LIMITATIONS ON DETECTION AND CLASSIFICATION FROM COMPRESSED IMAGES

R. Gupta, A. O. Hero III
Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor

Jan. 7, 2000

Outline of Presentation

- Overview of Vector Quantization (VQ)
- Motivation for Compression with Detection Criteria
- Optimal VQ for Detection vs. Estimation
- Examples
- Conclusions and Future Research

Scalar and Vector Quantization

Let X be a k-dimensional random process with density $q(x)$.

An N-point quantizer, (S, C) consists of:

- Cells $S = \{S_0, \ldots, S_{N-1}\}$, disjoint and exhaustive.
- Codebook $C = \{x_0, \ldots, x_{N-1}\}$, $x_i \in \mathbb{R}^k$.

Thus the quantization rule is:

$$\hat{x} = x_i, \text{ if } x \in S_i.$$

The MSE of the quantized source is

$$\mathcal{E} = \int (x - \hat{x})^2 q(x) dx = \sum_{i=0}^{N-1} \int_{S_i} (x - x_i)^2 q(x) dx$$

MSE-optimal (Lloyd-Max) quantizers have the following properties:

- x_i is the centroid of S_i:

$$x_i = \frac{\int_{S_i} x q(x) dx}{\int_{S_i} q(x) dx},$$

- Cells are partitioned such that quantizer is a nearest-neighbor mapping:

$$\text{if } x \in S_i,\ \|x - x_i\| \leq \|x - x_j\|, \forall j \neq i$$

For large N (using Bennet's integral),

$$\mathcal{E} \geq v(N,k) \int \frac{q(x)}{\zeta(x)^{2/k}} dx$$

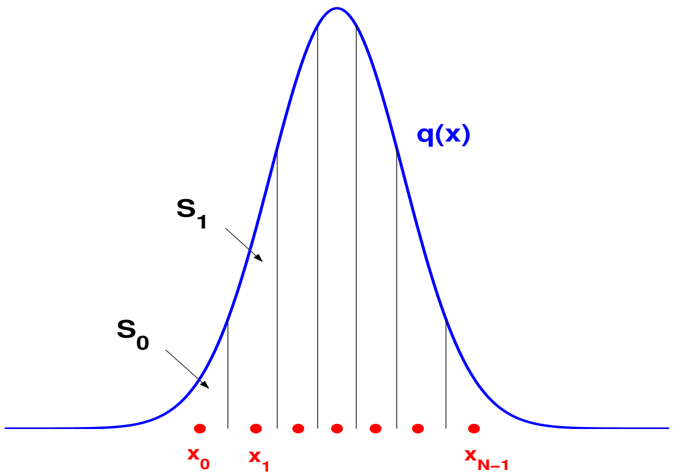
where $\zeta(x)$ is the "point density"

$$\zeta(x_i) \approx \frac{1}{NV(S_i)}, \qquad V(S_i) = \int_{S_i} dx.$$

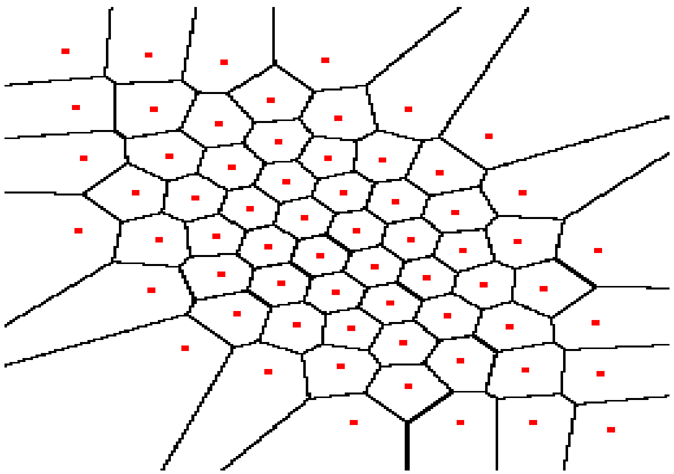
Hence, <u>quantizer can be designed by optimizing $\zeta(x)$</u>.

Examples of MSE-Optimal Quantizers for Gaussian Sources

Scalar Quantizer



Vector Quantizer



Hypothesis Testing (Detection)

Suppose the goal is to determine which of the two hypotheses is true:

$$\begin{aligned} H_0 &: X \sim q_0(x) \\ H_1 &: X \sim q_1(x). \end{aligned}$$

Define

$$\begin{aligned} \alpha &= P(\text{decide } H_1 | H_0) \quad \text{(False alarm probability)} \\ \beta &= P(\text{decide } H_0 | H_1) \quad \text{(Probability of miss).} \end{aligned}$$

The optimal (Neyman-Pearson) decision rule is

$$\begin{aligned} &\text{Decide } H_0 \;\text{ if } \Lambda(x) \geq T \\ &\text{Decide } H_1 \;\text{ if } \Lambda(x) < T \end{aligned}$$

where $\Lambda(x) = \log \frac{q_0(x)}{q_1(x)}$ is the log-likelihood ratio.

Discrimination between Sources

The (differential) discrimination (a.k.a. relative entropy, Kullback-Leibler distance) between $q_0(x)$ and $q_1(x)$ is defined as

$$L(q_0, q_1) = \int q_0(x) \log \frac{q_0(x)}{q_1(x)} dx.$$

Define the "tilted" density

$$q_\lambda(x) = \frac{q_0(x)^{1-\lambda} q_1(x)^\lambda}{\int q_0(x)^{1-\lambda} q_1(x)^\lambda dx}, \qquad \lambda \in [0, 1].$$

For large k (Blahut, 1987),

$$\alpha \approx e^{-kL(q_\lambda, q_0)} \qquad \beta \approx e^{-kL(q_\lambda, q_1)}.$$

The discrimination <u>can not increase</u> with processing (e.g. quantization).

Discrimination between Quantized Sources

Define

$$\Delta L_0 = L(q_\lambda, q_0) - L(\hat{q}_\lambda, \hat{q}_0), \qquad \Delta L_1 = L(q_\lambda, q_1) - L(\hat{q}_\lambda, \hat{q}_1)$$

where $\hat{q}_0$ and $\hat{q}_1$ are the densities of the quantized source under hypotheses H_0 and H_1, respectively, and $\hat{q}_\lambda$ is the corresponding tilted density. Also define

$$\begin{aligned} \hat{\alpha} &= \text{False alarm probability with quantized data} \\ \hat{\beta} &= \text{Probability of miss with quantized data.} \end{aligned}$$

Then:

$$\begin{aligned} \hat{\alpha} &= \alpha e^{\lambda k \Delta} \\ \hat{\beta} &= \beta e^{(1-\lambda) k \Delta} \end{aligned}$$

where

$$\Delta = \Delta L_0 + \Delta L_1$$

VQ Design for Maximum Discrimination

For a many-point VQ (large N), we find that

$$\Delta \;=\; \lim_{N\to\infty} \frac{1}{2} \sum_{i=0}^{N-1} q_\lambda(x_i) \int_{S_i} (x - x_i)^T \mathcal{F}_{x_i} (x - x_i) dx$$

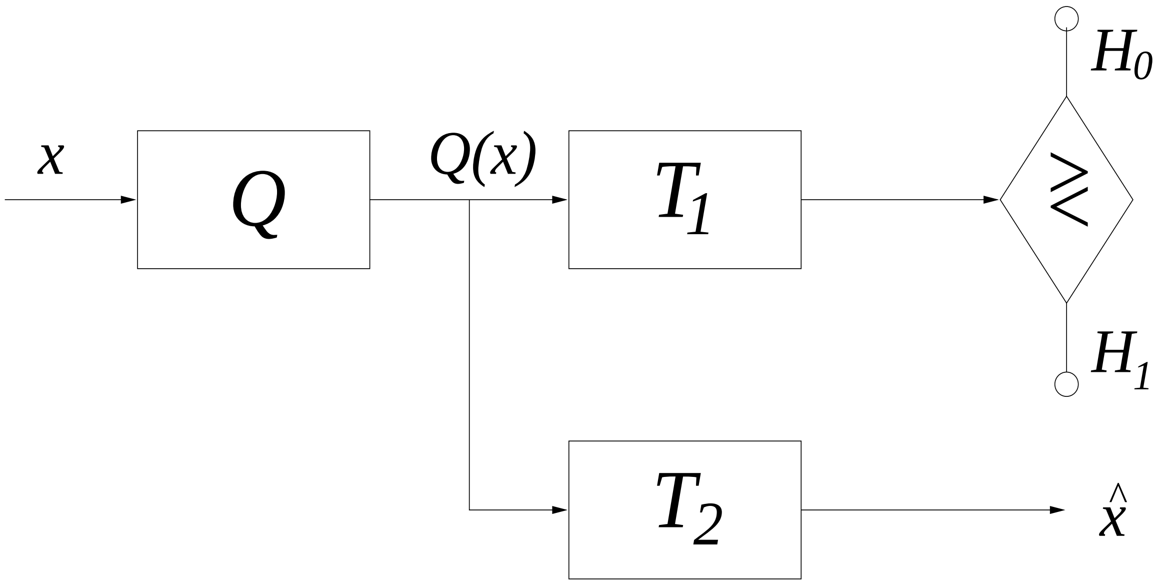
where

$$\mathcal{F}_y \;=\; \nabla\Lambda(y)\nabla\Lambda(y)^T, \qquad \Lambda(x) = \log \frac{q_0(x)}{q_1(x)}$$

$$\Rightarrow \Delta = u(N,k) \int \frac{q_\lambda(x)\|\nabla\Lambda(x)\|^2}{\zeta(x)^{2/k}} dx$$

Once again, <u>quantizer can be designed by optimizing $\zeta(x)$</u>.

Joint Detection / Estimation Design Criterion



Objective: Design Q to be optimal wrt detection and estimation performance simultaneously.

Quantizer Design Criterion 1: Average MSE

Optimal for <u>estimation</u>

$$\bar{\mathcal{E}} = \int (x - \hat{x})^2 \bar{q}(x) dx, \qquad \bar{q}(x) = P_0 q_0 + (1 - P_0) q_1$$

Optimal Point Density:

$$\zeta(x) = \frac{\bar{q}(x)^{\frac{k}{k+2}}}{\int \bar{q}(x)^{\frac{k}{k+2}} dx}$$

MSE Lower Bound:

$$\bar{\mathcal{E}}_1 \geq v(N, k) \int \bar{q}(x)^{\frac{2}{k+2}} dx \int \bar{q}(x)^{\frac{k}{k+2}} dx$$

Discrimination:

$$\Delta_1 = u(N, k) \left(\int \bar{q}(x)^{\frac{k}{k+2}} dx \right)^{2/k} \int \bar{q}(x)^{\frac{-2}{k+2}} q_\lambda(x) \|\nabla \Lambda(x)\|^2 dx$$

Quantizer Design Criterion 2: Tilted MSE

$$\mathcal{E}_\lambda = \int (x - \hat{x})^2 q_\lambda(x) dx$$

Optimal Point Density:

$$\zeta(x) = \frac{q_\lambda(x)^{\frac{k}{k+2}}}{\int q_\lambda(x)^{\frac{k}{k+2}} dx}$$

MSE Lower Bound:

$$\bar{\mathcal{E}}_2 \geq v(N,k) \left(\int q_\lambda(x)^{\frac{k}{k+2}} dx \right)^{2/k} \int \bar{q}(x) q_\lambda(x)^{\frac{-2}{k+2}} dx$$

Discrimination:

$$\Delta_2 = u(N,k) \left(\int q_\lambda(x)^{\frac{k}{k+2}} dx \right)^{2/k} \int q_\lambda(x)^{\frac{k}{k+2}} \|\nabla \Lambda(x)\|^2 dx$$

Quantizer Design Criterion 3: Discrimination

Optimal for <u>detection</u>

$$\Delta = \Delta L_0 + \Delta L_1$$

Optimal Point Density:

$$\zeta(x) = \frac{[q_\lambda(x)\|\nabla\Lambda(x)\|^2]^{\frac{k}{k+2}}}{\int [q_\lambda(x)\|\nabla\Lambda(x)\|^2]^{\frac{k}{k+2}} dx}$$

MSE Lower Bound:

$$\bar{\mathcal{E}}_3 \geq v(N,k) \left(\int [q_\lambda(x)\|\nabla\Lambda(x)\|^2]^{\frac{k}{k+2}} dx \right)^{2/k} \int \bar{q}(x)[q_\lambda(x)\|\nabla\Lambda(x)\|^2]^{\frac{-2}{k+2}} dx$$

Discrimination:

$$\Delta_3 = u(N,k) \left(\int [q_\lambda(x)\|\nabla\Lambda(x)\|^2]^{\frac{k}{k+2}} dx \right)^{\frac{k+2}{k}}$$

Quantizer Design Criterion 4: Combined $\mathcal{E}_\lambda$ and Δ

$$J = \rho\Delta + (1-\rho)\mathcal{E}_\lambda, \qquad \rho \in [0,1]$$

For large N,

$$J = ku(N,k)\int \frac{q_\lambda(x)\left(\frac{\rho}{k}\|\nabla\Lambda(x)\|^2 + 1 - \rho\right)}{\zeta(x)^{2/k}}dx$$

Optimal Point Density:

$$\zeta(x) = \frac{[q_\lambda(x)(\frac{\rho}{k}\|\nabla\Lambda(x)\|^2 + 1 - \rho)]^{\frac{k}{k+2}}}{\int[q_\lambda(x)(\frac{\rho}{k}\|\nabla\Lambda(x)\|^2 + 1 - \rho)]^{\frac{k}{k+2}}dx}$$

Quantizer Design Criterion 4: Combined $\mathcal{E}_\lambda$ and Δ (Contd.)

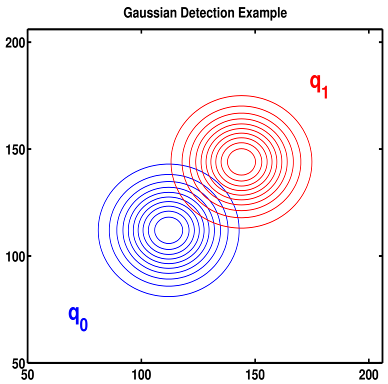
MSE Lower Bound:

$$\bar{\mathcal{E}}_4 \geq v(N,k)\left(\int \left[q_\lambda(x)\left(\frac{\rho}{k}\|\nabla\Lambda(x)\|^2+1-\rho\right)\right]^{\frac{k}{k+2}}dx\right)^{2/k}\cdot \int \bar{q}(x)\left[q_\lambda(x)\left(\frac{\rho}{k}\|\nabla\Lambda(x)\|^2+1-\rho\right)\right]^{\frac{-2}{k+2}}dx$$

Discrimination:

$$\Delta_4 = u(N,k)\left(\int \left[q_\lambda(x)\left(\frac{\rho}{k}\|\nabla\Lambda(x)\|^2+1-\rho\right)\right]^{\frac{k}{k+2}}dx\right)^{\frac{2}{k}}\cdot \int q_\lambda(x)^{\frac{k}{k+2}}\|\nabla\Lambda(x)\|^2\left(\frac{\rho}{k}\|\nabla\Lambda(x)\|^2+1-\rho\right)^{\frac{-2}{k+2}}dx$$

Comparison of VQ's: Gaussian Hypotheses



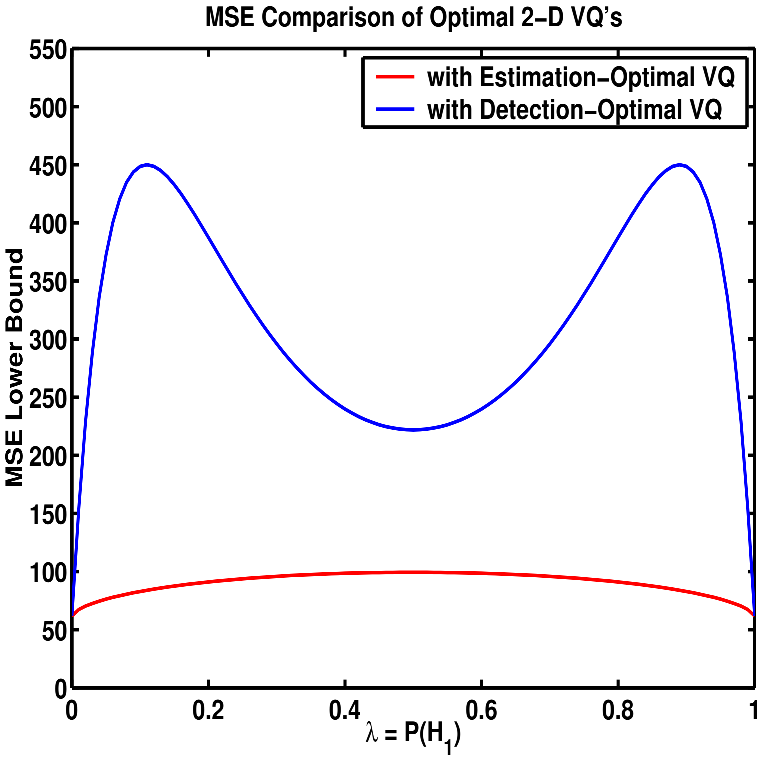
$$q_0 \sim N(\boldsymbol{\mu_0}, \mathbf{K_0}) \qquad q_1 \sim N(\boldsymbol{\mu_1}, \mathbf{K_1})$$

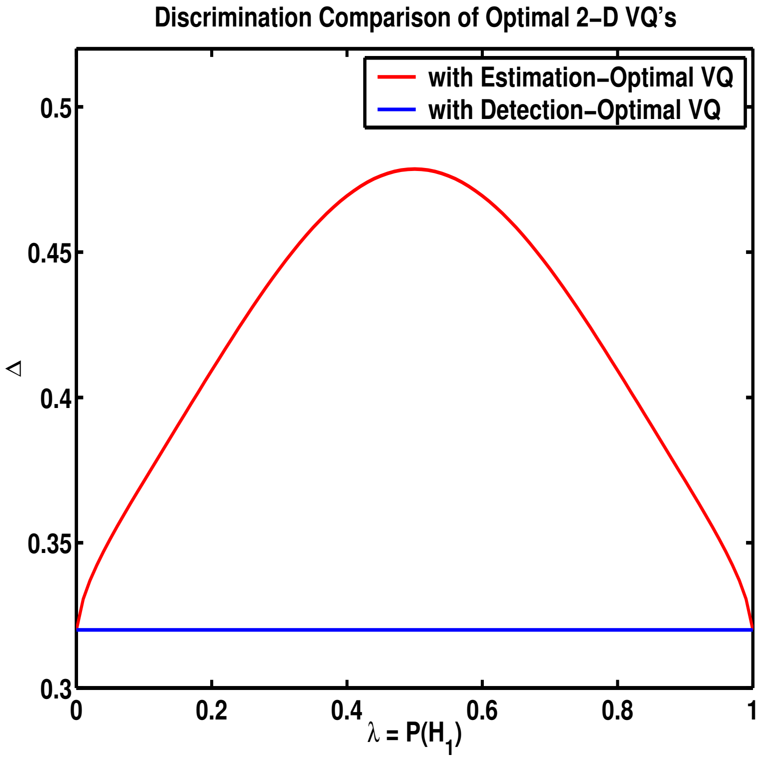
$$\boldsymbol{\mu_0} = [112, 112]^T \qquad \boldsymbol{\mu_1} = [144, 44]^T$$

$$\mathbf{K_0} = 200I \qquad \mathbf{K_1} = 200I$$

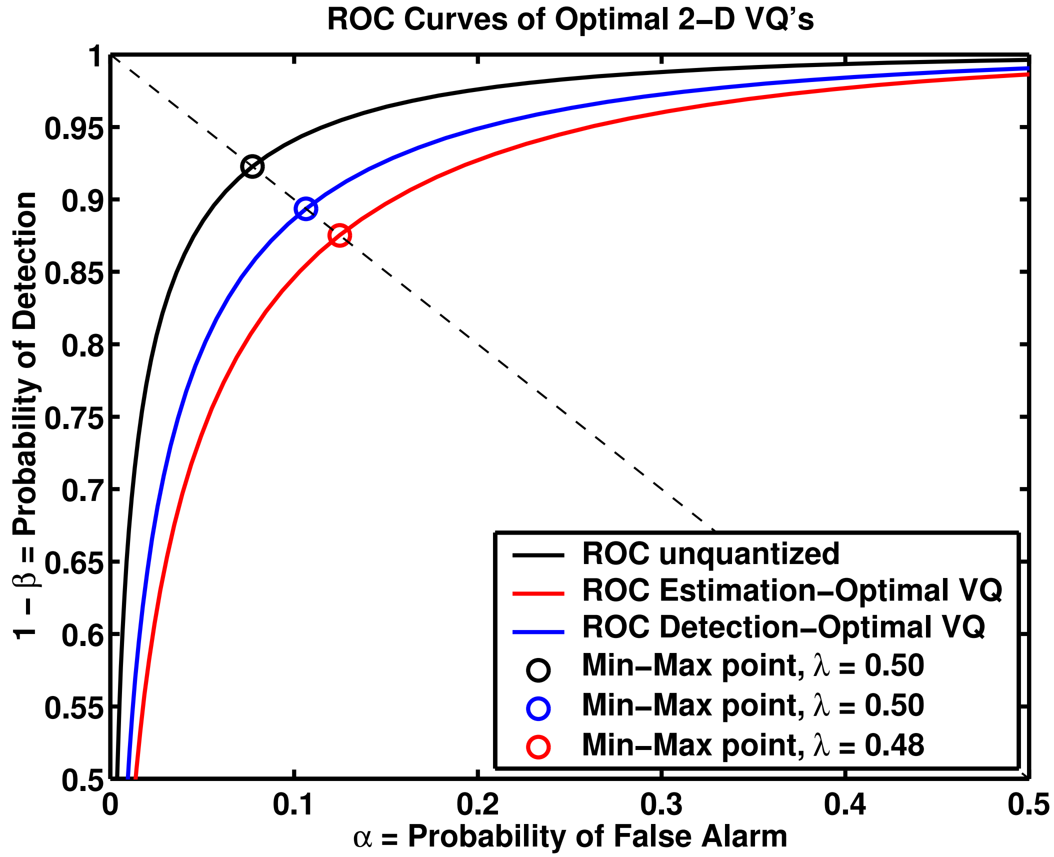
Consider optimal VQ's of dimension $k = 2$, with $N = 128$ cells.

Comparison of VQ's: MSE and Discrimination





Comparison of VQ's: ROC Curves



Implementation: Lloyd Algorithm

`Step 1:` Initialize codebook, C_1.

`Step 2:` Find optimal nearest-neighbor partition.

`Step 3:` Move codebook points to centroids (wrt pdf, $q(x)$) of their cells.

`Step 4:` Compute total distortion. If change in distortion is small enough, stop. Otherwise, go to `Step 2`.

- Results in MSE-Optimal quantizer (with correct initialization).
- If point density $\zeta(x)$ is known, can replace $q(x)$ with $\zeta(x)$ in Lloyd algorithm.

Gaussian Example

$$\mu_{\mathbf{0}} = 64, \quad \mu_{\mathbf{1}} = 192, \quad \mathbf{K_0} = 1000I, \quad \mathbf{K_1} = 500I, \quad N = 25, \quad \lambda = 1/2$$

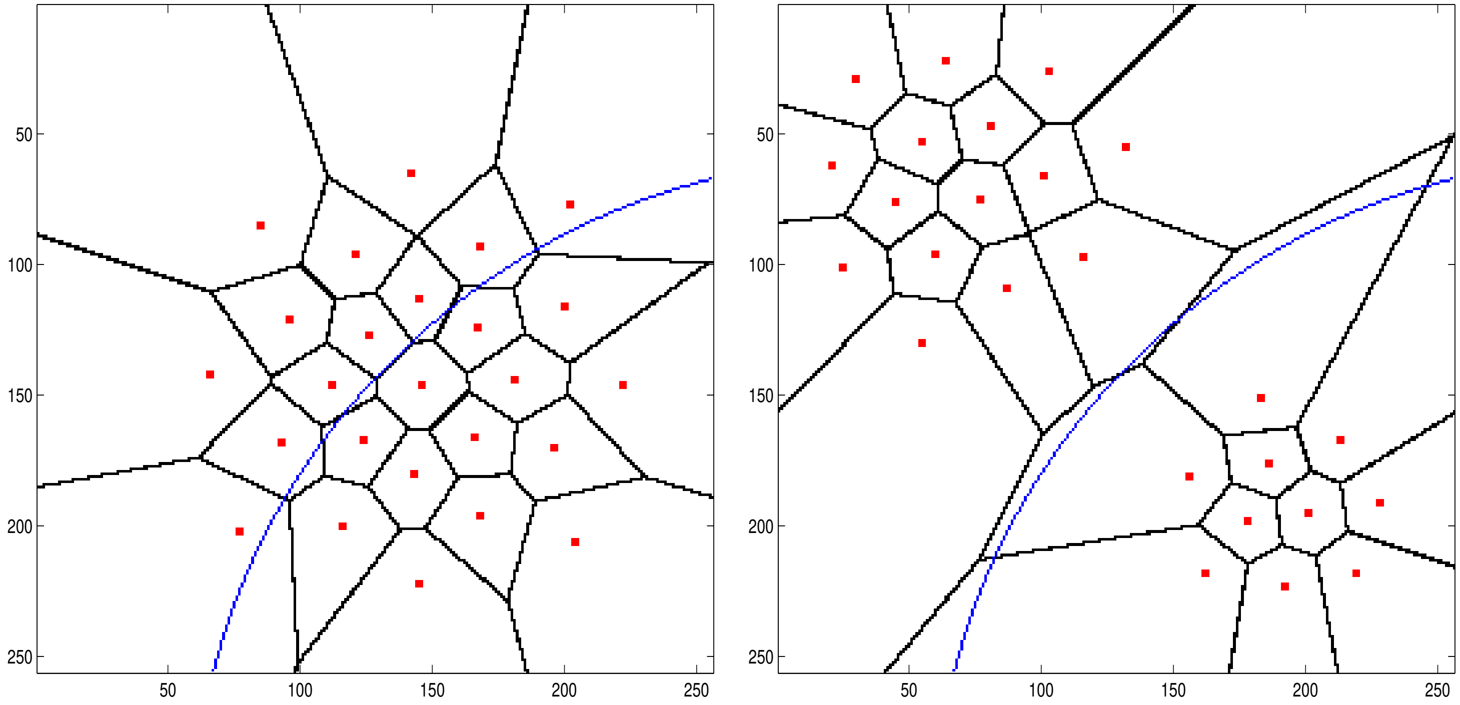
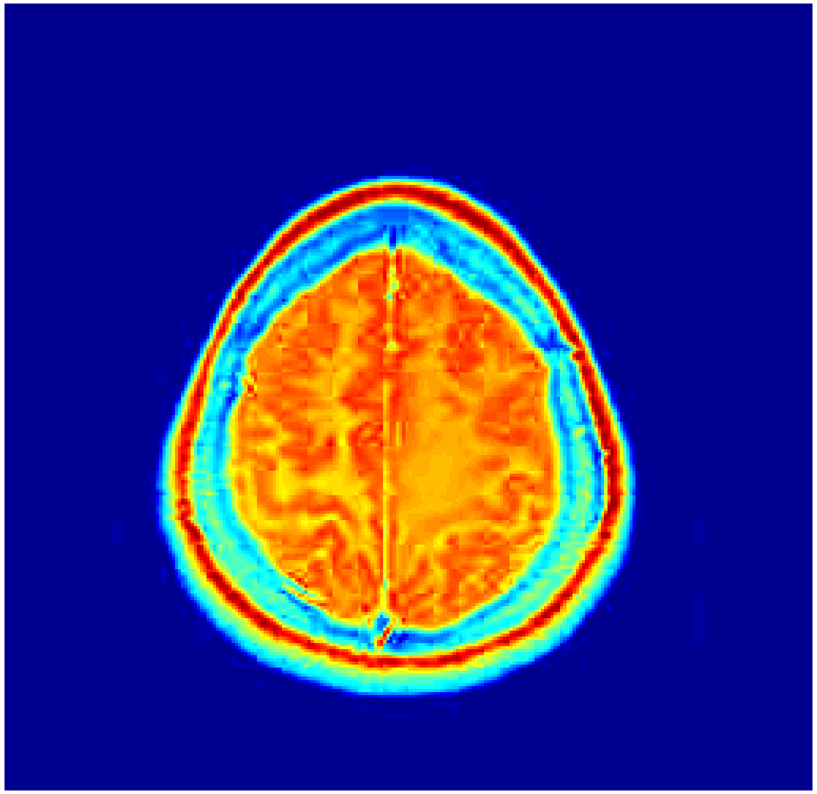
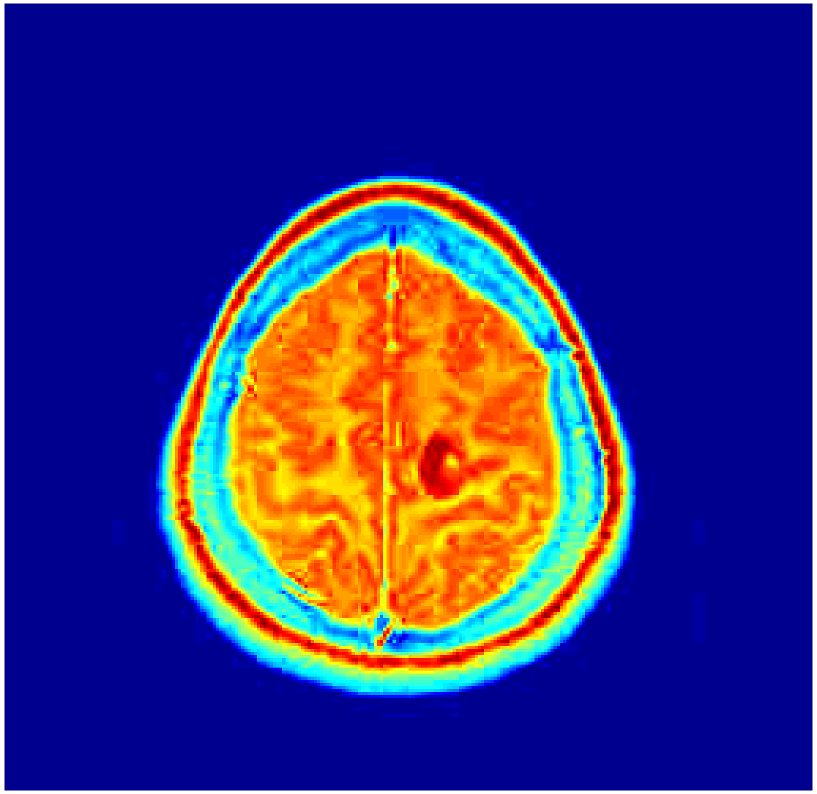
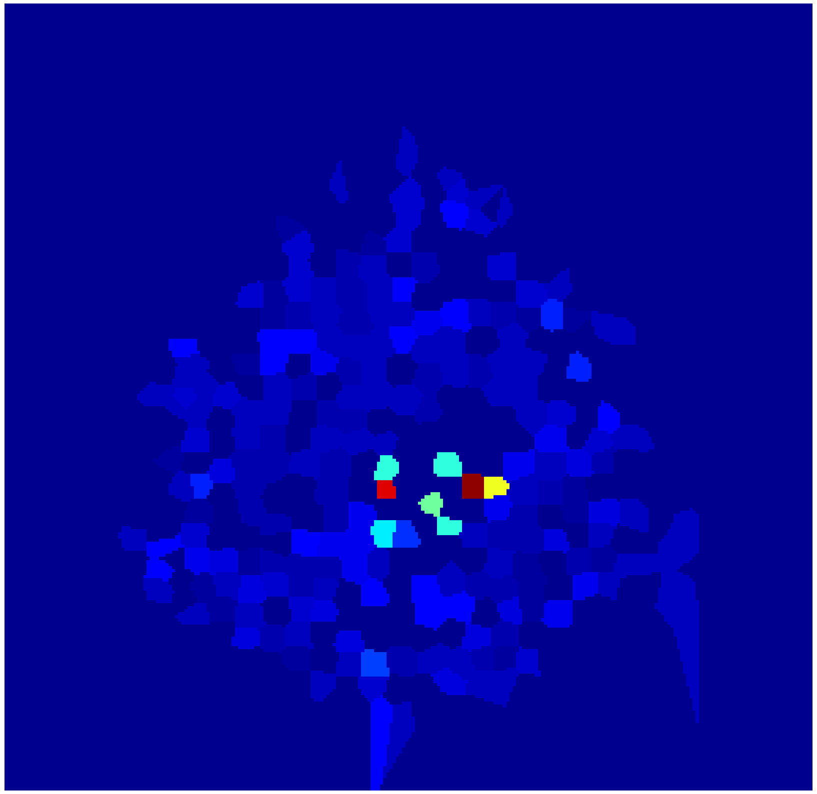


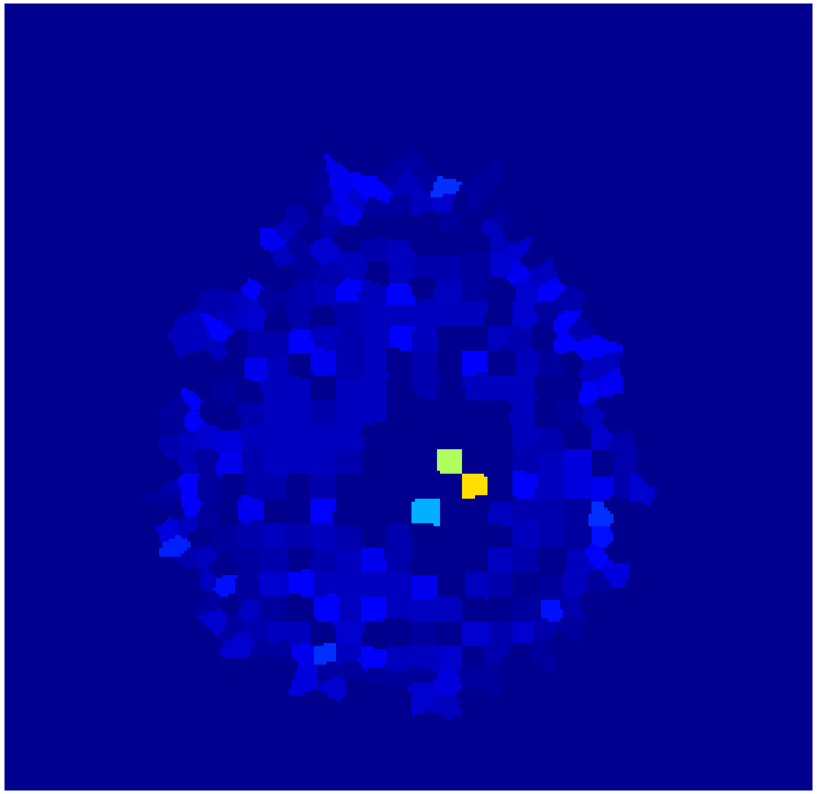
Image Compression Example



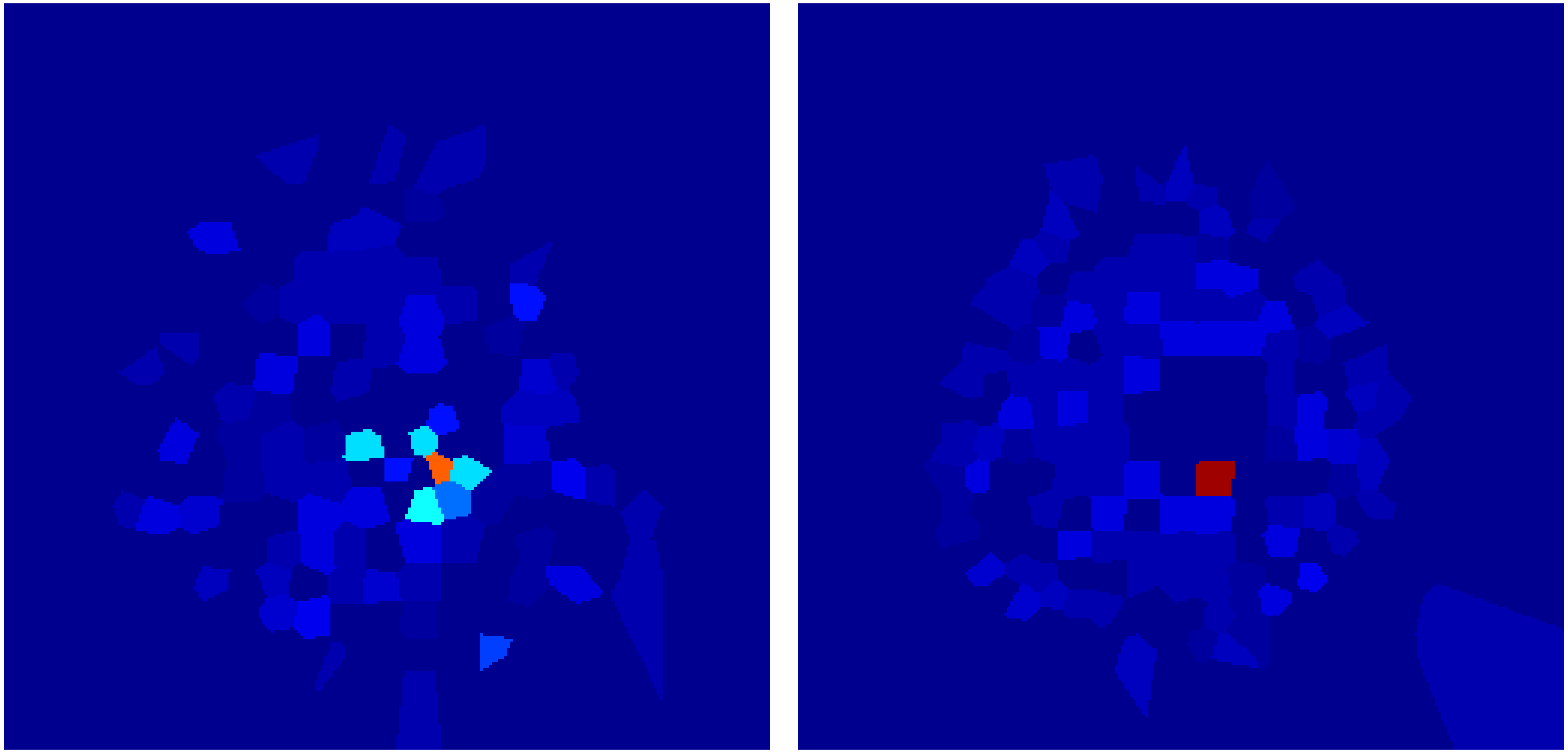


Difference Images with $N = 1024$ Cells, $\lambda = 1/2$





Difference Images with $N = 400$ Cells, $\lambda = 1/2$



Discrimination Distance

Recall that discrimination decrease is

$$\Delta = \lim_{N\to\infty} \frac{1}{2} \sum_{i=0}^{N-1} q_\lambda(x_i) \int_{S_i} (x - x_i)^T \mathcal{F}_{x_i} (x - x_i) dx.$$

Hence $D(x, x_i) = (x - x_i)^T \mathcal{F}_{x_i}(x - x_i)$ can be used as distortion metric (as opposed to Euclidean distance) in Lloyd algorithm.

For Gaussian i.i.d. hypotheses with $\mathbf{K_0} = \mathbf{K_1} = \sigma^2 I$:

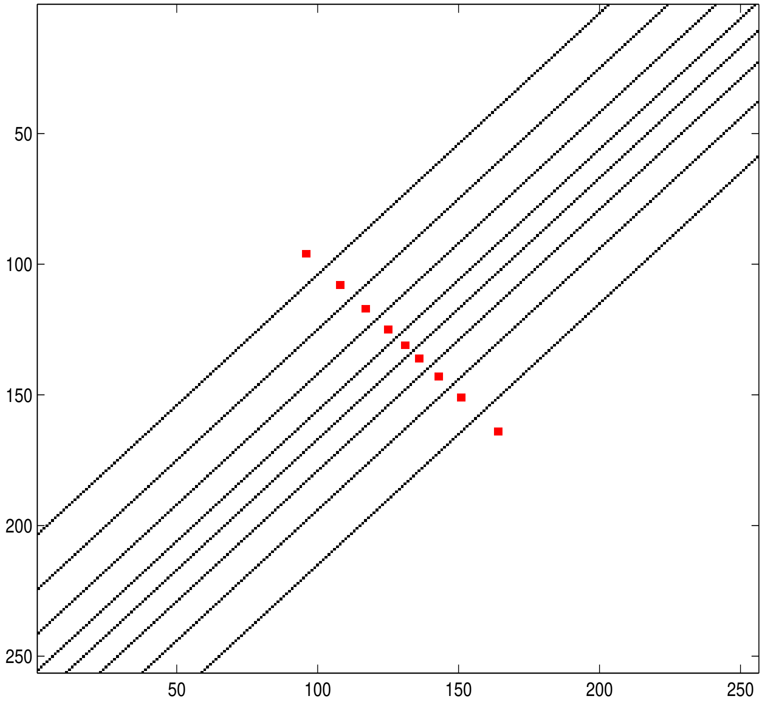
$$\mathcal{F}_x = \frac{(\mu_0 - \mu_1)^2}{\sigma^4} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\Rightarrow D(x, x_i) = \frac{(\mu_0 - \mu_1)^2}{\sigma^4} [(x + y) - (x_i + y_i)]^2$$

$\Rightarrow$ Quantize optimally w.r.t. MSE of $z = x + y$ under density q_λ.

$\boxed{\text{Example with } \lambda = 0}$

$$q_\lambda(x) = q_0(x) \sim N(128, 500)$$



Conclusions

- Optimal VQ's for detection can differ greatly from MSE-optimal VQ's.
- Optimization method: point density
- Penalty for not using detection-optimal quantizer: significant loss in discrimination.
- For some cases, VQ is optimal over all prior probabilities, $P(H_0)$.

Future Research

- Application: Compression of large databases with Detection vs. Estimation criteria
- Non-asymptotic theory