Probabilities on Graphs: Directed Graphs II

Alan Yuille and Dan Kersten

May 6, 2015

Examples of strong coupling

We now give two examples of strong coupling. The first example deals with coupling different modalities, while the second example concerns the perception of texture.

Multisensory cue coupling

- Human observers are sensitive to both visual and auditory cues.
- Sometimes these cues have a common cause, e.g., you see a barking dog. But in other situations, the auditory and visual cues have different causes, e.g., a nearby cat moves and a dog barks in the distance.
- Ventriloquists are able to make the audience think that a puppet is talking by making it seem that visual cues (the movement of the puppet's head) and auditory cues (words spoken by the ventriloquist) are related. The ventriloquism effect occurs when visual and auditory cues have different causes – and so are in conflict – but the audience perceives them as having the same cause.

Multisensory cue coupling: The model (I)

- We describe an ideal observer for determining whether two cues have a common cause or not (Kording et al., 2007), which gives a good fit to experimental findings.
- The model is formulated using a meta-variable C, where $C = 1$ means that the cues x_A, x_V are coupled.
- More precisely, they are generated by the same process S by a distribution $P(x_A, x_V|S) = P(x_A|S)P(x_V|S)$.
  $P(x_A|S)$ and $P(x_V|S)$ are normal distributions $N(x_A|S, \sigma_A^2)$, $N(x_V|S, \sigma_V^2)$ – with the same mean S and variances σ_A^2, σ_V^2.
- It is assumed that the visual cues are more precise than the auditory cues, so that $\sigma_A^2 > \sigma_V^2$. The true position S is drawn from a probability distribution $P(S)$, which is assumed to be a normal distribution $N(0, \sigma_p^2)$.

Multisensory cue coupling: The model (II)

- $C = 2$ means that the cues are generated by two different processes S_A and S_4.
- In this case, the cues x_A and x_V are generated respectively by $P(x_A|S_A)$ and $P(x_V|S_V)$, which are both Gaussian $N(S_A, \sigma_A^2)$ and $N(S_V, \sigma_V^2)$. We assume that S_A and S_V are independent samples from the normal distribution $N(0, \sigma_p^2)$.
- Note that this model involves model selection, between $C = 1$ and $C = 2$, and so, in vision terminology, it is a form of strong coupling with model selection (Yuille & Bulthoff, 1996).

Multisensory cue coupling: Illustration

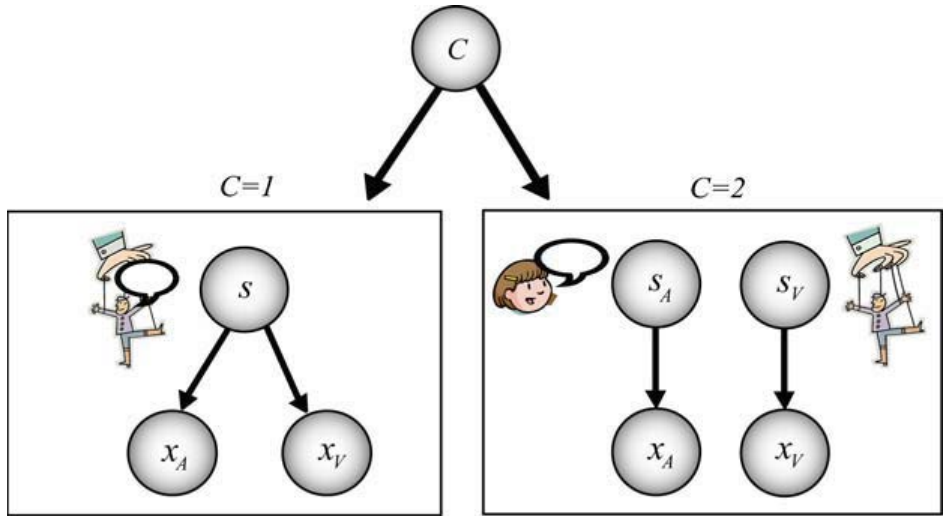

Figure 1: The subject is asked to estimate the position of the cues and to judge whether the cues are from a common cause – i.e., at the same location – or not. In Bayesian terms, the task of judging whether the cause is common can be formulated as model selection: are the auditory and visual cues more likely generated by a single cause (left) or by two independent causes (right)? Figure adapted from Kording et al. (2007).

Multisensory cue coupling: Comparison with experiments (I)

- This model was compared to experiments in which brief auditory and visual stimuli were presented simultaneously, with varying amounts of spatial disparity.
- Subjects were asked to identify the spatial location of the cue and/or whether they perceived a common cause (Wallace et al., 2004).
- The closer the visual stimulus was to the audio stimulus, the more likely subjects would perceive a common cause.
- In this case subjects' estimate of the stimuli's position was strongly biased by the visual stimulus (because it is considered more precise with $\sigma_V^2 > \sigma_A^2$).
- But if subjects perceived distinct causes, then their estimate was pushed away from the visual stimulus, and exhibited *negative bias*.

Multisensory cue coupling: Comparison with experiments (II)

- Körding et al. (2007) argue that this negative bias is a selection bias stemming from restricting to trials in which causes are perceived as being distinct.
- For example, if the auditory stimulus is at the center and the visual stimulus at 5 degrees to right of center, then sometimes the (very noisy) auditory cue will be close to the visual cue and hence judged to have a common cause, while in other cases, the auditory cause is farther away (more than 5 degrees).
- Hence the auditory cue will have a truncated Gaussian (if judged to be distinct) and will yield negative bias.

Multisensory cue coupling: Results and figure

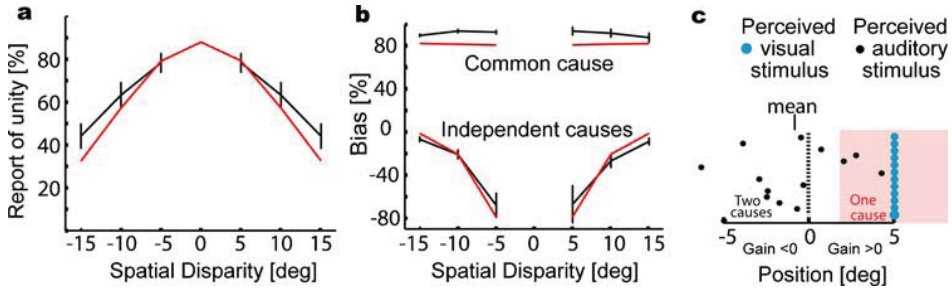

Figure 2: Reports of causal inference. (a) The relative frequency of subjects reporting one cause (black) is shown, with the prediction of the causal inference model (red). (b) The bias, i.e., the influence of vision on the perceivedauditory position, is shown (gray and black). The predictions of the model are shown in red. (c) A schematic illustration explaining the finding of negative biases. Blue and black dots represent the perceived visual and auditory stimuli, respectively. In the pink area, people perceive a common cause. Reprinted with permission from Kording et al. (2007)

Multisensory cue coupling: The mathematics (I)

More formally, the beliefs $P(C|x_A, x_V)$ in these two hypotheses $C = 1, 2$ are obtained by summing out the estimated positions s_A, s_4 of the two cues as follows:

$$P(C|x_A, x_V) = \frac{P(x_A, x_V|C)P(C)}{P(x_A, x_V)}$$

$$= \frac{\int dS P(x_A|S)P(x_V|S)P(S)}{P(x_A, x_V)}, \quad \text{if } C = 1,$$

$$= \frac{\int\int dS_A dS_V P(x_A|S_A)P(x_V|S_V)P(S_A)P(S_V)}{P(x_A, x_V)}, \quad \text{if } C = 2.$$

Multisensory cue coupling: The mathematics (II)

- There are two ways to combine the cues. The first is model selection. This estimates the most probable model $C^* = \arg\max P(C|x_V, x_A)$ from the input x_A, x_V and then uses this model to estimate the most likely positions s_A, s_V of the cues from the posterior distribution:

$$P(s_V, s_A) \approx P(s_V, s_A|x_V, x_A, C^*) = \frac{P(x_V, x_A|s_V, s_A, C^*)P(s_V, s_A|C^*)}{P(x_V, x_A|C^*)}.$$

- The second way to combine the cues is by *model averaging*. This does not commit itself to choosing C^* but instead averages over both models:

$$P(s_V, s_A|x_V, x_A) = \sum_C P(s_V, s_A|x_V, x_A, C)P(C|x_V, x_A)$$

$$= \sum_C \frac{P(x_V, x_A|s_V, s_A, C)P(s_V, s_A|C)P(C|x_V, x_A)}{P(x_V, x_A|C)},$$

where $P(C = 1|x_V, x_A) = \pi_C$ (the posterior mixing proportion).

Multisensory cue coupling: Extension

- Natarajan et al. (2008) showed that a variant of the model could fit the experiments even better.
- They replaced the Gaussian distributions with alternative distributions that are less sensitive to rare events. Gaussian distributions are non-robust because the tails of their distributions fall off rapidly, which gives very low probability to rare events.
- More precisely Natarajan et al. (2008) assumed that the data is distributed by a mixture of a Gaussian distribution, as above, and a uniform distribution (yielding longer tails).
- More formally, they assume $x_A \ \pi N(x_A : s_A, \sigma_A^2) + \frac{(1-\pi)}{r_l}$ and $x_V \ \pi N(x_V : s_V, \sigma_V^2) + \frac{(1-\pi)}{r_l}$, where π is a mixing proportion, and $U(x) = 1/r_1$ is a uniform distribution defined over the range r_1.

Homogeneous and isotropic texture

- The second example is by Knill and concerns the estimating of orientation in depth (slant) from texture cues (Knill, 2003).
- There are alternative models for generating the image, and the human observer must infer which is most likely. In this example, the data could be generated by isotropic homogeneous texture or by homogeneous texture only.
- Knill's finding is that human vision is biased to interpret image texture as isotropic, but if enough data are available, the system turns off the isotropy assumption and interprets texture using the homogeneity assumption only.

Homogeneous and isotropic texture: Illustration

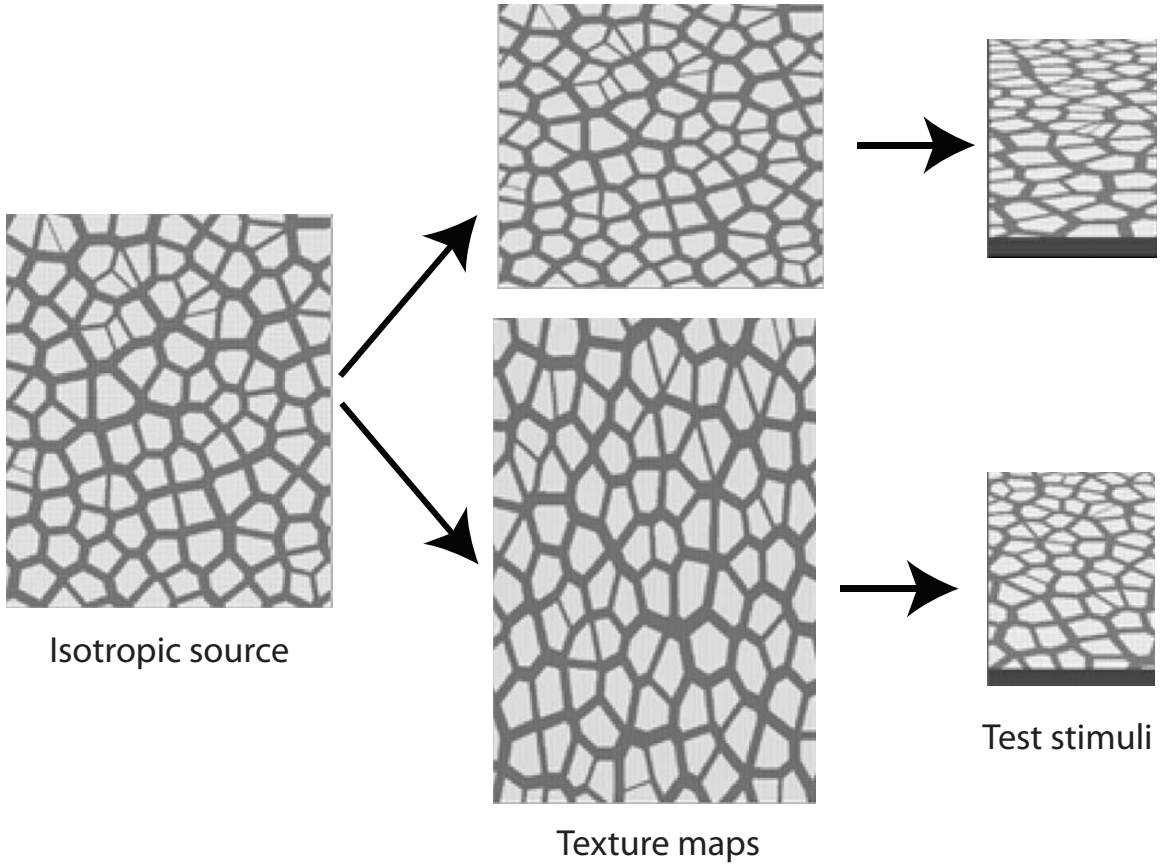

Figure 3: Generating textures that violate isotropy. An isotropic source image is either stretched (top middle) or compressed (bottom middle), producing texture maps that get applied to slanted surfaces shown on the right. A person that assumes surface textures are isotropic would overestimate the slant of the top stimulus and underestimate the slant of the bottom one. Figure adapted from from Knill (2003).

Homogeneous and isotropic texture: Theory (I)

- The posterior probability distribution for S is given by:

$$P(S|I) = \frac{P(I|S)P(S)}{P(I)}, \quad P(I|S) = \sum_{i=1}^{n} \phi_i P_i(I|S),$$

where ϕ_i is prior probability of model i, and $p_i(I|S)$ is corresponding likelihood function.

- More specifically, texture features T can be generated by either an isotropic surface or a homogeneous surface. The surface is parameterized by tilt and slant σ, τ. Homogenous texture is described by two parameters α, θ, and isotropic texture is a special case where $\alpha = 1$. This gives two likelihood models for generating the data:

$$P_h(T|(\sigma,\tau),\alpha,\theta), \quad P_i(T|(\sigma,\tau),\theta)$$

Here, $P_i(T|(\sigma,\tau),\theta) = P_h(T|(\sigma,\tau),\alpha = 1,\theta)$.

Homogeneous and isotropic texture: Theory (II)

- Isotropic textures are a special case of homogenous textures.
- The homogeneous model has more free parameters and hence has more flexibility to fit the data, which suggests that human observers should always prefer it. But the Occam factor (MacKay, 2003) means that this advantage will disappear if we put priors $P(\alpha)P(\theta)$ on the model parameters and integrate them out. This gives:

$$P_h(T|(\sigma,\tau)) = \int\int d\alpha d\theta P_h(T|(\sigma,\tau),\alpha,\theta),$$

$$P_i(T|(\sigma,\tau)) = \int d\theta P_h(T|(\sigma,\tau),\theta).$$

- Integrating over the model priors smooths out the models. The more flexible model, P_h, has only a fixed amount of probability to cover a large range of data (e.g., all homogeneous textures) and hence has lower probability for any specific data (e.g., isotropic textures).

Homogeneous and isotropic texture: The mathematics

- Knill describes how to combine these models using model averaging. The combined likelihood function is obtained by taking a weighted average:

$$P(T|(\sigma,\tau)) = p_h P_h(T|(\sigma,\tau)) + p_i P_i(T|(\sigma,\tau)), \tag{1}$$

where (p_h, p_i) are prior probabilities that the texture is homogeneous or isotropic. We use a prior $P(\sigma,\tau)$ on the surface and finally achieve a posterior:

$$P(\sigma,\tau|I) = \frac{P(I|(\sigma,\tau))P(\sigma,\tau)}{P(I)}. \tag{2}$$

- This model has a rich interpretation. If the data are consistent with an isotropic texture, then this model dominates the likelihood and strongly influences the perception. Alternatively, if the data are consistent only with homogeneous texture, then this model dominates. This gives a good fit to human performance (Knill, 2003).

Motion and time

- The perception of motion can be strongly influenced by its history and not merely by the change of image from frame to frame. For example, Anstis and Ramachandran(1987) demonstrated perceptual phenomena where motion perception seems to require a temporal coherence prior in addition to the slow and smoothness priors described earlier in this section. Similarly, Watamaniuk et al. (1995) demonstrated that humans could detect a coherently moving dot despite the presence of many incoherently moving dots.
- These classes of phenomena can be addressed by models that make prior assumptions about how motion changes over time. These can be performed (Yuille et al., 1998) by adapting the Bayes-Kalman filter (Kalman, 1960; Ho & Lee, 1964) filter which gives an optimal way to combine information over time.

Bayes-Kalman filter (I)

- The task of the Bayes-Kalman filter is to estimate the state x_t of a system at time t dependent on a set of observations $y_t, ..., y_1$ (e.g., x_t could be the position of an airplane and y_t a noisy measurement of the airplane's position at time t). The model assumes a probability distribution $P(x_{t+1}|x_y)$ for how the state changes over time and a likelihood function $P(y_t|x_t)$ for the observation.
- The task is to estimate the state x_t of a system at time t dependent on a set of observations $y_t, ..., y_1$ (e.g., x_t could be the position of an object and y_t a noisy measurement of the object position at time t). The model assumes a probability distribution $P(x_{t+1}|x_y)$ for how the state changes over time and a likelihood function $P(y_t|x_t)$ for the observation. This can be formulated by a Markov model, where the observations $y_t, ..., y_1$ and states $x_t, ..., x_1$ are represented by the blue and red dots, respectively (the lower and upper dots if viewed in black and white).

Bayes-Kalman filter (II)

- The purpose of Bayes-Kalman is to estimate the distribution $P(x_t|Y_t)$ of the state x_t conditioned on the measurements $Y_t = \{y_t, ..., y_1\}$ up to time t. It performs this by repeatedly performing the following two steps, which are called prediction and correction. The prediction uses the prior $P(x_{t+1}|x_t)$ to predict distribution $P(x_{t+1}|Y_t)$ of the state at $t+1$:

$$P(x_{t+1}|Y_t) = \int dx_t P(x_{t+1}|x_t)P(x_t|Y_t). \tag{3}$$

- The correction step integrates the new observation y_{t+1} to estimate $P(x_{t+1}|Y_{t+1})$ by:

$$P(x_{t+1}|Y_{t+1}) = \frac{P(y_{t+1}|x_{t+1})P(x_{t+1}|Y_t)}{P(y_{t+1}|Y_t)}. \tag{4}$$

- Bayes-Kalman is initialized by setting $P(x_1|y_1) = P(y_1|x_1)P(x_1)/P(y_1)$ where $P(x_1)$ is the prior for the original position of the object at the start of the sequence. Then equations (3, 4) are run repeatedly. The effect of prediction is to introduce uncertainly about the state x_t, while correction reduces uncertainty by providing a new measurement.

Bayes-Kalman filter: Figures

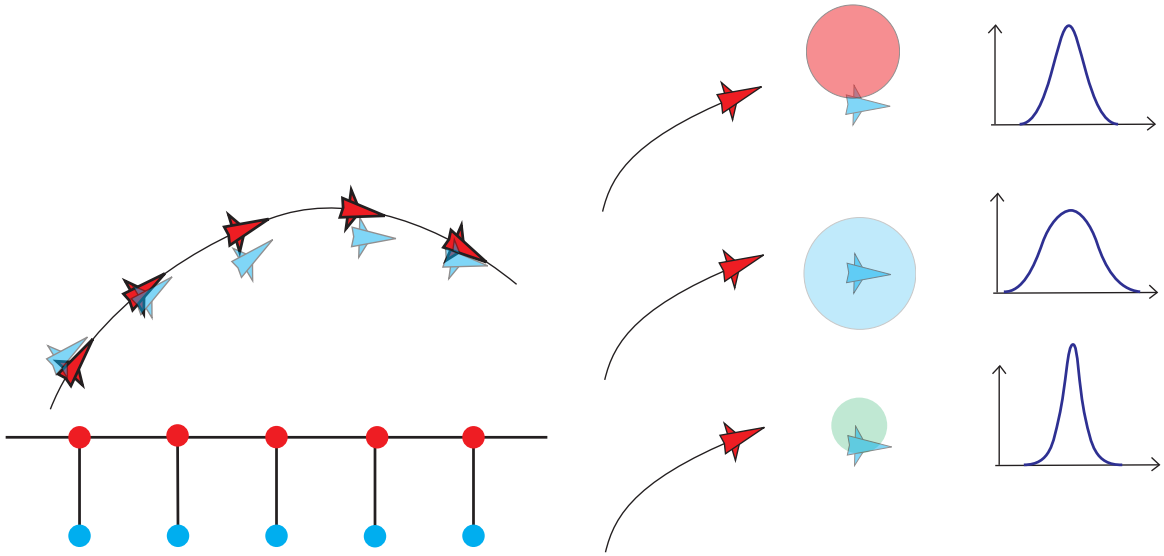

Figure 4: Left: Graph illustrating the unobserved states (red) and the observed states (blue) as a function of time. The airplanes true positions are shown in red, and their observations (biased) are shown in blue. The Bayes-Kalman filter integrates observations to make estimate the true state using prior probabilities. Right: Bayes-Kalman updates a probability distribution for the estimated position of the target. The variance of the distribution is illustrated by the one-dimensional figure (on the right) and the size of the circle (red, blue, or green). In the prediction stage (middle) the variance becomes large, and after the measurement, the variance becomes smaller.

Bayes-Kalman filter: Special Case

- But this is an important special case where Bayes-Kalman can be estimated by algebraic equations (Kalman 1960). This occurs if the prior $P(x_1)$, the distribution $P(x_{t+1}|x_t)$, and the observation model $P(y_t|x_t)$ are all Gaussian models. Then it follows that $P(x_t|Y_t), P(x_{t+1}|Y_t)$ and $P(x_{t+1}|Y_{t+1})$ are all Gaussian distributions.
- $$P(y_t|x_t) = \mathcal{N}(x_t, \sigma_m^2),\ P(x_{t+1}|x_t) = \mathcal{N}(x_t + \mu, \sigma_p^2),\ P(x_1) = \mathcal{N}(\mu_1, \sigma_1^2) \tag{5}$$

  . Here σ_m^2 is the variance of the observation, μ is the mean distance traveled by the object from t to $t+1$ with variance σ_p^2, and μ_1 is the mean position of the object at time $t = 1$ with variance σ_1^2.
- Suppose the distribution $P(x_t|Y_t) = \mathcal{N}(\mu_t, \sigma_t^2)$. Then we can use equation (5) to re-express the prediction and correction update equations (3,4) as:

$$P(x_{t+1}|Y_t) = \mathcal{N}(\mu + \mu_t, \sigma_p^2 + \sigma_t^2),\ P(x_{t+1}|Y_{t+1}) = \mathcal{N}(\mu_{t+1}, \sigma_{t+1}^2), \tag{6}$$

$$\mu_{t+1} = \mu + \mu_t - \frac{(\sigma_p^2 + \sigma_t^2)\{(\mu + \mu_t) - y_{t+1}\}}{\sigma_m^2 + (\sigma_p^2 + \sigma_t^2)},$$

$$\sigma_{t+1}^2 = \frac{\sigma_m^2(\sigma_p^2 + \sigma_t^2)}{\sigma_m^2 + (\sigma_p^2 + \sigma_t^2)}.) \tag{7}$$

Bayes-Kalman filter: Special Cases

- Observe that if the object is at the mean predicted position – i.e. $y_{t+1} = \mu + \mu_{t+1}$ - then the prediction part disappears. Also note that the Kalman update combines the different sources of information – the observation y_{t+1}, the mean estimated positions $\mu + \mu_{t+1}$ by a linear weighted average similar to that used for coupling cues by linear weighted averaging (previous chapter).
- You can get better understanding of the Kalman filter by considering special cases.
- If the observations are noiseless – $\sigma_m = 0$ – then it follows that $\mu_{t+1} = y_{t=1}$, so we should forget the history and just use the current observation as our estimate of x_{t+1}. If $\sigma_p^2 = 0$ then we have perfect prediction and so $\mu_{t+1} = \frac{\sigma_t^2}{\sigma_m^2+\sigma_t^2} y_{t+1} + \frac{\sigma_m^2(\mu+\mu_{t+1})}{\sigma_m^2+\sigma_t^2}$ with $\sigma_{t+1}^2 = \frac{\sigma_m^2\sigma_t^2}{\sigma_m^2+\sigma_t^2}$, which corresponds to taking the weighted average of y_{t+1} with $\sigma^2 + \sigma_t^2$.
- If we also require that $\mu = 0$ (i.e. the object does not move) then we obtain $\mu_{t+1} = \frac{\sigma_t^2}{\sigma_m^2+\sigma_t^2} y_{t+1} + \frac{\sigma_m^2(\mu_{t+1})}{\sigma_m^2+\sigma_t^2}$, which is simply an online method for computing the MAP estimate of a static object at position x (as described in the first paragraph).

Bayes-Kalman filter: Extension to Actions

- Bayes-Kalman is a directed graphical models for integrating information over time and is part of the BDT framework. It can be used for visual tasks like tracking objects over time.
- Another directed graphical model for modeling phenomena over time are Hidden Markov Models (HMMs) developed for speech processing, They have been used for visual tasks like detecting types of tennis strokes, or analyzing baseball plays.
- BDT can be extended to Control theory where the system makes a series of decisions over time. This was used by Dickmanns's group for designing automated cars that could drive 1,000 kilometers on an autobahn in 1990. The control ensures that the cars stay in lane, avoid other cars, and can drive from A to B,
- World models, described later in the course enable these ideas to be extended by putting probability distributions on generating images conditioned on previous images, immagining, and taking action. See GeneX: https://generative-world-explorer.github.io/

Divisive normalization: Probabilistic Graphical Model

- An important example is the use of probabilistic models (Wainwright & Simoncelli, 2000) to account for divisive normalization. This is a mechanism whereby cells mutually inhibit one another, effectively normalizing their responses with respect to stimulus inputs. Originally developed to explain nonlinear responses to contrast in V1 (Heeger, 1992), divisive normalization has been proposed as a basic cortical computation that underlies various effects of context, as well as higher-level processe,s such as attention (Carandini & Heeger, 2011) .
- The probabilistic approach gives a theoretical justification for divisive normalization in V1. The main idea is that filters with similar preferences for orientation representing nearby spatial locations in a scene have striking statistical dependencies, which can be removed by divisive normalization. Specifically, if we plot the statistics of two linear filters f_c, f_s (center and surround), then the magnitudes of f_c, f_s are coordinated in a straightforward way, which has a characteristic shape of a bow tie.

Modeling divisive normalization using hidden variables

This can be modeled by assuming there are hidden variables ν that affect both responses and hence induces correlation between the responses. For example, ν could represent the local average image intensity, which could affect the response of both filters, but after the filter response, it could be made independent by conditioning on the average intensity. Suppose ν has a prior distribution $P(\nu) = \nu \exp\{-\nu^2/2\}$ for $\nu \geq 0$. We have a pair of filters $\{I_i : i = 1, 2\}$ that are related to Gaussian models $\{g_1 : i = 1, 2\}$. Then we can model the activation of the set of filter responses:

$$P(I_1, I_2) = \int d\nu P(\nu) \prod_{i=1}^{2} P(I_i|\nu, g_i))P(g_i), \tag{8}$$

where $P(I_i|\nu, g_i) = \delta(I_i - \nu g_i)$. In this model the filter responses are generated by independent processes, g_1, g_2, but then are multiplied by the common factor ν. This is illustrated in the next figure.

Figure for divisive normalization model

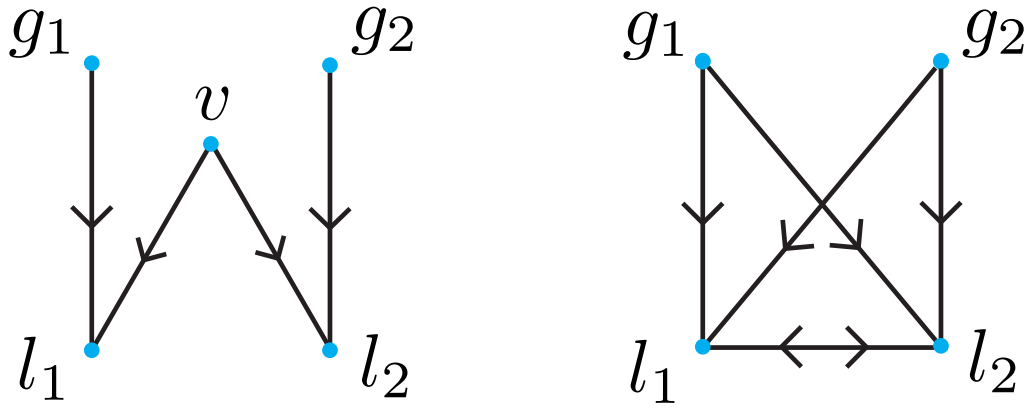

Figure 5: Left: The graphical structure of the divisive normalization model. The filter responses l_1, l_2 are generated from stimuli g_1, g_2 and by the common factor ν. The distributions of l_1, l_2 are factorized if we condition on ν. Right: But if we integrate out ν, then almost all the variables become dependent, as reflected by the complexity of the graph structure.

Divisive normalization model

- In particular, for each filter we can compute $P(g_i|I_1, I_2)$. After some algebra, this is computed to be:

$$P(g_1|I_1, I_2) = \frac{g_1^{-1} \exp\{-\frac{g_1^2 I^2}{2\sigma^2 I_1^2} - \frac{I_1^2}{2g_1^2}\}}{B(0, I/\sigma)}, \tag{9}$$

  where $I = \sqrt{I_1^2 + I_2^2}$, and $B(.,.)$ is a Bessel function. To get intuition, note that $g_1 = I_1/\nu$ and $g_1 = I_1/\nu$. So if ν is small, then $|I_1|$ and $|I_2|$ are likely to be small together, while if ν is large, then $|I_1|$ and $|I_2|$ are both likely to be large.

- Assume that the goal of a model unit is to estimate the g_i from the observed filter responses $\{I_i : i = 1, 2\}$, which gives the nonlinear response of the cell. It follows, from analysis above, that

$$E(g_1|I_1, I_2) \propto \text{sign}\{I_1\}\sqrt{|I_1|}\sqrt{\frac{|I_1|}{\sqrt{I_1^2 + I_2^2 + k}}}. \tag{10}$$

  The $\sqrt{I_1^2 + I_2^2 + k}$ term sets the gain and performs the divisive normalization.

Application to the tilt illusion

- The model has also been applied to explain the classic tilt illusion in perception (Schwartz et al., 2009; Qiu et al., 2013). In the "simultaneous" tilt illusion, a set of vertically oriented lines appears to tilt right when surrounded by an annulus of lines tilted left–an effect called "repulsion." But for large differences between the center orientation and the surround (tilted left), the center vertical lines can appear to tilt left–an effect called "attraction." In the model, the population of neurons responding to the surround tilted lines contributes to divisive normalizing of the neurons responding to the center stimulus. This results in a change of their neural tuning curves, which, together with the degree of coupling between center and surrounds, accounts for repulsion and attraction.
- The suppressive effect of surround contrast on a central region is an example of local spatial context.

The tilt illusion

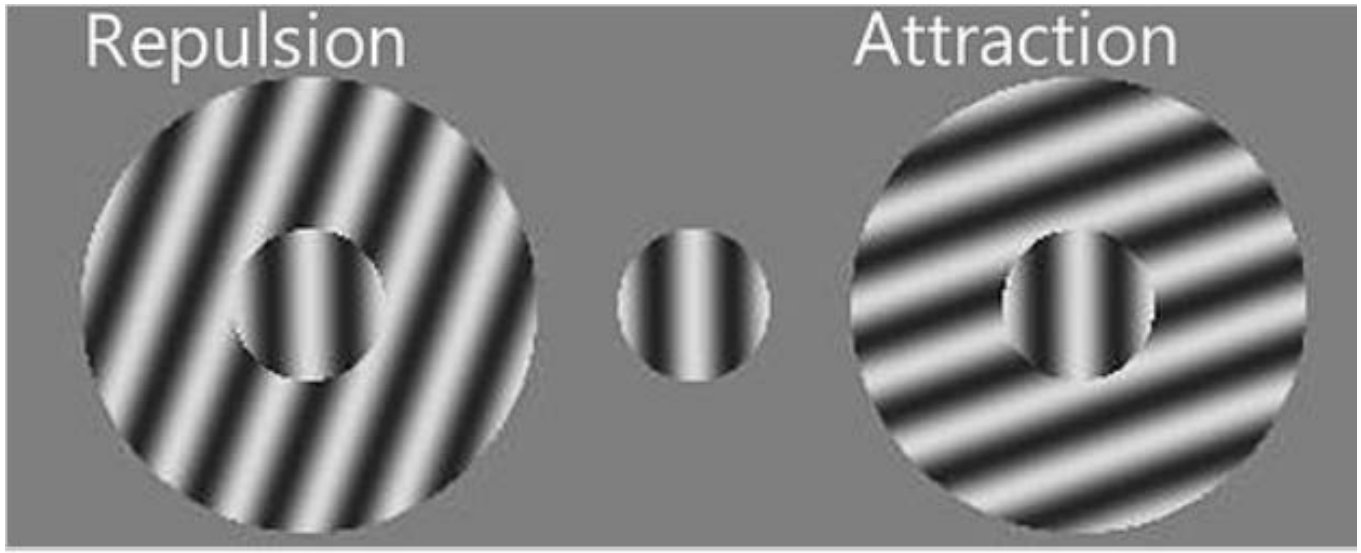

Figure 6: The perceived orientation of a grating pattern can appear to be tilted away from its true orientation due to the presence of surrounding gratings with different orientations. The central circular grating (Center Panel) appear to be tilted to the left (Left Panel) because it is *repulsed* from the orientation of the larger background grating (because the relative orientation is greater than 0 but less than 50 degrees). Conversely it is tilted slightly to the right (Right panel) when it is *attracted* to the background grating (where relative orientation is between 50 and 90 degrees).

Context and spatial interactions between neurons

- There is considerable evidence that low-level vision involves long-range spatial interaction,s so that human perception of local regions of an image can be strongly influenced by their spatial context. Psychophysicists have discovered many perceptual phenomena demonstrating spatial interactions.
- For example, local image regions that differ from their neighbors tend to "pop out" and attract attention, while, conversely, similar image features that form spatially smooth structures tend to get "grouped" together to form a coherent percept, see chapter figure 12.26 (left panel). Image properties such as color tend to spread out, or fill in regions, until they hit a boundary (Grossberg & Mingolla, 1985; Sasaki et al., 2004) as shown in chapter figure 12.26 (right panel).