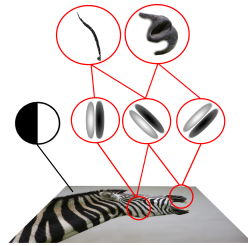


The Component Scale Mixture model of images

- The visual system is representing a hierarchical **generative model** of the environment.
- V1 simple cell responses are organised by **latent variables** representing higher-order statistics of sensory input.
- The latent structure determining covariance structure of V1 cells corresponds to **Gestalt principles**.
- Full **Bayesian inference** is assumed in the model, posteriors are represented by stochastic **samples**.



$$p(v | g) = \mathcal{N}(v; 0, \sum_{j=1}^K g_j C_j)$$

$$p(x | v, z) = \mathcal{N}(x; zAv, \sigma_x I)$$

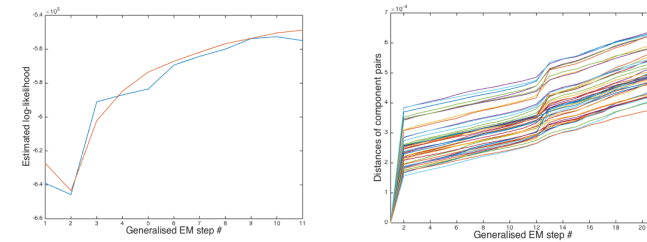
Learning the components

- Generalised EM scheme with gradient ascent
- Averaging over posterior samples in the E-step

$$C_v = \sum_{k=1}^K g_k U_k^T U_k \quad [U_k]_{i,j}^{new} = [U_k]_{i,j}^{old} + \epsilon \frac{\partial \mathcal{L}}{\partial [U_k]_{i,j}}$$

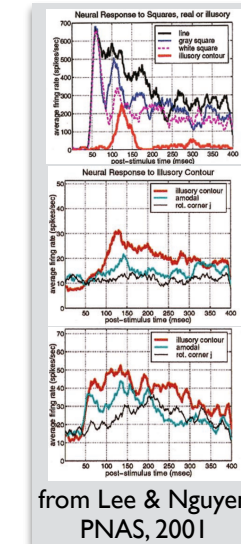
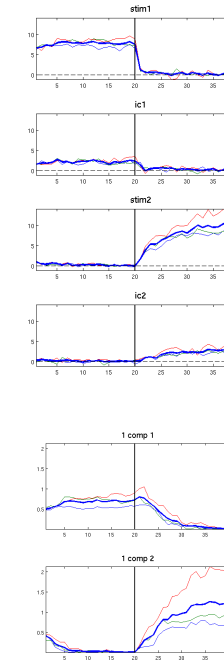
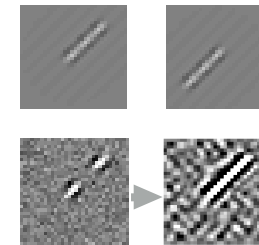
$$\frac{\partial \mathcal{L}}{\partial [U_k]_{i,j}} = \sum_{l=1}^{NL} \text{Tr} \left[\frac{\partial \log p(x^l, v^l, g^l | U_{1...K})}{\partial C_v^l} \frac{\partial C_v^l}{\partial [U_k]_{i,j}} \right]$$

- Increasing log-likelihood and separation of components



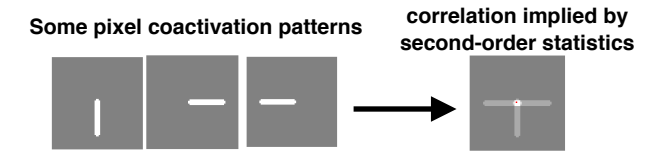
Predicted response to illusory contours

- IC responses are elicited by top-down effects of covariance component activations
- Temporal ordering of activation in latent layers are reproduced by sampling the posterior
- Measured firing rate ratios are reproduced by the model

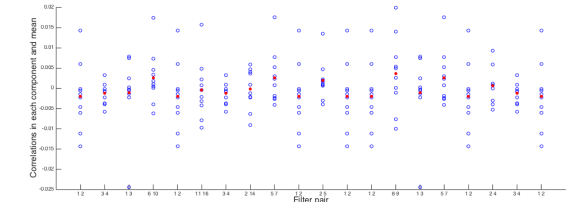


Relation to GSM

- Wainwright & Simoncelli, NIPS, 2000
- GSM describes the second-order statistics of coefficients
- Selective activation of correlation patterns is only possible with latent variables switching between them
- Learnability of context-dependent correlations also relies on a latent layer preventing the correlations to average out



Dispersion of learned correlations around the mean over all components



Sampling the posterior

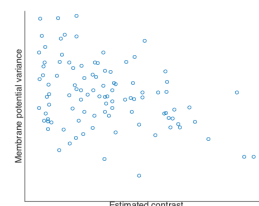
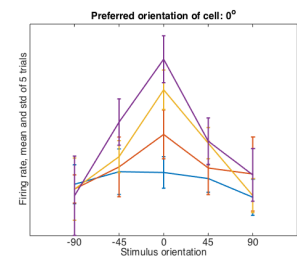
- Gibbs sampling
- Hamiltonian sampling

$$p(v | x, g, z) = \mathcal{N}(v; \frac{z}{\sigma_x} C_{cp} A^T x, C_{cp}), \quad C_{cp} = \left[\frac{z^2}{\sigma_x} A^T A + \sum_{j=1}^K g_j C_j \right]^{-1}$$

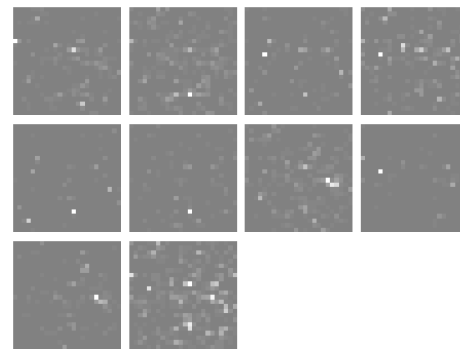
$$\log p(g | x, v, z) \sim -\frac{1}{2} \left[\log \left(\det \left(\sum_{k=1}^K g_k C_k \right) \right) + v^T \left(\sum_{k=1}^K g_k C_k \right)^{-1} v \right] + \log p(g)$$

$$\log p(z | x, v, g) \sim -\frac{1}{2} \left[D_x \log(\sigma_x) + \frac{1}{\sigma_x} (x - zAv)^T (x - zAv) \right] + \log p(z)$$

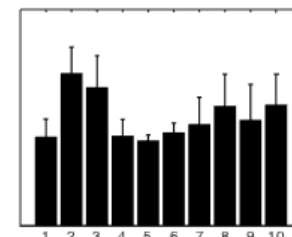
- Contrast dependence of response statistics



Correlations implied by natural statistics

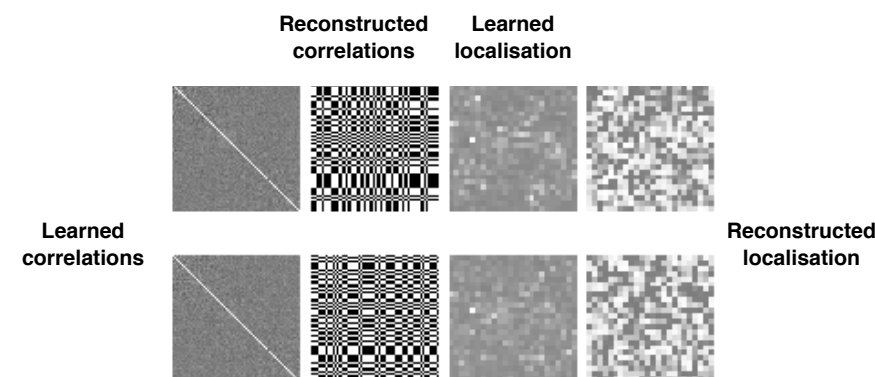


- Vanhateren database, whitened patches of 24x24 pixels
- Localisation of the correlation structure in the pixel domain
- Selectivity of the latent components to specific image patterns



Relation to component models

- Karklin & Lewicki, Nature, 2008, similar model structure
- In CSM we do full posterior sampling instead of giving a point estimate of the latent values, and explicitly represent contrast similarly to GSM
- Parametrisation allows the independent learning of variances and correlations in components
- If we reconstruct correlations from learned variances assuming K&L type parametrisation, we obtain different correlations than through learning them explicitly



Conclusions

- Contextual effects on perception are formalised in a generative model of images
- Sampling from the full posterior enables predictions about variance and covariance
- The model gives predictions for noise correlations between V1 simple cells when fitted to natural image statistics
- The model predicts V1 responses to illusory contours
- CSM generalises GSM and previous component-based image models

Acknowledgement

This work was supported by an MTA Lendület Fellowship