Learning à la Bayes

—From image reconstruction to inverse quantum theory —

Part 3: Image reconstruction, prior mixtures, and hyperparameters

Jörg Lemm

Institut für Theoretische Physik
Universität Münster

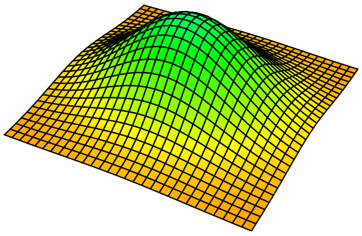
`lemm@uni-muenster.de`

`http://pauli.uni-muenster.de/~lemm`

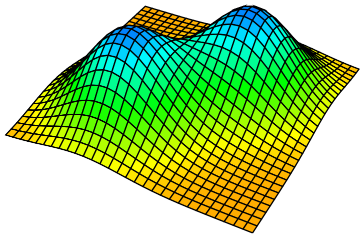
J.L., ICANN99; J.L., physics/9912005; J.L., Uhlig, Weiguny, quant-ph/0005122

Explicit modeling of complex priors with prior mixtures, hyperparameters, and hyperfields

Gaussian process

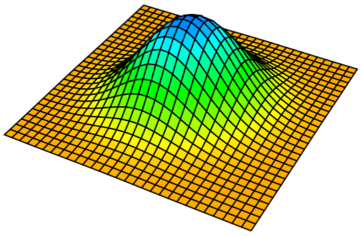


mixture of two Gaussians

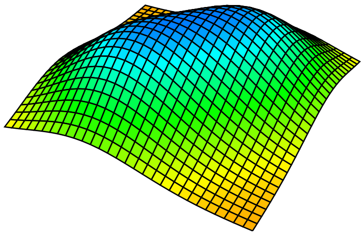


low temperature case

mixture of two Gaussians

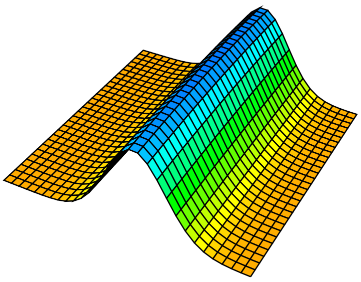


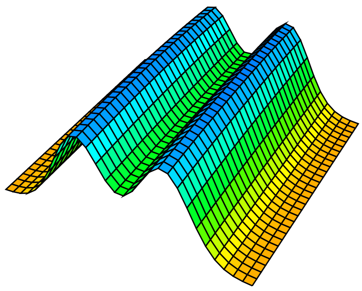
high temperature case



near „critical point“

hyperparameter θ or hyperfiels $\theta(x)$

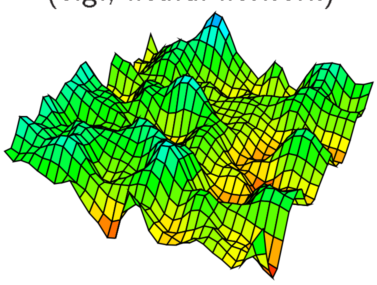




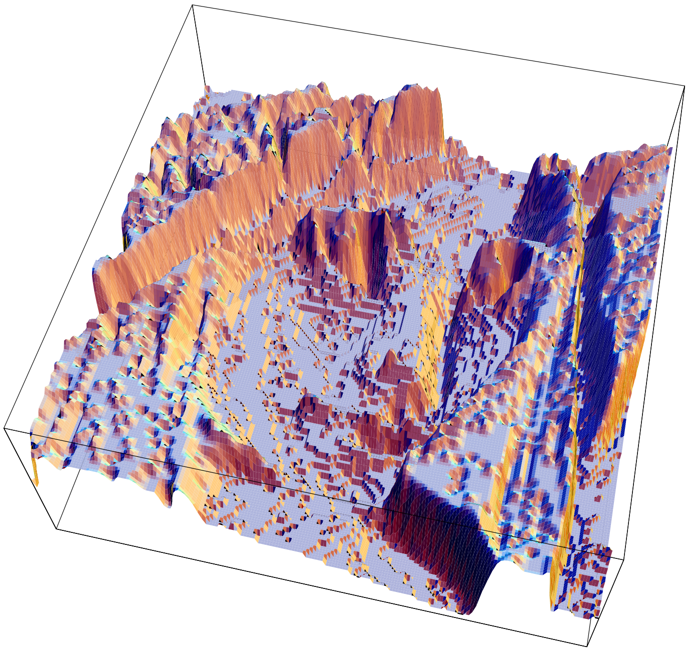
mixture with hyperparameters/–fields

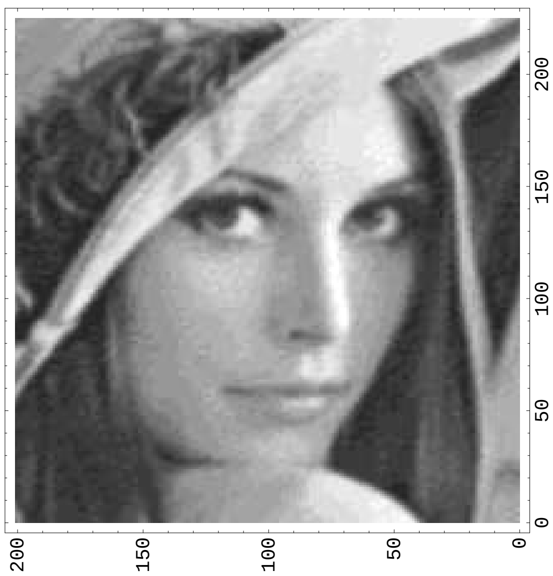
typical error surface of a parametric model

(e.g., neural network)



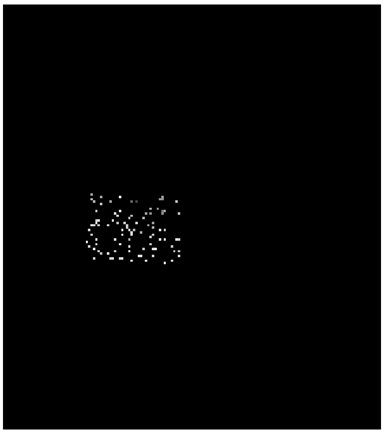
Example of a two–dimensional regression function



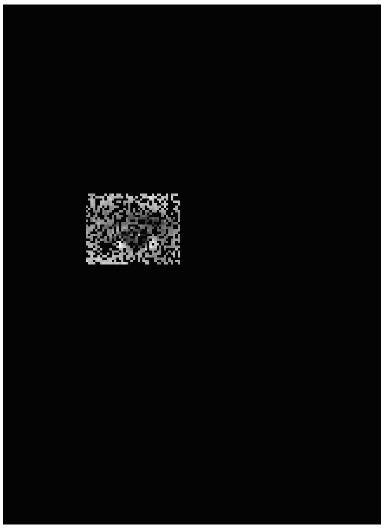


Bayesian image reconstruction

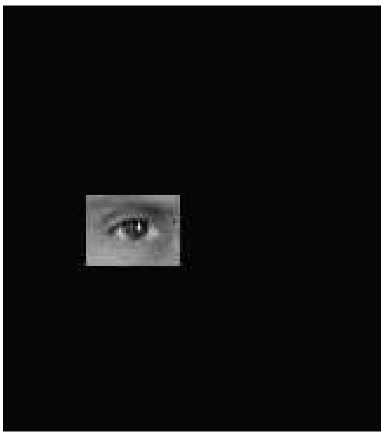
Example data:



100 data points

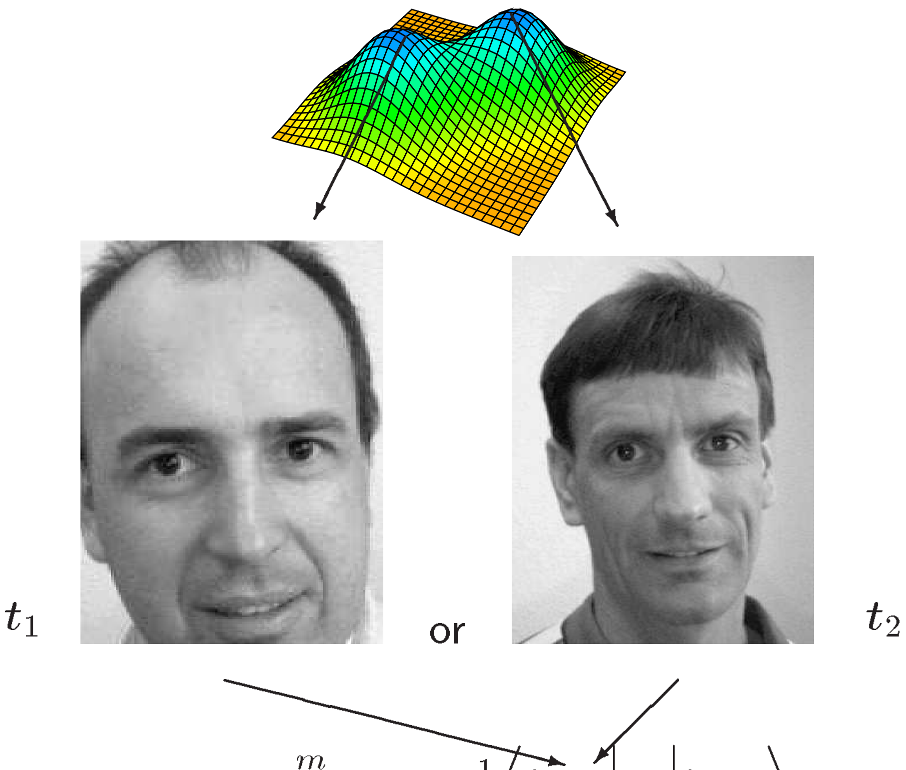


1000 data points



10000 data points

Priors as mixtures of Gaussian prozesses



$$p(h) \propto \sum_{j=1}^{m} a_j \; e^{-\frac{1}{2}\left\langle h - t_j \middle| \mathbf{K}_j \middle| h - t_j \right\rangle}$$

Expectation of predictive density

$$\bar{y}(x) = \int dy \; y \, p(y|x, D) = \sum_j b_j \, \bar{t}_j$$

with component solution $\bar{t}_j = t_j + \mathbf{K}_j^{-1}\mathbf{K}_{t_D}(j)(t_D - t_j)$ and mixture coefficient

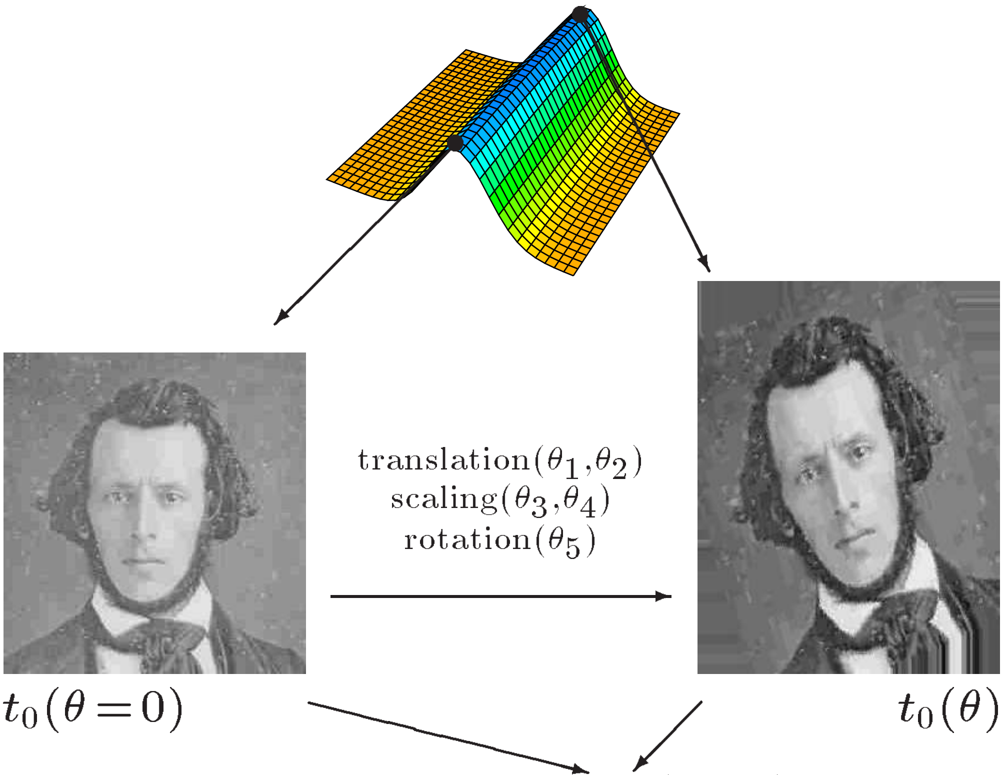
$$b_j = \frac{p(j)p(y_D|x_D, D_0, j)}{\sum_j p(j)p(y_D|x_D, j)},$$

$$p(y_D|x_D, j) = e^{-\frac{\beta}{2}\left\langle t_D - t_j \middle| \mathbf{K}_{t_D}(j) \middle| t_D - t_j \right\rangle + \frac{1}{2}\ln\det\left(\frac{\beta}{2\pi}\mathbf{K}_{t_D}(j)\right)}.$$

The $\tilde{n} \times \tilde{n}$ matrix $\mathbf{K}_{t_D}(j) = (\mathbf{K}_D^{-1} + \mathbf{K}_{j,DD}^{-1})^{-1}$ is defined in the space of data points, $\mathbf{K}_{j,DD}^{-1}$ being the projection of the prior covariance $\mathbf{K}_j^{-1}$ into that space.

J.L., ICANN99 (physics/9911077) [J.L., EUFIT'98 + FNS'98]

Flexible priors using hyperparameters



$$p(h) \propto \int d\theta\; a(\theta)\; e^{-\frac{1}{2}\left\langle h - t_0(\theta) \middle| \mathbf{K}_0(\theta) \middle| h - t_0(\theta) \right\rangle}$$

Expectation of predictive density (for combination with mixture model)

$$\bar{y}(x) = \int dy\, y\, p(y|x,D) = \sum_j \int d\theta\; b_j(\theta)\, \bar{t}_j(\theta)$$

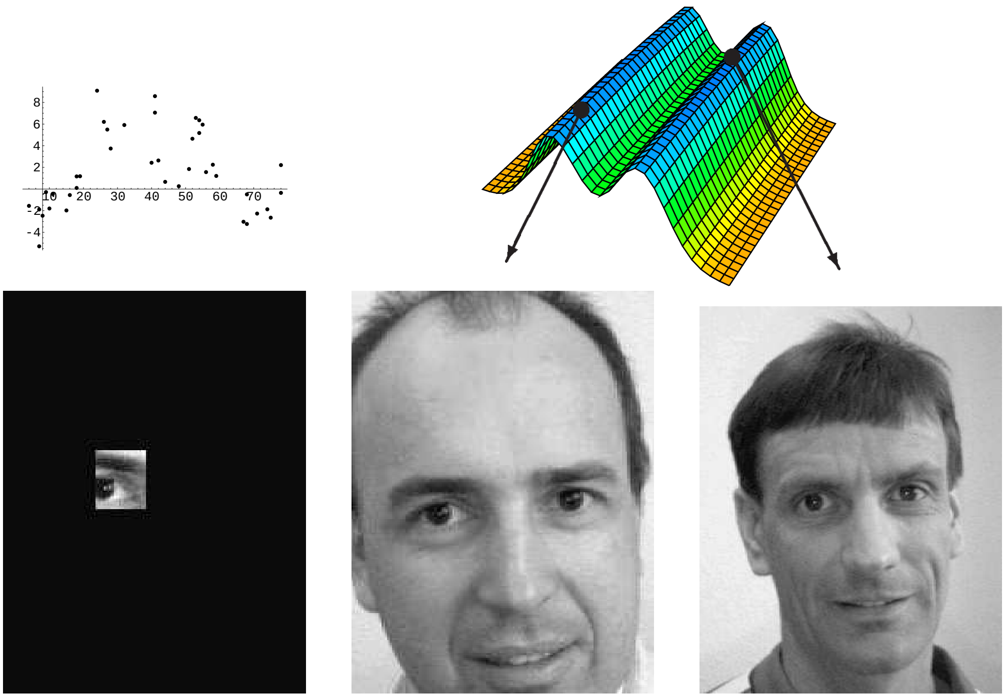
with mixture coefficients

$$b_j(\theta) = \frac{p(\theta,j)\, p(y_D|x_D,\theta,j)}{\sum_j \int d\theta\, p(\theta,j)\, p(y_D|x_D,\theta,j)}$$

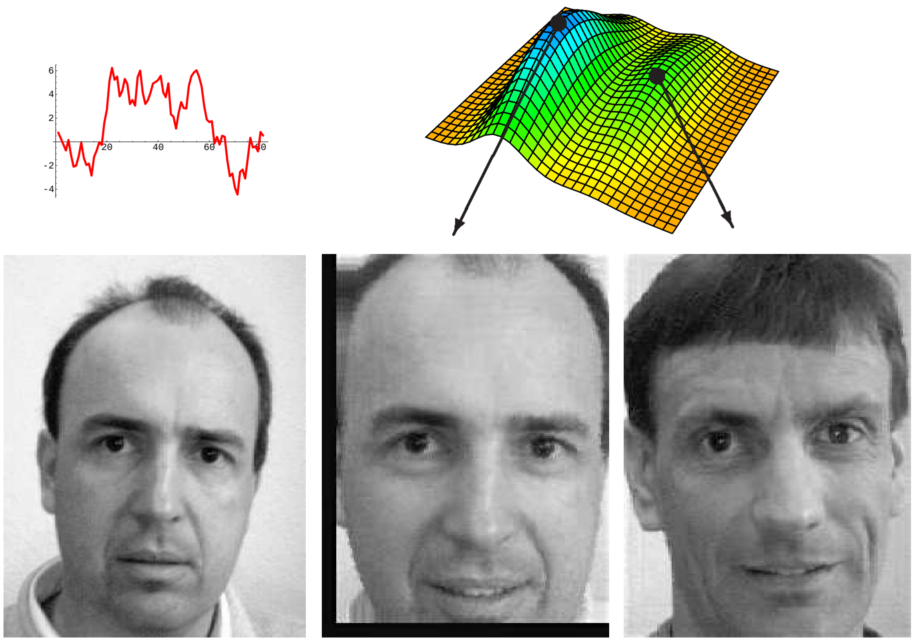
The MAP–Stationarity equation for θ becomes

$$\begin{aligned}
0 \quad = \quad & \sum_j \Bigg[\left\langle \frac{\partial t_j(\theta)}{\partial\theta} \middle| \mathbf{K}_{t_D}(\theta,j) \middle| t_j(\theta) - t_D \right\rangle \\
& + \frac{1}{2} \left\langle t_D - t_j(\theta) \middle| \frac{\partial \mathbf{K}_{t_D}(\theta,j)}{\partial\theta} \middle| t_D - t_j(\theta) \right\rangle \\
& - \mathrm{Tr}\left(\mathbf{K}_{t_D}^{-1}(\theta,j) \frac{\partial \mathbf{K}_{t_D}(\theta,j)}{\partial\theta} \right) - \frac{1}{p(\theta,j)} \frac{\partial p(\theta,j)}{\partial\theta} \Bigg]
\end{aligned}$$

Image reconstruction: Low–temperature results

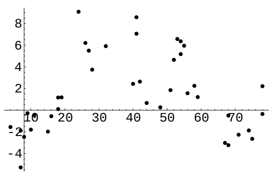


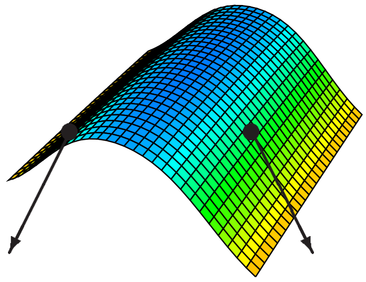
data t_1 t_2

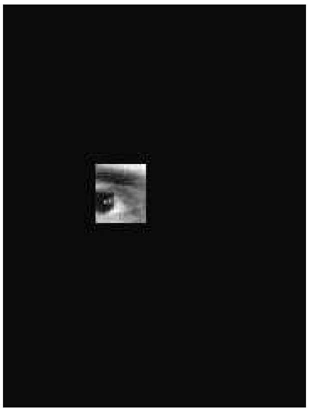


original reconstructed local maximum

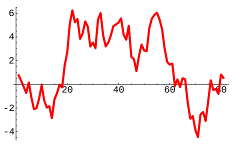
Image reconstruction: High–temperature results

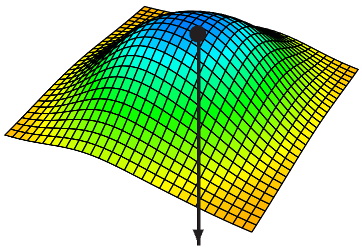






data t_1 t_2





original

high–temperature reconstruction