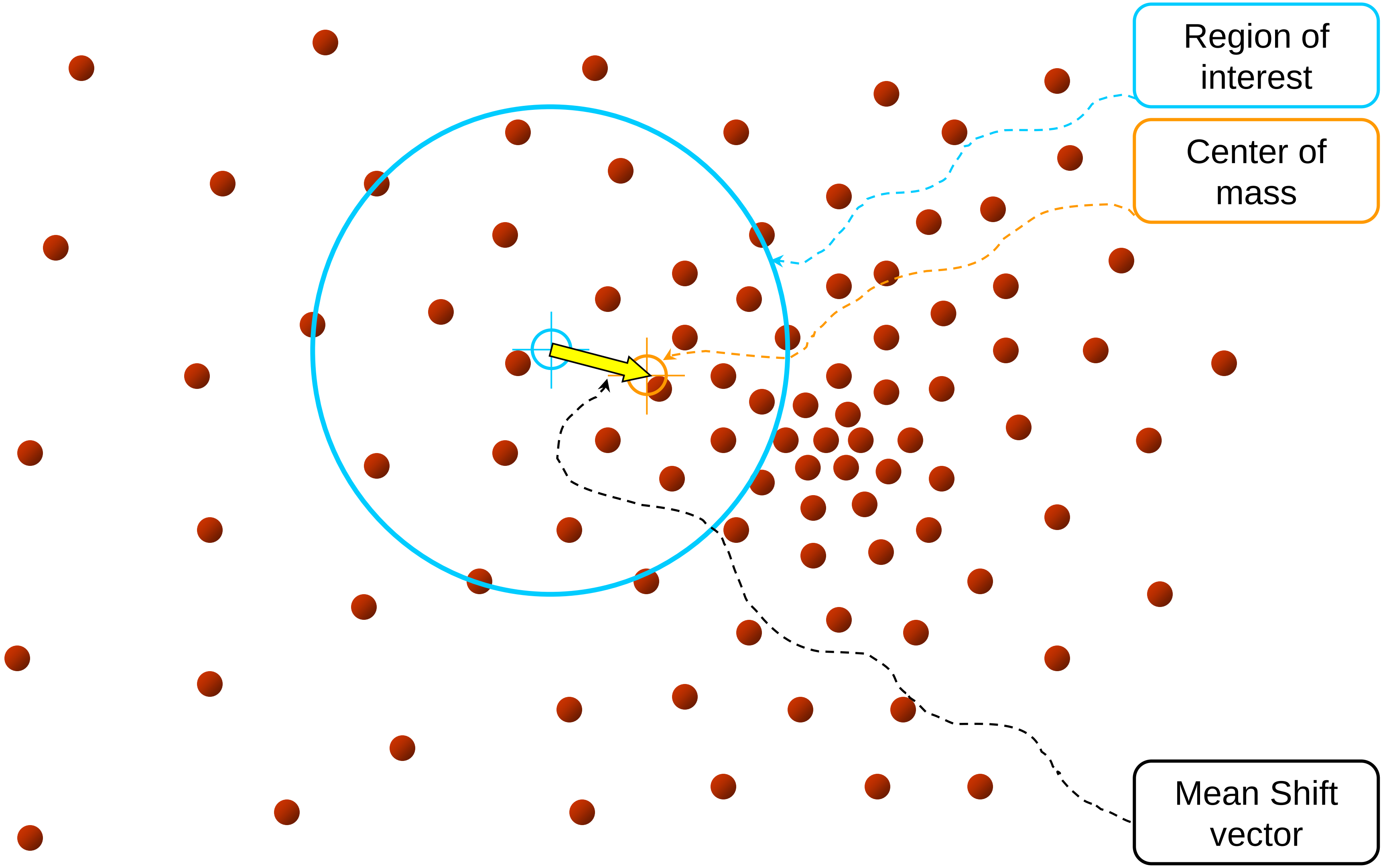
S. Wren et al. **Pfinder: Real-Time Tracking of the Human Body.** TPAMI 1997

# Mean Shift Tracking

Goal: Mark object in first frame and locate it in all frames

D. Comaniciu. **Real-Time Tracking of Non-Rigid Objects using Mean Shift.** CVPR 2000

# Recall: Mean Shift

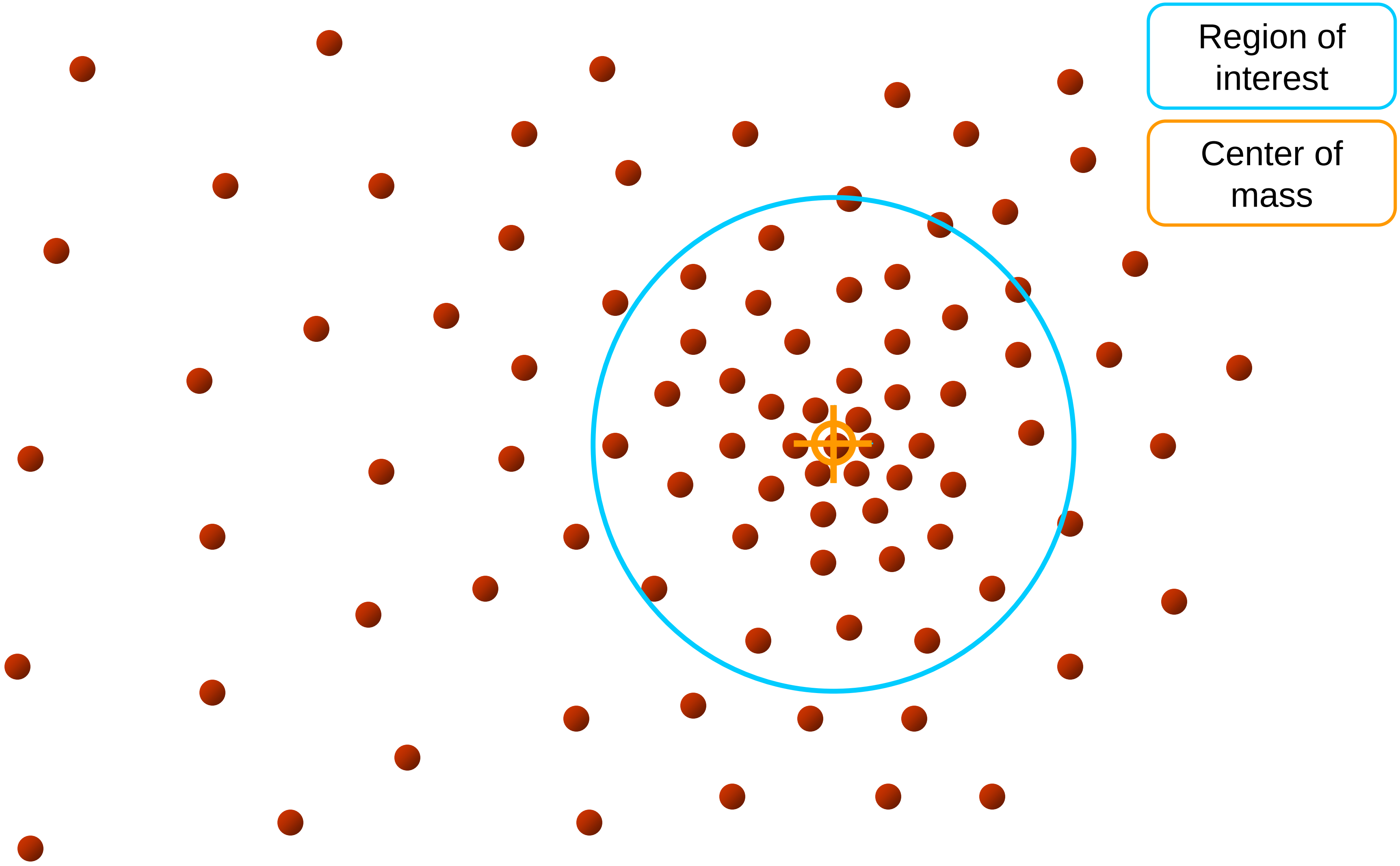



**<u>Objective</u> : Find the densest region**

Distribution of identical billiard balls

Source: Y. Ukrainitz and B. Sarel

# Recall: Mean Shift




**<u>Objective</u> : Find the densest region**

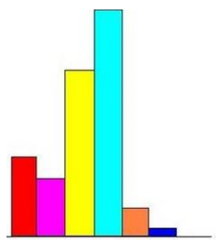
Distribution of identical billiard balls

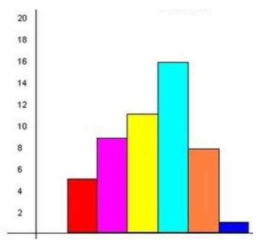
Source: Y. Ukrainitz and B. Sarel

# Template

Distance between template q (histogram of color) and target p(y) (histogram of color), given by Bhattacharyya coefficient:

q






p(y)

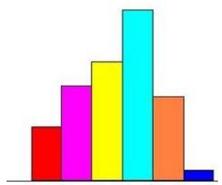
$$d(\mathbf{y}) = \sqrt{1 - \rho\left[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}\right]} \qquad \hat{\rho}(\mathbf{y}) \equiv \rho\left[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}\right] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\mathbf{y})\hat{q}_u}.$$

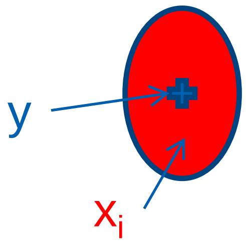
# Mean shift

Mean shift for $x_i$ data points:

$$\mathbf{y}_{j+1} = \frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\left\|\frac{\mathbf{y}_j - \mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{y}_j - \mathbf{x}_i}{h}\right\|^2\right)} \qquad g(\mathbf{x}) = -k'(\mathbf{x})$$

Look at pixel around y: $x_i$ with histogram bin $b(x_i)$ to get histogram $\sum_{u=1}^{m} \hat{p}_u = 1$






$$\hat{p}_u(\mathbf{y}) = C_h \sum_{i=1}^{n_h} k\left(\left\|\frac{\mathbf{y} - \mathbf{x}_i}{h}\right\|^2\right) \delta\left[b(\mathbf{x}_i) - u\right]$$

$$C_h = \frac{1}{\sum_{i=1}^{n_h} k\left(\left\|\frac{\mathbf{y} - \mathbf{x}_i}{h}\right\|^2\right)}$$

# Mean shift tracking

UNIVERSITÄT BONN

Maximize Bhattacharyya coefficient:

$$\hat{\rho}(\mathbf{y}) \equiv \rho\left[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}\right] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\mathbf{y})\hat{q}_u}.$$

$$f(a) + \frac{f'(a)}{1!}(x-a)$$

Taylor expansion around $p(y_0)$:

$$\rho\left[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}\right] \approx \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{\mathbf{y}}_0)\hat{q}_u} + \frac{1}{2}\sum_{u=1}^{m} \hat{p}_u(\mathbf{y})\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{\mathbf{y}}_0)}} \qquad \hat{p}_u(\hat{\mathbf{y}}_0) > 0$$

# Mean shift tracking

Maximize Bhattacharyya coefficient:

$$\hat{\rho}(\mathbf{y}) \equiv \rho\left[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}\right] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\mathbf{y})\hat{q}_u}.$$

Taylor expansion around $p(y_0)$:

$$\rho\left[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}\right] \approx \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{\mathbf{y}}_0)\hat{q}_u} + \frac{1}{2}\sum_{u=1}^{m} \hat{p}_u(\mathbf{y})\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{\mathbf{y}}_0)}} \qquad \hat{p}_u(\hat{\mathbf{y}}_0) > 0$$

Using

$$\hat{p}_u(\mathbf{y}) = C_h \sum_{i=1}^{n_h} k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h}\right\|^2\right)\delta\left[b(\mathbf{x}_i) - u\right]$$

We get:

$$\rho\left[\hat{\mathbf{p}}(\mathbf{y}), \hat{\mathbf{q}}\right] \approx \frac{1}{2}\sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{\mathbf{y}}_0)\hat{q}_u} + \frac{C_h}{2}\sum_{i=1}^{n_h} w_i k\left(\left\|\frac{\mathbf{y}-\mathbf{x}_i}{h}\right\|^2\right) \qquad w_i = \sum_{u=1}^{m} \delta\left[b(\mathbf{x}_i) - u\right]\sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{\mathbf{y}}_0)}}$$

# Mean shift tracking

Start with position from previous frame $y_0$ and template q

$$\rho\left[\hat{\mathbf{p}}(\hat{\mathbf{y}}_0), \hat{\mathbf{q}}\right] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{\mathbf{y}}_0)\hat{q}_u}$$

Compute weights:

$$w_i = \sum_{u=1}^{m} \delta\left[b(\mathbf{x}_i) - u\right] \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{\mathbf{y}}_0)}}$$

Mean shift:

$$\hat{\mathbf{y}}_1 = \frac{\sum_{i=1}^{n_h} \mathbf{x}_i w_i g\left(\left\| \frac{\hat{\mathbf{y}}_0 - \mathbf{x}_i}{h} \right\|^2\right)}{\sum_{i=1}^{n_h} w_i g\left(\left\| \frac{\hat{\mathbf{y}}_0 - \mathbf{x}_i}{h} \right\|^2\right)}$$

Update:

$$\rho\left[\hat{\mathbf{p}}(\hat{\mathbf{y}}_1), \hat{\mathbf{q}}\right] = \sum_{u=1}^{m} \sqrt{\hat{p}_u(\hat{\mathbf{y}}_1)\hat{q}_u}$$

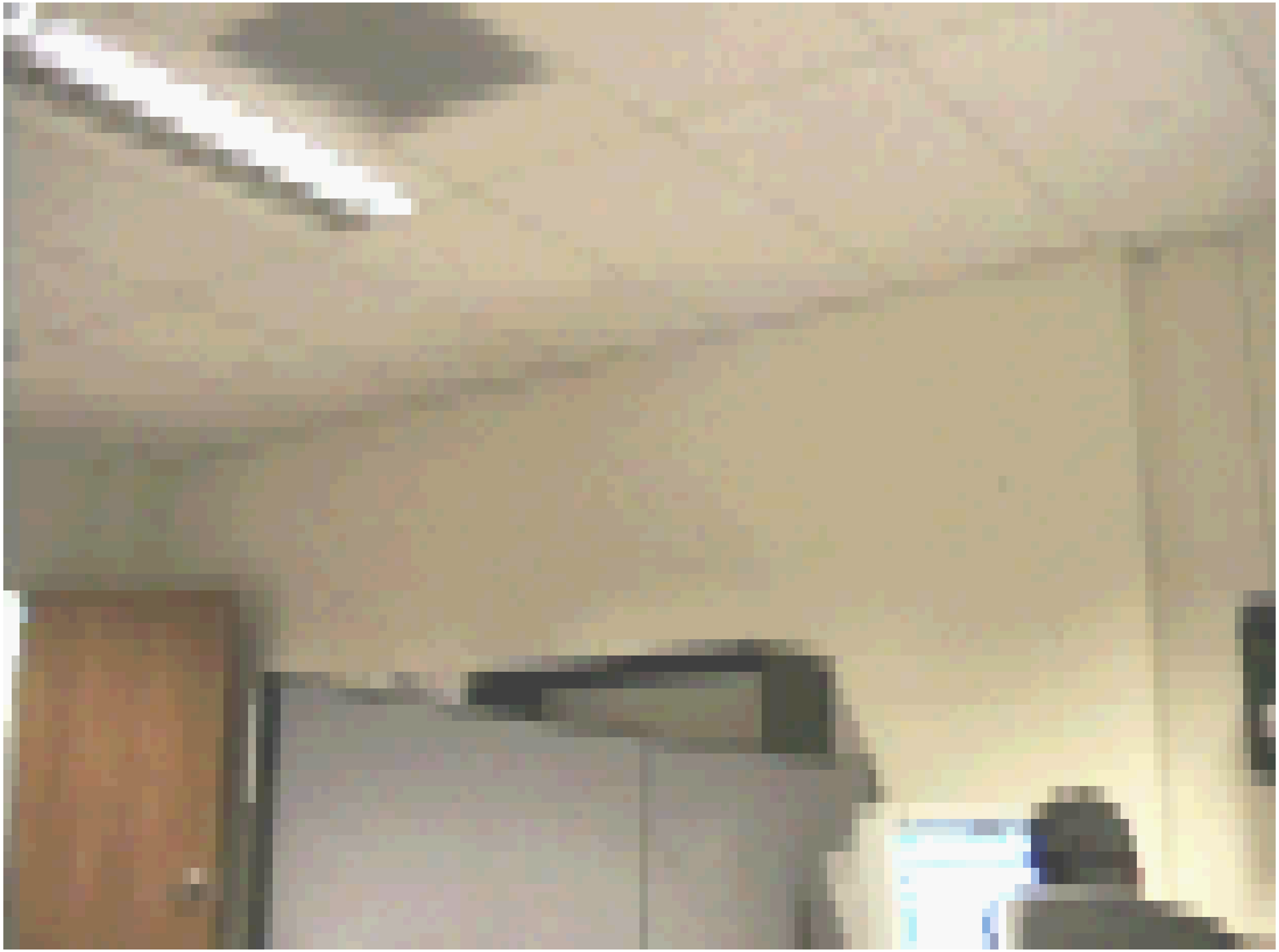
$$\begin{array}{ll} \text{While} & \rho\left[\hat{\mathbf{p}}(\hat{\mathbf{y}}_1), \hat{\mathbf{q}}\right] < \rho\left[\hat{\mathbf{p}}(\hat{\mathbf{y}}_0), \hat{\mathbf{q}}\right] \\ \text{Do} & \hat{\mathbf{y}}_1 \leftarrow \frac{1}{2}(\hat{\mathbf{y}}_0 + \hat{\mathbf{y}}_1). \end{array}$$

$$\|\hat{\mathbf{y}}_1 - \hat{\mathbf{y}}_0\| < \epsilon$$

# Mean shift tracking

# Mean shift tracking

# Mean shift tracking



# Mean shift tracking

- Simple and fast
- Initialization matters
- Get stuck in local optima
- Temporal information is ignored
- Cannot handle occlusions

UNIVERSITÄT BONN