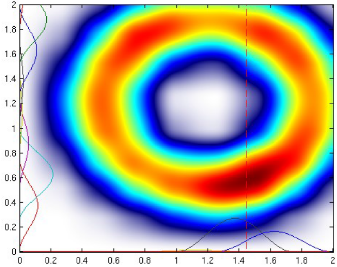
Channel Representations for Machine Learning

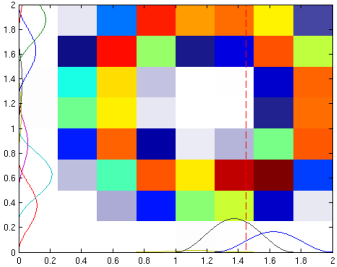
Michael Felsberg and Klas Nordberg

Department of Electrical Engineering
Computer Vision Laboratory
Linköping University

Associative Mappings

[Granlund, Invited Talk AFPAC 2000]

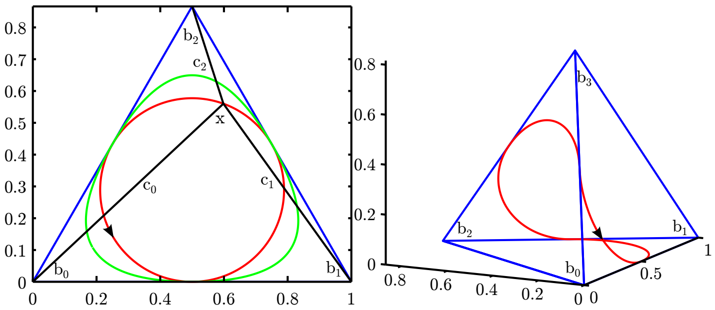




Channel Smoothing [Felsberg et al., TPAMI 2006]

Basis functions form a tight frame, enabling reconstruction

Geometric Interpretation



Curves with constant distance to origin (constant l_2-norm) on the surface of a simplex (constant l_1-norm)

Decoding of Channel Representations

The $\cos^2$-kernel

- Is a tight frame (the only tight channel basis with overlap 3)
- Decoding with an orthogonal matrix and argument calculation [Forssén, PHD 2004] (up to scaling of r_2)

$$\begin{pmatrix} r_1 \cos(\frac{2\pi}{3}x) \\ r_1 \sin(\frac{2\pi}{3}x) \\ r_2 \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{2}{3}} & -\sqrt{\frac{1}{6}} & -\sqrt{\frac{1}{6}} \\ 0 & \sqrt{\frac{1}{2}} & -\sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{3}} & \sqrt{\frac{1}{3}} & \sqrt{\frac{1}{3}} \end{pmatrix} \mathbf{c} = \begin{pmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \mathbf{w}_3 \end{pmatrix} \mathbf{c}$$

For more than 3 channels, consider the 3-window $\mathbf{c}_l$ shifted by l

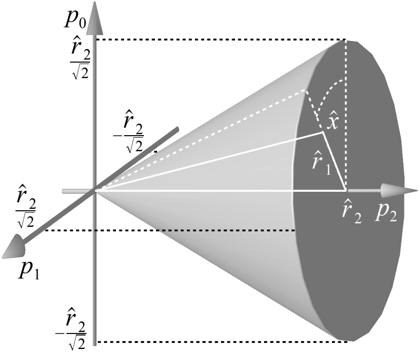
Decoding of Channel Representations

$r_1 \exp(i2\pi\hat{x}/3) =$
$(\mathbf{w}_1 + i\mathbf{w}_2)^T \mathbf{c}$

value estimate $\hat{x}$

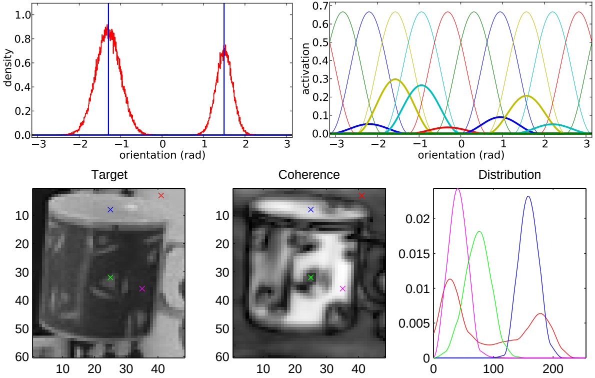
evidence estimate
$r_2 = \mathbf{w}_3^T \mathbf{c}$

coherence estimate
r_1^2/r_2^2 (max. $1/2$)



Most of subsequent material: [Felsberg et al., Frontiers in Robotics and AI, 2015]

Extending EDFT: Coherence



[Öfjäll&Felsberg, ECCVWS VOT 2014]

Weighted Distance

coherence-based distance

$$\mathrm{coh}(\mathbf{c}_l) = \frac{r_1^2}{r_2^2} = \frac{1}{\mathbf{1}^\mathrm{T}\mathbf{c}_l\mathbf{c}_l^\mathrm{T}\mathbf{1}}\ \mathbf{c}_l^\mathrm{T} \begin{pmatrix} 4 & -2 & -2 \\ -2 & 4 & -2 \\ -2 & -2 & 4 \end{pmatrix} \mathbf{c}_l$$

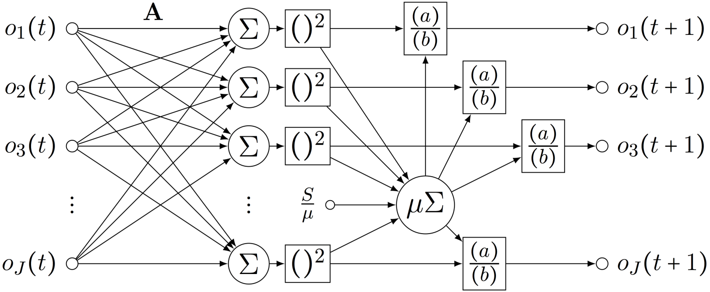
$$\begin{aligned} D_{\mathrm{coh}}(\mathbf{C}_{\mathrm{model}},\ \mathbf{C}_f) \quad &= \quad \sum_{i,j,k} |[\mathbf{C}_{\mathrm{model}}]_{ijk} - [\mathbf{C}_f]_{ijk}| \cdot \\ & \quad\quad \cdot (\mathrm{coh}([\mathbf{C}_{\mathrm{model}}]_{ij}) + \kappa) \end{aligned}$$

[Öfjäll&Felsberg, ECCVWS VOT 2014]

Denève's Decoding Approach

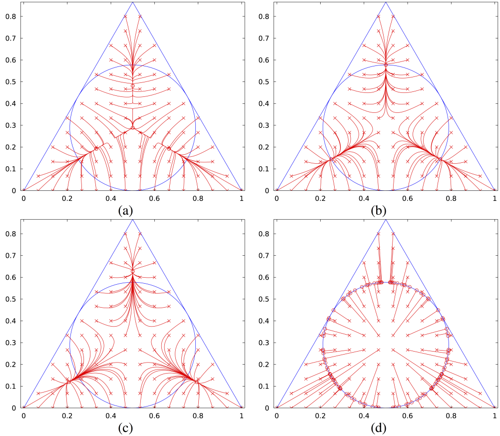
[Denève et al., Nature Neuroscience 1999]

$$\mathbf{u} = \mathbf{A}\mathbf{o} \qquad \text{and for all } j : \qquad o_j = u_j^2/(S + \mu|\mathbf{u}|^2)$$



claimed to be bias-free

Synthetic 3-Channel Example



Theorems on Decoding

Theorem

(Bias of Denève's scheme) Denève's scheme is biased unless we choose (overlap 3)

$$\mathbf{A} = \sqrt{\frac{S}{3(1-\mu)}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \quad , \text{ where } \mu < 1 \ . \qquad (1)$$

Theorems on Decoding

Theorem

(Constraint on channel coefficients) The decoding of a 3-channel vector $\mathbf{c}$ *is invariant under a change of channel coefficients* $\nabla_{\mathbf{c}} = (\partial_{c_0}, \partial_{c_1}, \partial_{c_2})^{\mathrm{t}}$ *iff*

$$(c_0 - c_1)\partial_{c_2} + (c_2 - c_0)\partial_{c_1} + (c_1 - c_2)\partial_{c_0} = 0 \ . \quad (2)$$

Proof is based on assuming no rotation in the $p_0 - p_1$-plane.

ML Decoding of $J > 3$ Channels

Assuming Gaussian distributed noise of the channel coefficients c_k, we obtain the MLE of x as ($[x]$ is the closest integer to x)

$$\min_x \quad \|\mathbf{c} - (\ldots, 0, \underbrace{K(x-[x]+1)}_{\text{coefficient}[x]-1}, \underbrace{K(x-[x])}_{\text{coefficient}[x]}, \underbrace{K(x-[x]-1)}_{\text{coefficient}[x]+1}, 0, \ldots)\|_2^2$$

Theorem

(Maximum-likelihood decoding) Assuming independent Gaussian noise on the channel coefficients, the decoding 3-window with maximum likelihood is given at the index $\hat{j}$

$$\hat{j} = \arg\max_j \tilde{r}_1(j) + \sqrt{2} r_2(j) \quad , \tag{3}$$

where $r_i(j)$ *is computed as* r_i *of* $\mathbf{c}_j := (c_{j-1}, c_j, c_{j+1})^{\mathrm{t}}$,

$$\tilde{r}_1(j) = \begin{cases} r_1(j) & \textit{if } |\alpha(\mathbf{c}_j)| \leq \pi/3 \\ r_1(j)\cos(|\alpha(\mathbf{c}_j)| - \pi/3) & \textit{else} \end{cases} ,$$

and $\alpha(\mathbf{c}_j) = \arg(\mathbf{c}_j^{\mathrm{t}}(\mathbf{w}_0 + i\mathbf{w}_1)) - \frac{2\pi}{3}$. *The MLE is given as* $\hat{x} = \max(\min(\frac{3}{2\pi}\alpha(\mathbf{c}_{\hat{j}}), \frac{1}{2}), -\frac{1}{2}) + \hat{j}$.

ML Decoding of $J > 3$ Channels

Using the constant l_1 norms of K and $\mathbf{c}$, the constant l_2 norm of K, and the fact that the $\mathbf{w}_{1,2,3}$ are orthonormal, we obtain an equivalent discrete optimization problem and obtain the optimal position j of the decoding window as

$$\hat{j} = \arg\max_j r_1(j) + \sqrt{2} r_2(j) \qquad (4)$$

except for the case that the argument α of the current decoding window is outside the valid range $(r_1(j) \mapsto \tilde{r}_1(j))$.

ML Decoding of $J > 3$ Channels

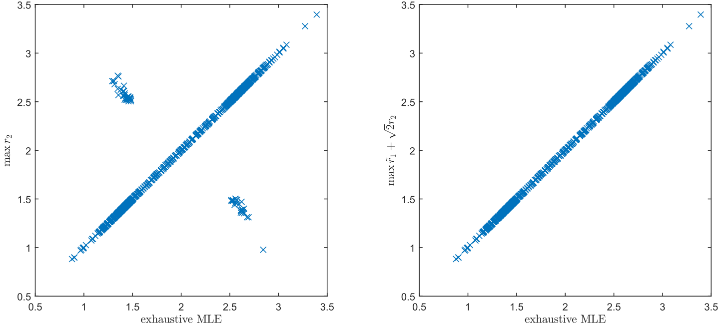
This discrete optimization problem possibly leads to other decoding windows than maximizing either r_1 or r_2 (as suggested in [Forssén, PHD 2004]).
Only for joint minimization we get $\hat{j} = [\hat{x}]$ and the MLE of x given $\mathbf{c}$ is thus

$$\hat{x} = \hat{j} + \arg((c_{\hat{j}-1}, c_{\hat{j}}, c_{\hat{j}+1})(\mathbf{w}_1 + i\mathbf{w}_2)) \qquad (5)$$

In practice, the coefficients are correlated and not i.i.d., thus an iterative scheme is required.

ML Decoding of $J > 3$ Channels

Experiment: Monte-Carlo simulation with 1000 samples



A New Recurrent Scheme

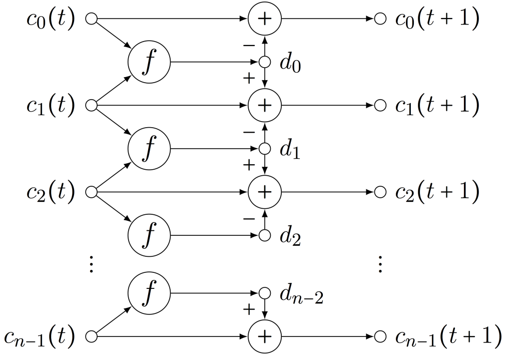
Mode invariance and evidence constancy: inverse linear diffusion (3 channels)

$$\mathbf{c}_0 = \mathbf{c} + \alpha \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix} \mathbf{c} \tag{6}$$

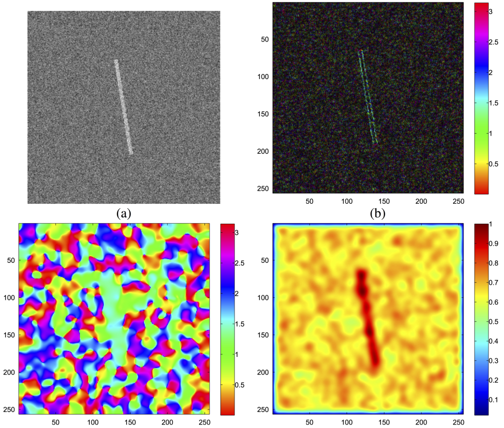
Note similarity to coherence formula: This is *coherence enhancing diffusion of channel coefficients*

A New Recurrent Scheme

In general: inverse non-linear diffusion, where d_j depend on the noise

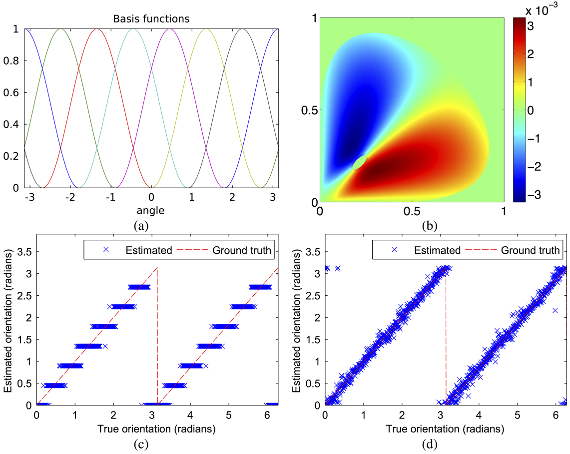


Example Stimulus



(a) (b) (c) (d)

Results



(a) (b) (c) (d)

Maximum-Entropy Decoding

- Problematic: compare distributions by absolute differences (DFT) or squared differences (MLE)
- Intuitive better: *simplest* distribution (maximum entropy) [Jonsson&Felsberg, SCIA2005]

$$H(p) = -\int_{-\infty}^{\infty} p(x) \log p(x) \, dx \qquad (7)$$

- that generates the same channel coefficients

$$\int_{-\infty}^{\infty} p(x) K(x - x_j) \, dx = c_j, \quad 1 \leq j \leq N \qquad (8)$$

$$\int_{-\infty}^{\infty} p(x) \, dx = 1 \qquad (9)$$

Maximum-Entropy Decoding

- Variational approach with Lagrange multipliers $0 \leq \lambda_j \leq N$, we obtain:

$$p(x) = \exp \lambda_0 \exp \left(\sum_{j=1}^{N} \lambda_j K(x - x_j) \right) \qquad (10)$$

- Note that non-negativity is implicitly achieved
- Explicit solution of λ_j impossible; apply Newton method instead

Approximative Solution

- Maximize approximate entropy (Taylor expansion) with same constraints

$$H_T(p) = \int_{-\infty}^{\infty} \frac{3}{2} p(x)(1 - p(x))\, dx \qquad (11)$$

- Using a variational approach with Lagrange multipliers $0 \leq \lambda_j \leq N$, we obtain:

$$p(x) = \frac{\lambda_0}{3} + \frac{1}{2} + \frac{1}{3} \sum_{j=1}^{N} \lambda_j K(x - x_j) \qquad (12)$$

Approximative Solution

Define $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_N)^T$ and $\tilde{\mathbf{c}} = (1, c_1, \ldots, c_N)^T$,

$$\mathbf{A}\boldsymbol{\lambda} = 3\tilde{\mathbf{c}} - \frac{3}{2} \qquad (13)$$

$$\mathbf{A} = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 & \ldots & 1 \\
1 & \frac{1}{2} & \frac{1}{6} + \frac{\sqrt{3}}{8\pi} & \frac{1}{12} - \frac{\sqrt{3}}{8\pi} & 0 & \ldots & 0 \\
1 & \frac{1}{6} + \frac{\sqrt{3}}{8\pi} & \ddots & \ddots & \ddots & \ddots & \vdots \\
1 & \frac{1}{12} - \frac{\sqrt{3}}{8\pi} & \ddots & \ddots & \ddots & \ddots & 0 \\
1 & 0 & \ddots & \ddots & \ddots & \ddots & \frac{1}{12} - \frac{\sqrt{3}}{8\pi} \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \frac{1}{6} + \frac{\sqrt{3}}{8\pi} \\
1 & 0 & \ldots & 0 & \frac{1}{12} - \frac{\sqrt{3}}{8\pi} & \frac{1}{6} + \frac{\sqrt{3}}{8\pi} & \frac{1}{2}
\end{pmatrix}$$

Approximative Solution

- Problem: $\hat{p}(x)$ can become negative
- Positivity requires for all $1 \leq j \leq N$ either

$$\lambda_j \geq 0 \quad \text{or} \quad \sqrt{\sum_{k=j-1}^{j+1} \lambda_k^2} \leq \sum_{k=j-1}^{j+1} \lambda_k \qquad (14)$$

- If this is not fulfilled, some function $\tilde{p}$ need to be added that does not influence constraints (8) & (9)
- Thus $\tilde{p}$ must be orthogonal to all K and without DC component
- Still looking for a solution!

Credits and Acknowledgments

- Researchers@CVL: Gösta, Per-Erik, Fahad, Klas, Reiner, Jörgen, Johan H
- Students@CVL: Amanda, Gustav, Kristoffer, Marcus, Martin
- Alumni@CVL: Erik J, Fredrik, Freddie, Johan W, Liam
- Funding@CVL: VR, SSF, EU

Questions?