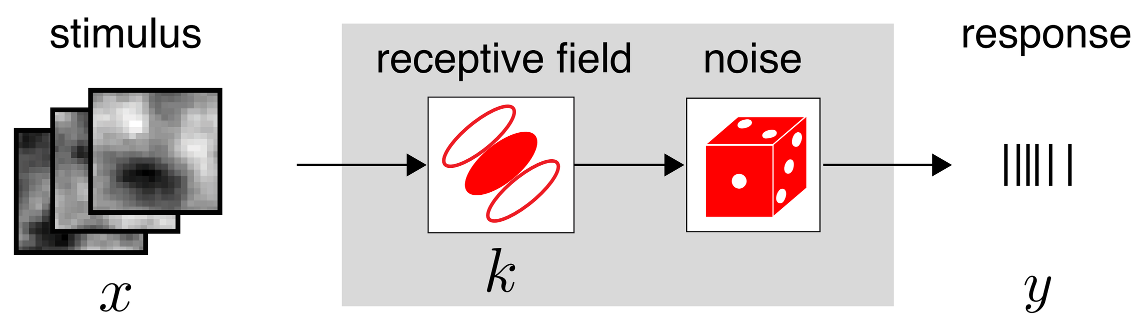
Empirical Bayes Methods for Sparse, Smooth, Localized Receptive Field Estimation

Mijung Park - Electrical and Computer Engineering , Jonathan W. Pillow - Center for Perceptual Systems, Departments of Psychology & Neurobiology, The University of Texas at Austin

COSYNE 2011
#III-88

1. Neural characterization problem



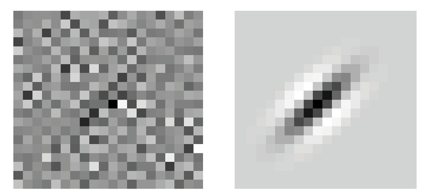
Goal: characterize the receptive field (RF) using neural responses to white noise or naturalistic stimuli

Problem: standard estimators are noisy, require lots of data

Maximum Likelihood Esimator

$$\mathbf{k}_{ML} = (\mathbf{X}^T\mathbf{X})^{-1}\underbrace{\mathbf{X}^T\mathbf{y}}_{\text{spike-triggered average}}$$

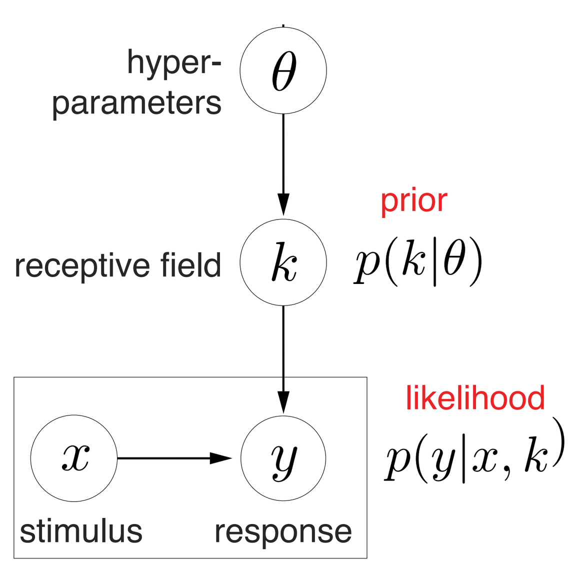
simulated example:



2. Empirical Bayes (EB)

- Use a prior to regularize RF estimate
- Set hyper-parameters governing that prior by maximum likelihood

generative model



2-stage estimation procedure

1. Set θ by maximizing the "evidence"

$$\hat{\theta} = \arg\max_{\theta} P(\mathbf{y}|\mathbf{X},\theta)$$
$$= \arg\max_{\theta} \int P(\mathbf{y}|\mathbf{X},\mathbf{k})P(\mathbf{k}|\theta)d\mathbf{k}$$

2. MAP estimate for $\mathbf{k}$:

$$\arg\max_{\mathbf{k}} \log P(\mathbf{y}|\mathbf{X},\mathbf{k}) + \log P(\mathbf{k}|\theta)$$

Gaussian case: zero-mean Gaussian prior + Gaussian likelihood
$\Longrightarrow$ evidence is easy to compute!

prior $\quad P(\mathbf{k}|\theta) = \mathcal{N}(0, C(\theta))$

posterior covariance $\left(\frac{\mathbf{X}^T\mathbf{X}}{\sigma^2} + C^{-1}\right)^{-1}$

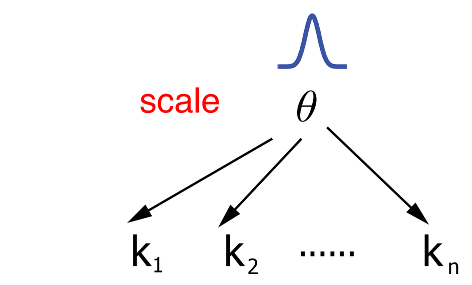
evidence $\quad P(\mathbf{y}|\mathbf{X}, C(\theta), \sigma^2) = \sqrt{\frac{|2\pi\Lambda|}{|2\pi\sigma^2||2\pi C|}} \exp\left[-\frac{1}{2}\mathbf{y}^T\left(\frac{I}{\sigma^2} - \frac{\mathbf{X}\Lambda\mathbf{X}^T}{\sigma^4}\right)\mathbf{y}\right]$

MAP estimate $\quad \mathbf{k}_{MAP} = (\mathbf{X}^T\mathbf{X} + \sigma^2 C^{-1})^{-1}\mathbf{X}^T\mathbf{y}$

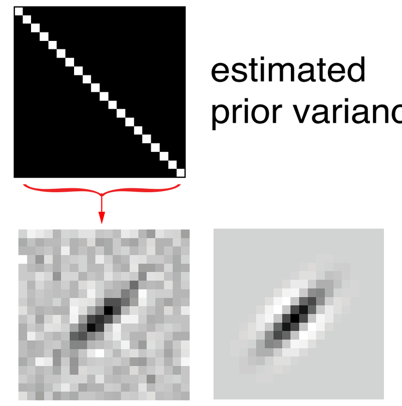
3. Prior methods (using empirical Bayes)

(A) ridge regression

- Gaussian prior over weights with a common variance
- standard regularization technique: "L2 shrinkage"

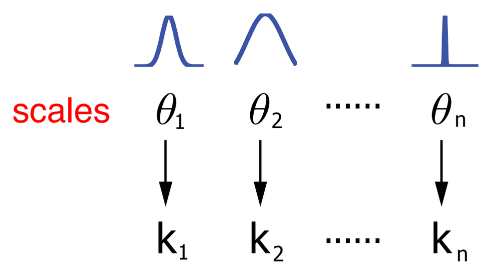


$$C(\theta) = \frac{1}{\theta}I$$

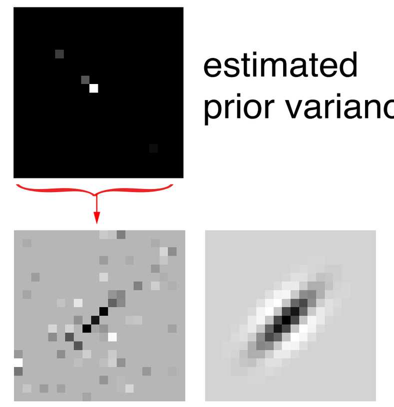


(B) automatic relevance determination (ARD) (Tipping, 2001)

- Gaussian prior with different variance for each weight
- produces sparse k

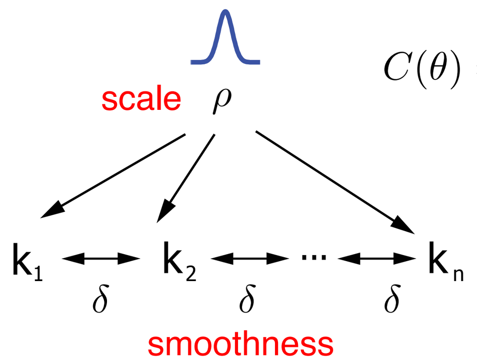


$$C(\theta) = diag(\frac{1}{\theta_i})$$



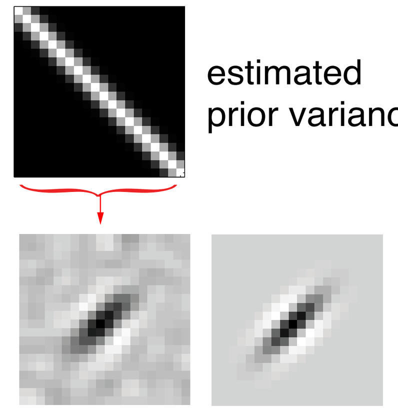
(C) automatic smoothness determination (ASD) (Sahani & Linden, 2002)

- Gaussian prior with distance-dependent correlation
- produces smooth k



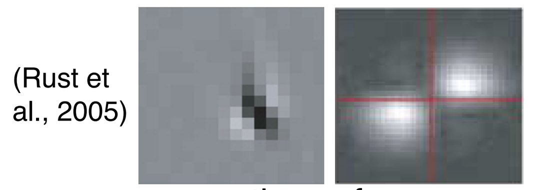
pixel distance

$$C(\theta) = \rho \exp\left(-\frac{(x_i - x_j)^2}{2\delta^2}\right)$$



4. Observation

macaque V1 receptive field



RFs tend to be localized in space-time and spatio-temporal frequency (not *just* sparse or smooth)

idea: design a prior covariance matrix to capture this structure

5. Automatic Locality Determination (ALD)

(A) spacetime-localized prior (ALDs)

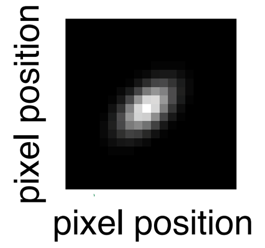
- diagonal prior with location-dependent variance
- allows large weights only within some space-time region

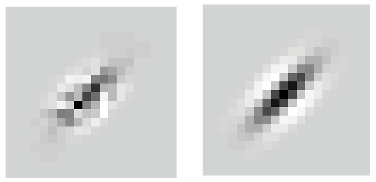
pixel position

$$C_s(\theta) = diag\left[\rho \exp\left(-\frac{1}{2}(x-\mu)^T\Delta^{-1}(x-\mu)\right)\right]$$

scale center extent

estimated prior variances





(B) frequency-localized prior (ALDf)

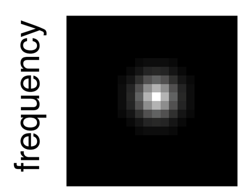
- diagonal prior in Fourier basis with frequency-dependent variance
- allow large weights only within some region of Fourier space

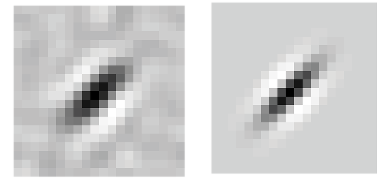
frequency

$$C_f(\theta) = diag\left[\rho \exp\left(-\frac{1}{2}(w-\mu_f)^T\Delta_f^{-1}(w-\mu_f)\right)\right]$$

scale center extent

estimated prior variances (in Fourier basis)



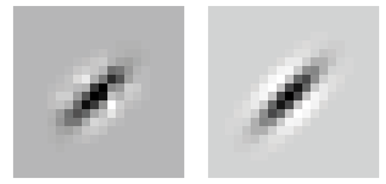


(C) spacetime & frequency-localized prior (ALDsf)

- "sandwich" together ALDs and ALDf prior covariance matrices
- weights localized in spacetime and frequency

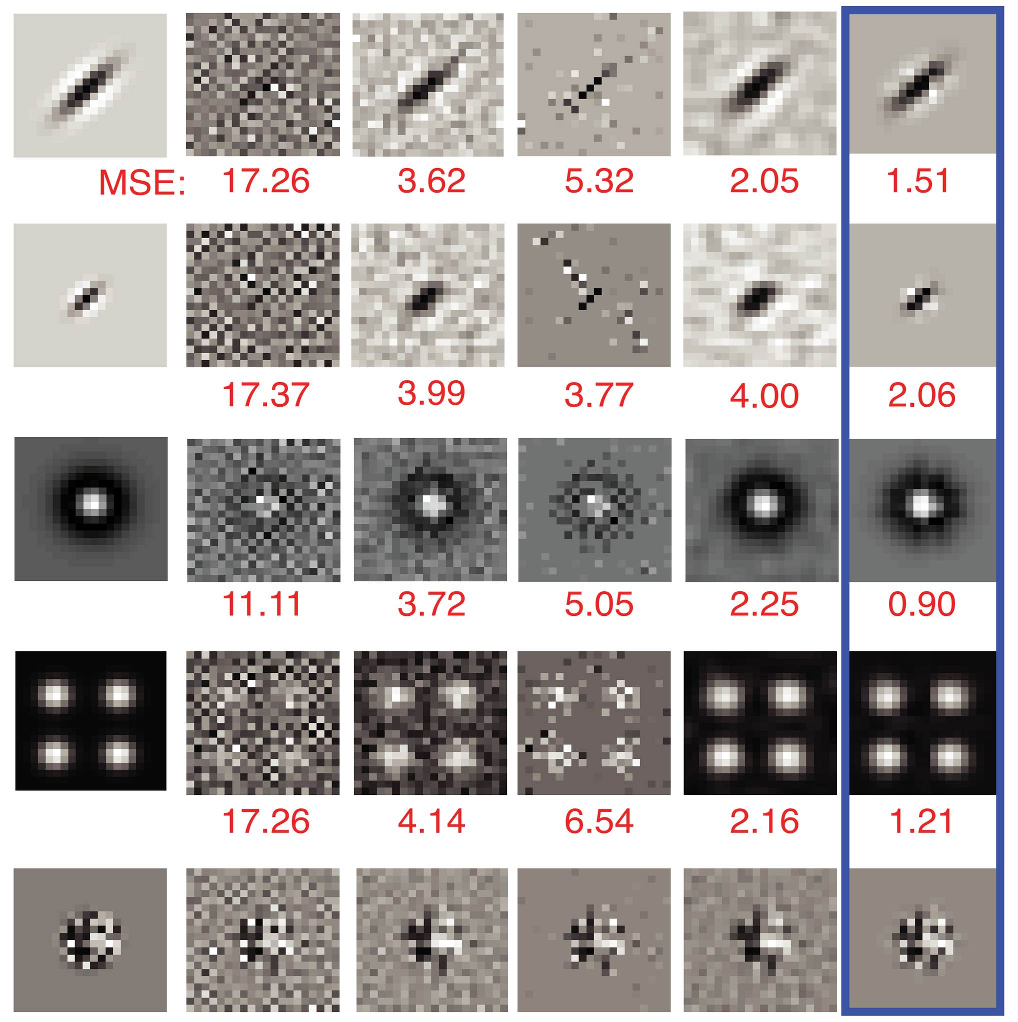
Fourier basis matrix

$$C_{sf}(\theta) = C_s^{\frac{1}{2}} B^T C_f B C_s^{\frac{1}{2}}$$



- RF estimates are sparse in both bases
- tend to be smooth

6. Simulations

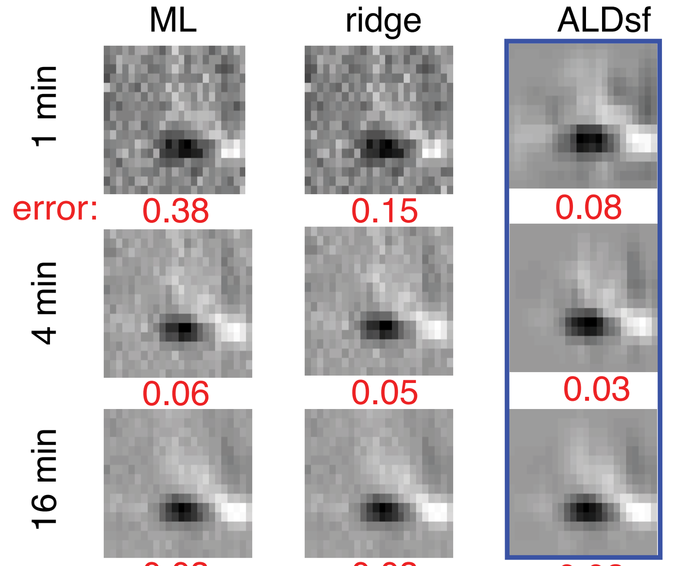


MSE:	ML	ridge	ARD	ASD	ALDsf
	17.26	3.62	5.32	2.05	1.51
	17.37	3.99	3.77	4.00	2.06
	11.11	3.72	5.05	2.25	0.90
	17.26	4.14	6.54	2.16	1.21
	17.29	6.54	6.13	6.55	4.68

7. V1 simple cell data

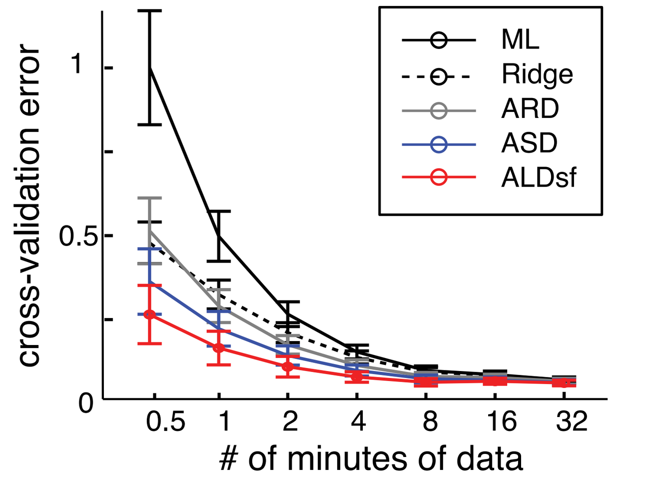
data: Rust et al Neuron 2005

example cell



error:	ML	ridge	ALDsf
1 min	0.38	0.15	0.08
4 min	0.06	0.05	0.03
16 min	0.03	0.03	0.02

average (16 cells)



8. Extension: fully Bayesian inference and error bars

- Empirical Bayes fails to take account of uncertainty in hyper-parameters

data

posterior: $P(\mathbf{k}|D) \approx P(\mathbf{k}|\hat{\theta}_{ML}, D)$

But: true posterior: $P(\mathbf{k}|D) = \int P(\mathbf{k},\theta|D)d\theta$

Algorithm for sampling from true posterior (Markov Chain Monte Carlo)

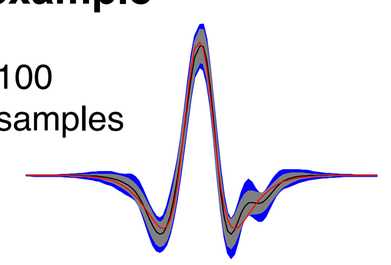
evidence

1. Sample $\theta^* \sim P(\theta|D) \propto P(D|\theta)P(\theta)$ by Metropolis Hastings
2. For each θ^*, sample $\mathbf{k}^* \sim P(\mathbf{k}|D,\theta^*)$

simulated 1-D example

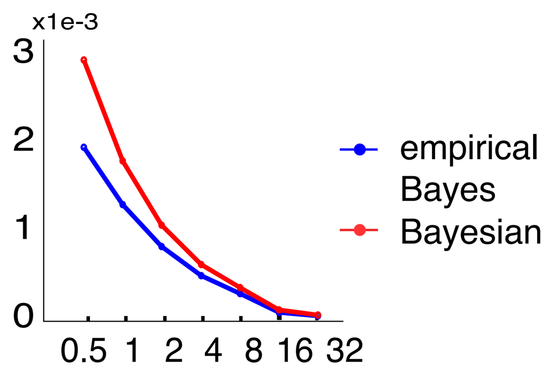
95 % credible intervals
- empirical Bayes
- Bayesian

100 samples

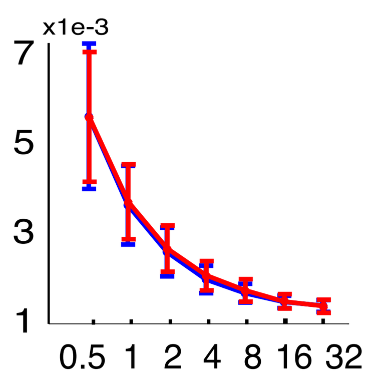


V1 simple cell data

posterior variance



cross-validation error



# of minutes of data

Conclusions

- novel priors capture localized structure of neural RFs
- automatic setting of hyper-params by empirical Bayes
- more accurate RF estimates from less data

Acknowledgements

We thank Nicole Rust & Tony Movshon for neural data. MP and JWP were supported by the Center for Perceptual Systems