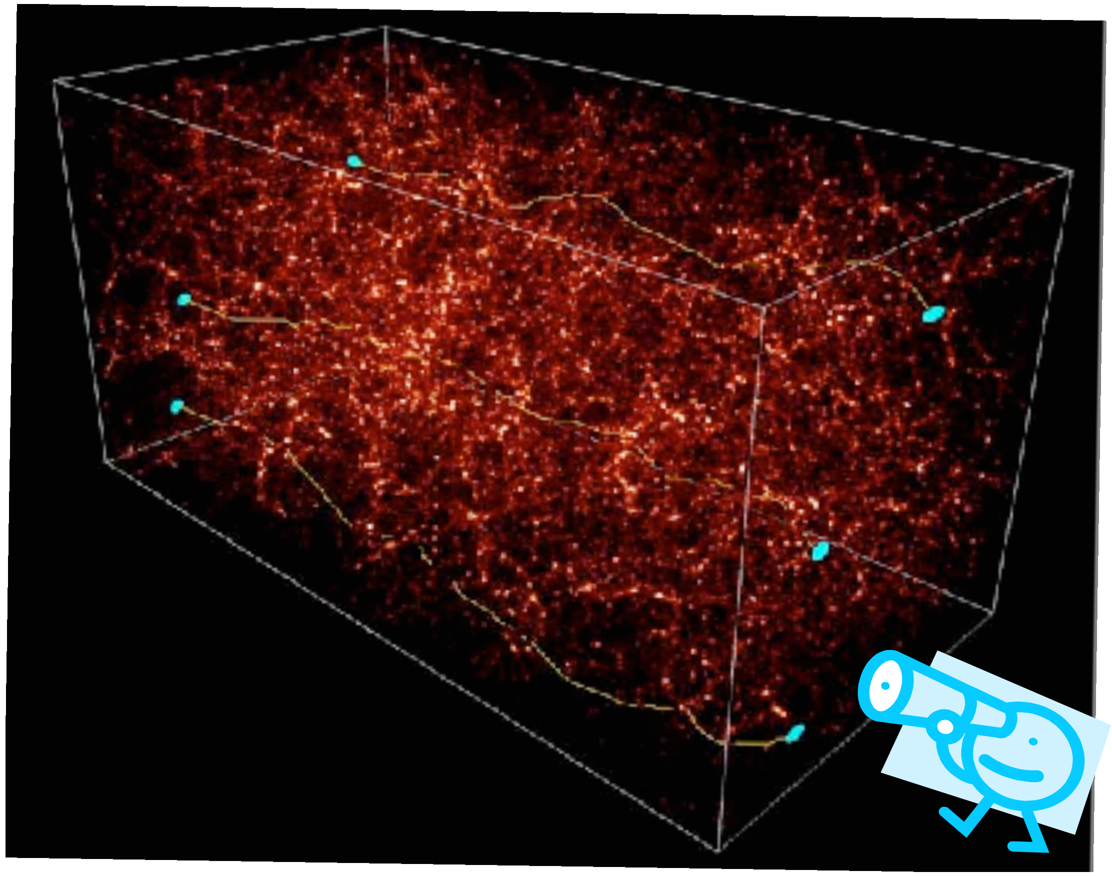
Discussion of bias in shape measurements with general adaptive moments

Patrick Simon

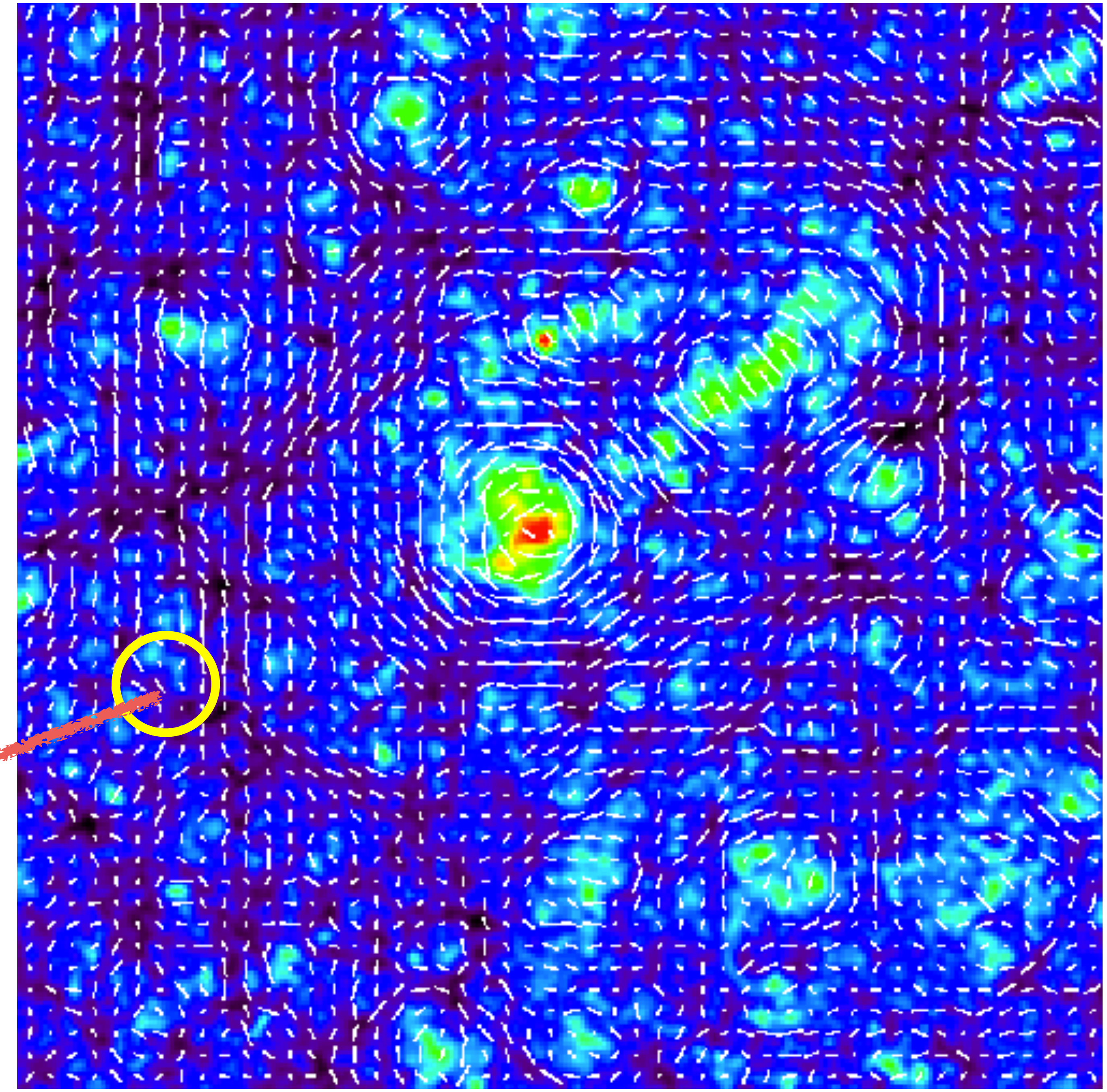
Cosmology / lens seminar
21+28/03/2017

Simon & Schneider (2017)
https://arxiv.org/abs/1609.07937 (v2)

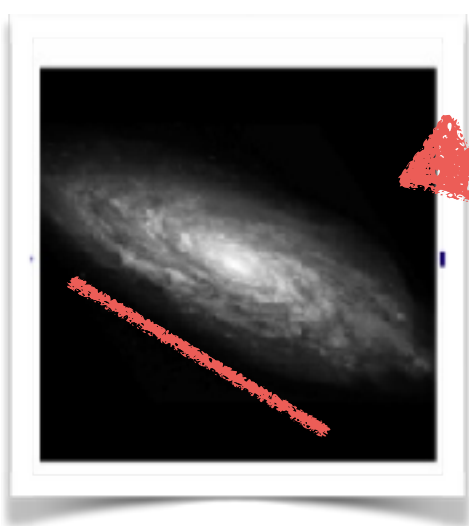
Credit: S. Colombi



Credit:T. Hamana



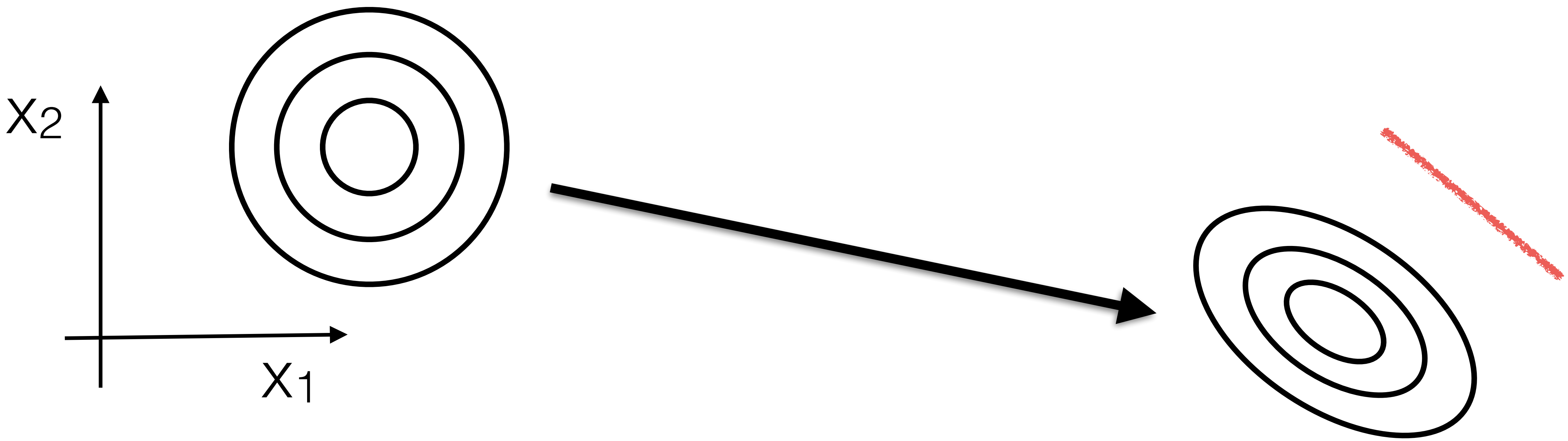
Gravitational shear pattern on piece of sky



shapes of galaxy images are correlated with shear

- change of isophotes under (reduced) shear

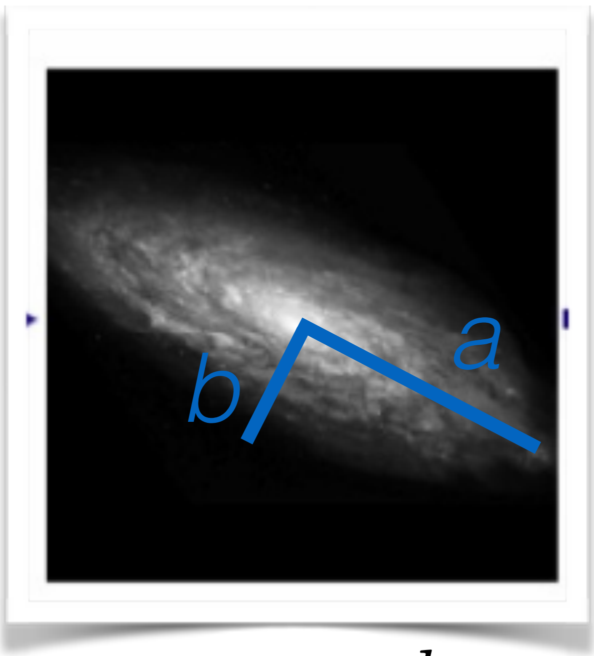
$$(\mathbf{x}-\mathbf{x}_0)^{\mathrm{T}} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} (\mathbf{x}-\mathbf{x}_0) = \mathrm{const}$$



$$(\mathbf{x}-\mathbf{x}_0)^{\mathrm{T}} \begin{pmatrix} 1+g_1 & g_2 \\ g_2 & 1-g_2 \end{pmatrix}^{-2} (\mathbf{x}-\mathbf{x}_0) = \mathrm{const}$$

define reduced shear of image by $g = g_1 + \mathrm{i}\, g_2$

- estimator of reduced shear



$$|\epsilon| = \frac{a-b}{a+b}$$

image surface-brightness $I(\mathbf{x})$

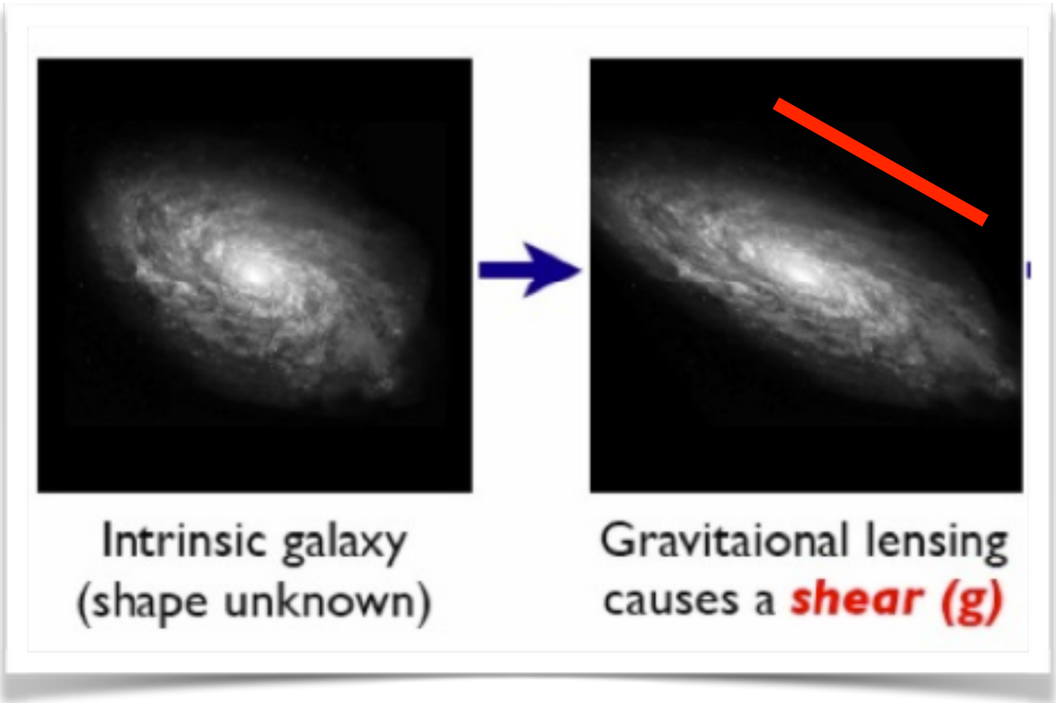
centroid position $$\mathbf{X}_0 = \frac{\int \mathrm{d}^2x \, \mathbf{x} \, I(\mathbf{x})}{\int \mathrm{d}^2x \, I(\mathbf{x})}$$

quadrupole moment $$Q_{ij} = \frac{\int \mathrm{d}^2x \, (x_i - X_{0,i})(x_j - X_{0,j}) \, I(\mathbf{x})}{\int \mathrm{d}^2x \, I(\mathbf{x})}$$

$$\epsilon := \epsilon_1 + \mathrm{i}\,\epsilon_2 = \frac{Q_{11} - Q_{22} + 2\mathrm{i}Q_{12}}{Q_{11} + Q_{22} + 2(Q_{11}Q_{22} - Q_{12}^2)^{1/2}}$$

"third flattening"

- estimator of reduced shear



intrinsic ellipticity $\epsilon_{\rm s}$

sheared ellipticity $\epsilon = \dfrac{\epsilon_{\rm s} + g}{1 + g^* \epsilon_{\rm s}}$

average for *any* isotropic distribution of intrinsic ellipticities:

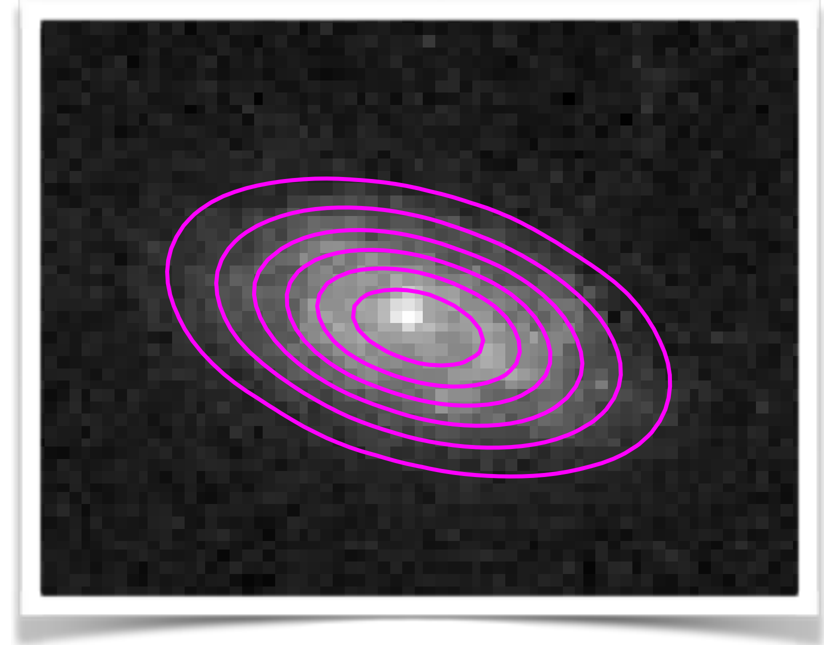
$$\langle \epsilon \rangle = g$$

Seitz & Schneider (1997)

typically $|\epsilon| \sim 0.4$

$|g| \sim 0.01$

- general adaptive moments (GLAM)



Problem: $\int \mathrm{d}^2x \; I(\mathbf{x})$ generally diverges.

One solution: introduce weights $w(\mathbf{x}) \geq 0$

$$\langle w(\mathbf{x}), I(\mathbf{x}) \rangle := \int \mathrm{d}^2x \; w(\mathbf{x}) \, I(\mathbf{x})$$

and use weighted moments

$$\mathbf{x}_0 = \frac{\langle w(\mathbf{x}), \mathbf{x} \, I(\mathbf{x}) \rangle}{\langle w(\mathbf{x}), I(\mathbf{x}) \rangle}$$

$$Q_{ij} = \frac{\langle w(\mathbf{x}), (x_i - x_{0,i})(x_j - x_{0,j}) \, I(\mathbf{x}) \rangle}{\langle w(\mathbf{x}), I(\mathbf{x}) \rangle}$$

- general adaptive moments (GLAM)

GLAM use adaptive weights by least-square fitting an elliptical template to the image, i.e., by minimising

$$E(\mathbf{p}|I) = \Big\langle I(\mathbf{x}) - \underline{A\,f(\rho)}, I(\mathbf{x}) - \underline{A\,f(\rho)} \Big\rangle$$

template

w.r.t. to $\mathbf{p} = (\mathbf{x}_0, \epsilon, t, A)$, where iso-contour is

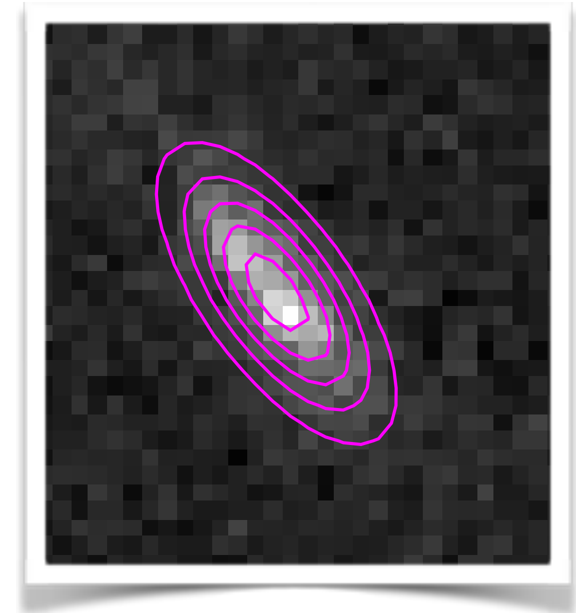
$$\rho := \frac{4}{t^2}(\mathbf{x}-\mathbf{x}_0)^{\mathrm{T}} \begin{pmatrix} 1+\epsilon_1 & \epsilon_2 \\ \epsilon_2 & 1-\epsilon_1 \end{pmatrix}^{-2} (\mathbf{x}-\mathbf{x}_0)$$

and at the *minimum* (if it exists — assumed here)

$$\mathbf{x}_0 = \frac{\langle f'(\rho), \mathbf{x}\, I(\mathbf{x})\rangle}{\langle f'(\rho), I(\mathbf{x})\rangle}$$

$$Q_{ij} = \frac{\langle f'(\rho), (x_i - x_{0,i})(x_j - x_{0,j})\, I(\mathbf{x})\rangle}{\langle f'(\rho), I(\mathbf{x})\rangle} \propto \begin{pmatrix} 1+\epsilon_1 & \epsilon_2 \\ \epsilon_2 & 1-\epsilon_1 \end{pmatrix}^{2}$$

Simon & Schneider (2017); Hirata & Seljak (2003)



★ GLAM ε (and centroid) of best-fit template is that of the adaptively weighted image;

★ the radial weight profile is $w(r) \propto \frac{1}{r} \frac{\mathrm{d}f(r^2)}{\mathrm{d}r}$

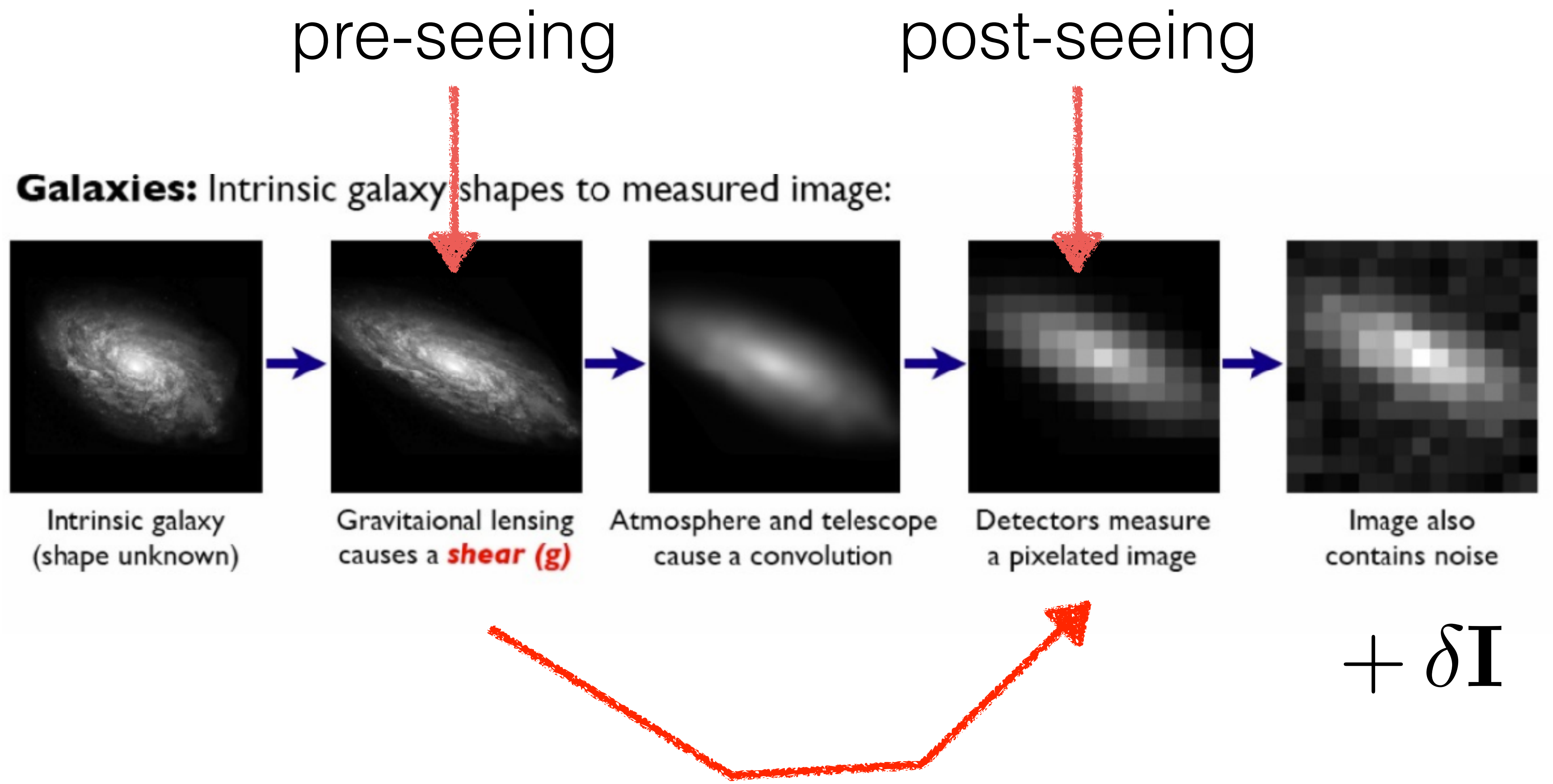
★ GLAM ε is independent of *w(r)* for elliptical galaxy images (is third flattening of iso-contours);

★ for *any* weight *w(r)* and any *I(**x**)*, the GLAM ε is an unbiased estimator of reduced shear;

★ unweighted moments are special case of GLAM, $f(\rho) = \rho$
$\rightarrow w(r) = \mathrm{const}$

★ adaptive, moment-based ellipticities are not fundamentally different to model-based ellipticities;

- Bummer: we need the pre-seeing GLAM ellipticity.



linear mapping $\mathbf{L}$

$$\mathbf{I}_{\rm post} = \mathbf{L}\, I_{\rm pre}(\mathbf{x})$$

- ignore pixel noise; optimal guess for pre-seeing ellipticity?

step 1: assume L is regular (no information loss)

pre-seeing profiles are very finely pixelated; cost function is

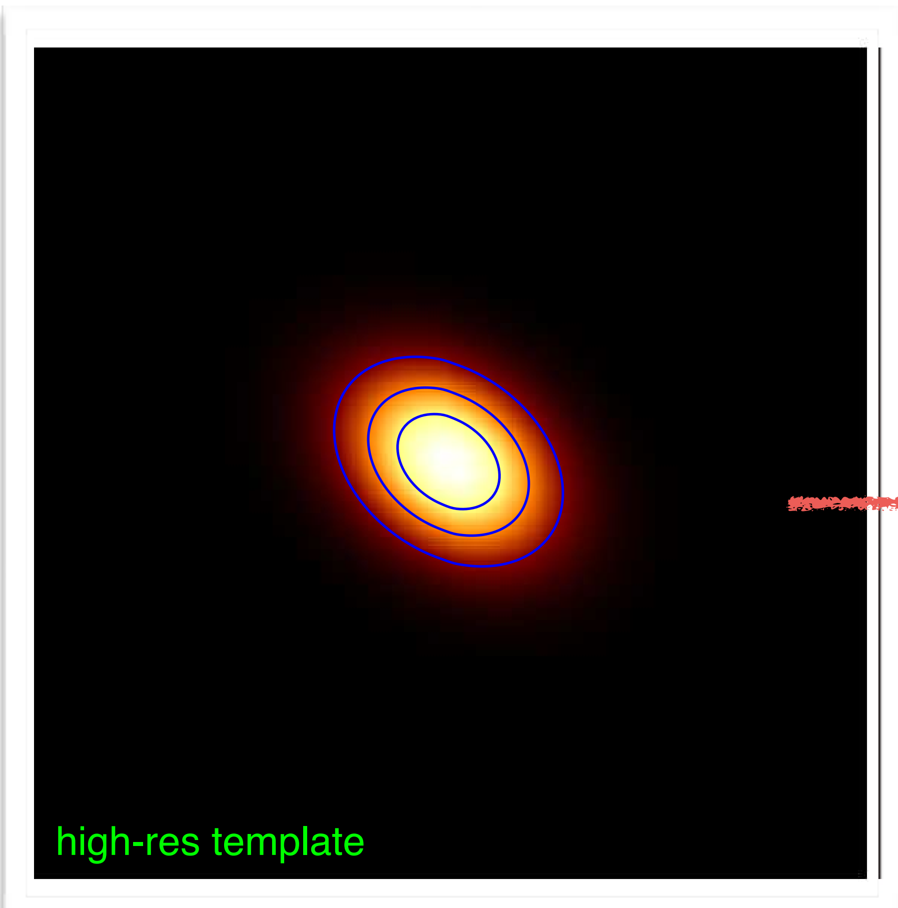
$$E(\mathbf{p}|\mathbf{I}_{\rm pre}) = \left(\mathbf{L}^{-1}\mathbf{I}_{\rm post} - A\,\mathbf{f}_\rho\right)^{\rm T}\left(\mathbf{L}^{-1}\mathbf{I}_{\rm post} - A\,\mathbf{f}_\rho\right)$$

then obtain pre-seeing ellipticity form post-seeing image:

$$\begin{aligned} E(\mathbf{p}|\mathbf{I}_{\rm pre}) &= \left(\mathbf{L}^{-1}\mathbf{I}_{\rm post} - A\,\mathbf{f}_\rho\right)^{\rm T}\left(\mathbf{L}^{-1}\mathbf{I}_{\rm post} - A\,\mathbf{f}_\rho\right) \\ &= \left(\mathbf{I}_{\rm post} - A\,\mathbf{L}\,\mathbf{f}_\rho\right)^{\rm T}\left(\mathbf{L}\mathbf{L}^{\rm T}\right)^{-1}\left(\mathbf{I}_{\rm post} - A\,\mathbf{L}\,\mathbf{f}_\rho\right) \end{aligned}$$

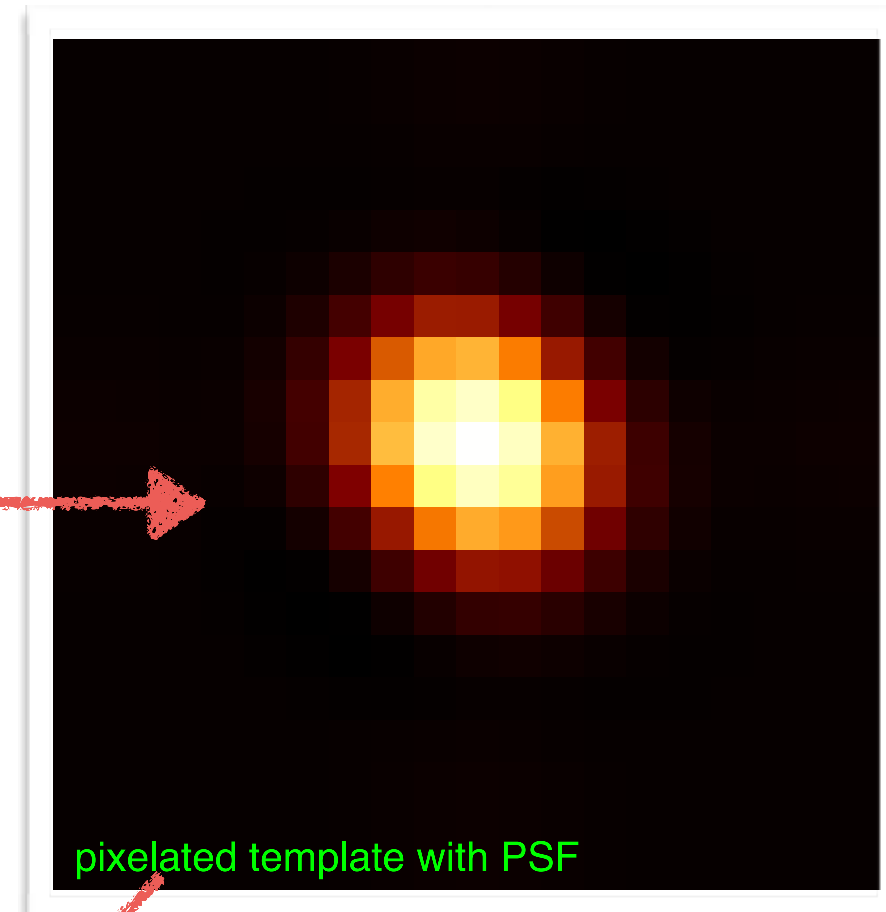
Therefore, we fit in post-seeing frame template $\mathbf{L}\mathbf{f}_\rho$ with metric $\mathbf{U} := (\mathbf{L}\mathbf{L}^{\rm T})^{-1}$ (no bias!).

- conceptional GLAM procedure (noiseless images)

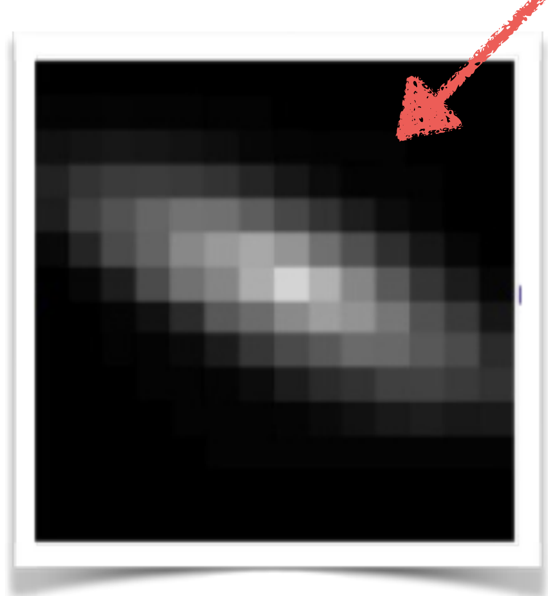


pre-seeing frame

$$\mathbf{f}_\rho \rightarrow \mathbf{L}\,\mathbf{f}_\rho$$



post-seeing frame



fit template to images with metric

$$\mathbf{U}_{\rm best} = (\mathbf{L}\mathbf{L}^{\rm T})^{+}$$

- ignore pixel noise; optimal guess for pre-seeing ellipticity?

step 2: assume L is singular (information loss: mainly pixelation)

Ansatz for cost function:

$$E(\mathbf{p}|\mathbf{I}_{\rm post}) = (\mathbf{I}_{\rm post} - A\,\mathbf{L}\,\mathbf{f}_\rho)^{\rm T}\;\mathbf{U}\;(\mathbf{I}_{\rm post} - A\,\mathbf{L}\,\mathbf{f}_\rho)$$

Unbiased if pre-seeing image $\mathbf{I}_{\rm pre}$ is prefectly fit by $\mathbf{f}_\rho$; but biased if best-fit in pre-seeing frame has residuals $\mathbf{R}_{\rm pre}$, namely

$$\delta\mathbf{p} = (\mathbf{G}\mathbf{U}_{\rm L}\mathbf{G})^{-1}\,\mathbf{G}^{\rm T}\mathbf{U}_{\rm L}\,\mathbf{R}_{\rm pre} + \mathcal{O}(R^2_{\rm pre})$$

GLAM exhibit “underfitting bias” if

(i) $\mathbf{I}_{\rm pre}$ is unmatchable with template;
(ii) *and* if **L** is singular;
(iii) and if $\mathbf{L}\mathbf{R}_{\rm pre}<>0$;

$$\mathbf{G}_i = \left.\frac{\partial(A\mathbf{f}_\rho)}{\partial p_i}\right|_{\mathbf{p}_{\rm true}}$$

$$\mathbf{U}_{\rm L} := \mathbf{L}^{\rm T}\mathbf{U}\mathbf{L}$$

$$\mathbf{U}_{\rm best} = (\mathbf{L}\mathbf{L}^{\rm T})^{+}$$

- Conclusions for images without pixel noise

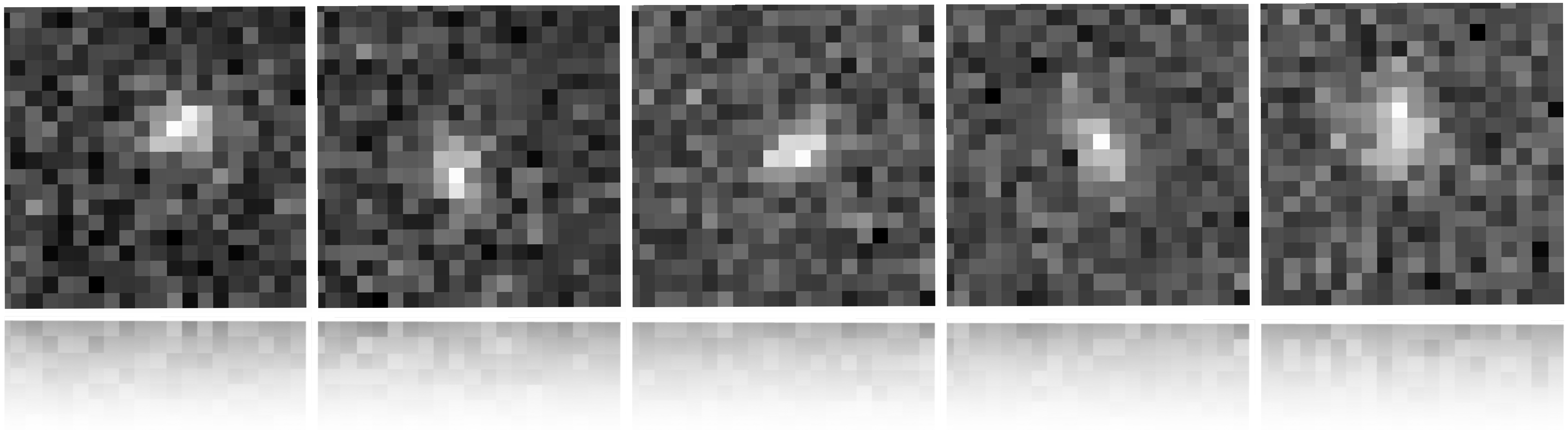
★ adaptive-moment ellipticities are parameters of best-fits with elliptical templates; adaptive weight is *derivative* of template;

★ no information loss in $\mathbf{I}_{post}$: estimator of GLAM ellipticity is unbiased; then also unbiased estimator of g;

★ information loss in $\mathbf{I}_{post}$: bias depends on residuals between best-fit template and image in *pre-seeing frame* (necessary for bias is $\mathbf{LR}_{pre} \neq 0$ *and* a singular $\mathbf{L}$);

★ bias is reduced by *template profile* that closely resembles that of pre-seeing image;

★ there is bias for pixelated and non-elliptical images;



There shall be noise!
(part II)

$$\langle \delta \mathbf{I} \rangle = 0$$

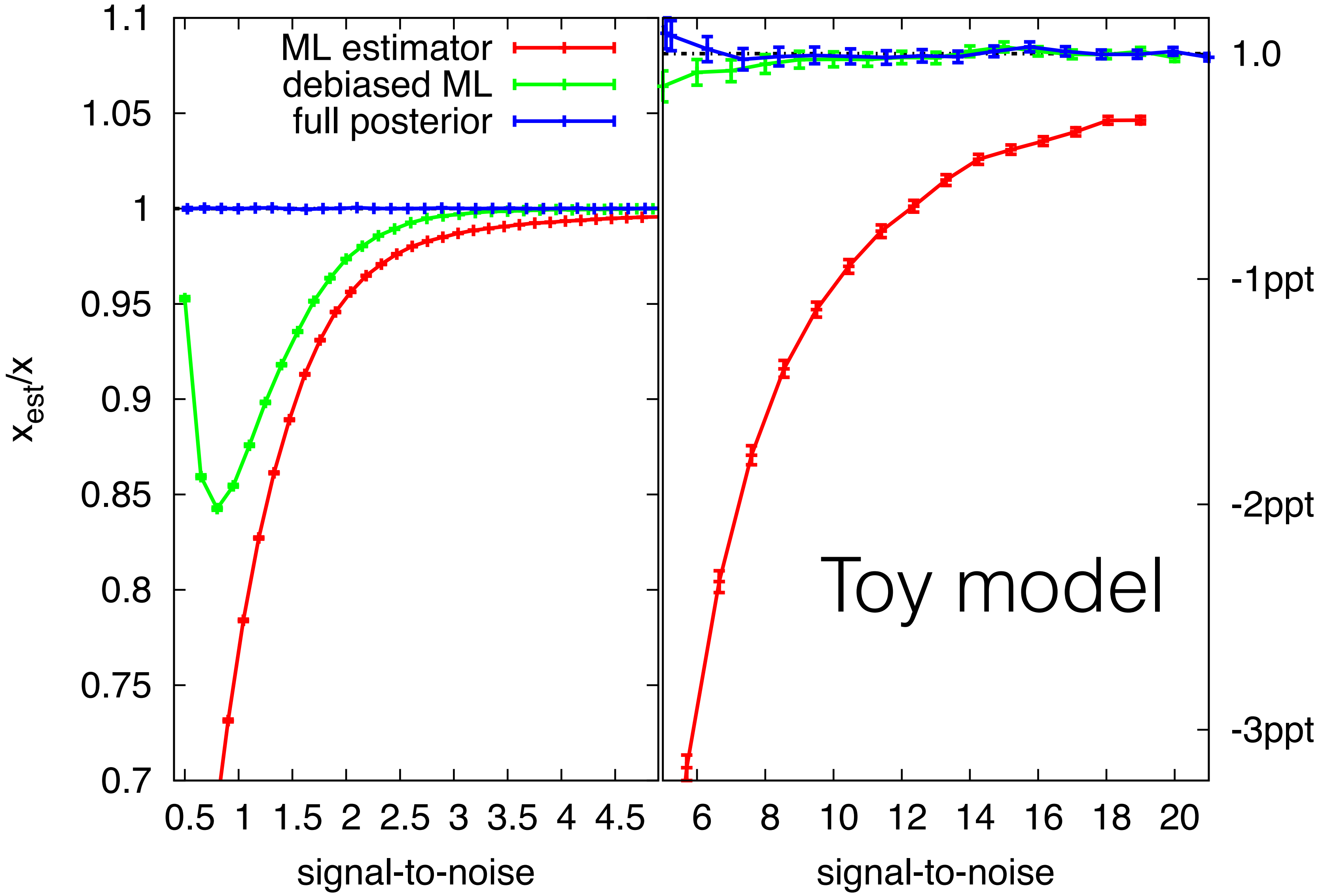
$$\langle \delta \mathbf{I}\, \mathbf{I}_{\rm post}^{\rm T} \rangle = 0$$

$$N = \langle \delta \mathbf{I}\, \delta \mathbf{I}^{\rm T} \rangle$$

$$\mathbf{I} = \mathbf{I}_{\rm post} + \delta \mathbf{I}$$

- illustration of noise bias

bias of estimator varies with S/N; not present without noise



What is x ?

$$y = x^3 + n \; ; \; n \curvearrowleft N(0, \sigma)$$

Maximum likelihood

$$x_{\rm ml} = y^{1/3}$$

$$\langle x_{\rm ml} \rangle = \sum_{i=1}^{N} y_i^{1/3} / N$$

Bayesian posterior

$$p(x|y) \propto \mathcal{L}(y|x)\, p(x)$$

Combine information!

$$p(x|y_1, y_2, \ldots) \propto \prod_{i=1}^{N} \mathcal{L}(y_i|x)\, p(x)$$

x: “ellipticity”; y: “image”

- Bayesian GLAM: do not use point estimates of ellipticity; keep full statistical information;

Likelihood function of GLAM ellipticity (Gaussian noise)?

Inferring adaptive moments by forward-fitting templates….

$$-2\ln\mathcal{L}(\mathbf{I}|\mathbf{p},\mathbf{R}_{\rm pre})+{\rm const}=\Big(\mathbf{I}-A\,\mathbf{L}\,\mathbf{f}_\rho-\mathbf{L}\,\mathbf{R}_{\rm pre}\Big)^{\rm T}\mathbf{N}^{-1}\Big(\mathbf{I}-A\,\mathbf{L}\,\mathbf{f}_\rho-\mathbf{L}\,\mathbf{R}_{\rm pre}\Big)$$
$$=:\ \|\mathbf{I}-A\,\mathbf{L}\,\mathbf{f}_\rho-\mathbf{L}\,\mathbf{R}_{\rm pre})\|^2_{\mathbf{N}}$$

We do not know the pre-seeing residuals.

Option 1: hope that they are not relevant, and use a *misspecified* likelihood ,

$$-2\ln\mathcal{L}(\mathbf{I}|\mathbf{p})+{\rm const}=\|\mathbf{I}-A\,\mathbf{L}\,\mathbf{f}_\rho\|^2_{\mathbf{N}}$$

Option 2: quantify your ignorance by a *prior* density for residuals,

$$\mathcal{L}(\mathbf{I}|\mathbf{p})=\int{\rm d}\mathbf{R}_{\rm pre}\;p(\mathbf{R}_{\rm pre}|\mathbf{p})\,\mathcal{L}(\mathbf{I}|\mathbf{p},\mathbf{R}_{\rm pre})$$

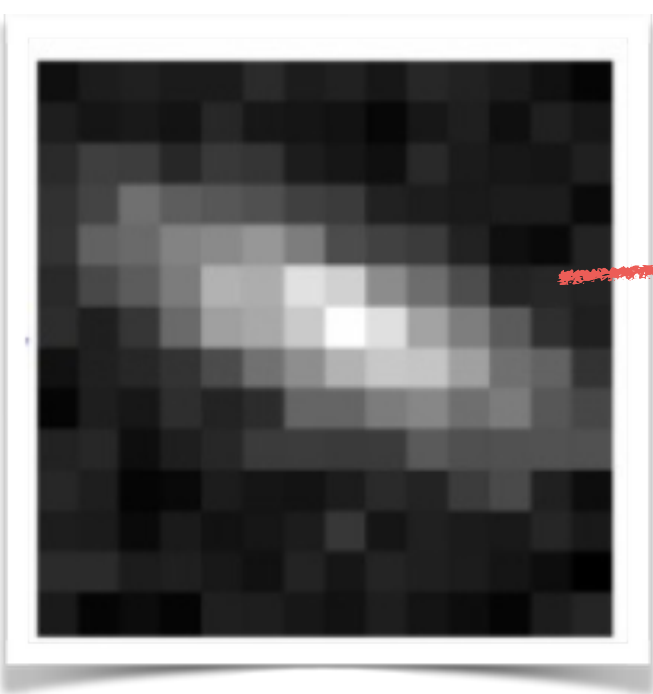
- posterior density of GLAM ellipticity

prior

$$p(\epsilon|\mathbf{I}) \propto \int \underbrace{\mathrm{d}A\,\mathrm{d}t\,\mathrm{d}\mathbf{x}_0}_{} \overbrace{p(A,t,\mathbf{x}_0,\epsilon)}^{} \mathcal{L}(\mathbf{I}|\mathbf{p})$$

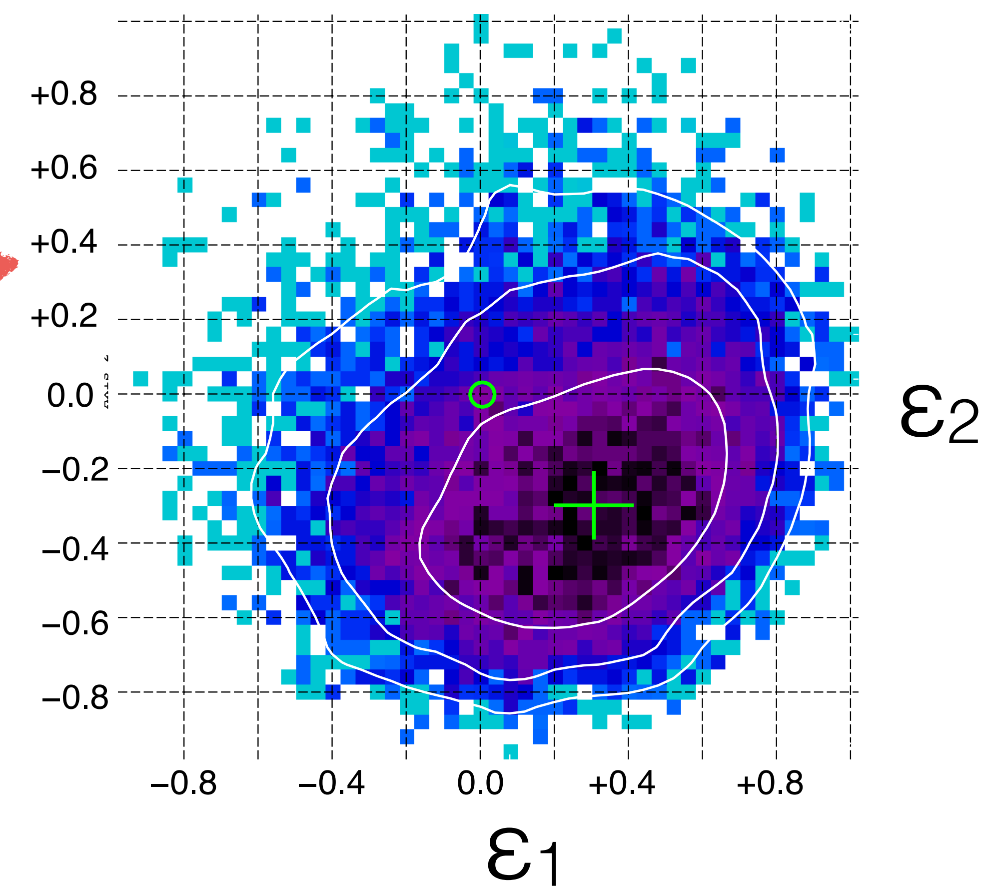
nuisance parameters

Adopting relaxed, uniform prior. Prior details should not matter?

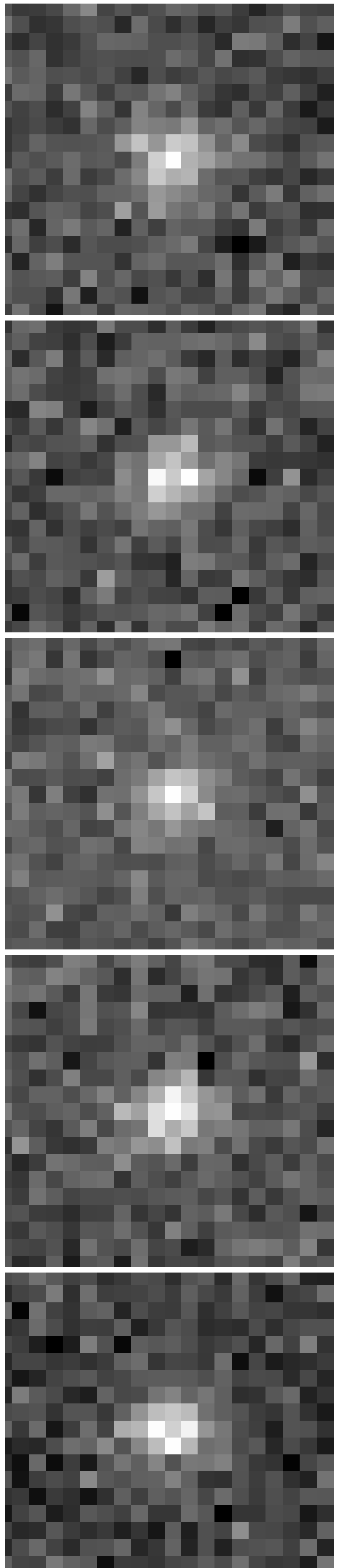


Monte-Carlo sampling

$p(\epsilon|\mathbf{I})$



- Experiment 1: i.i.d. exposures of *same* pre-seeing galaxy $\mathbf{I}_{\mathrm{post}}$



Combine likelihoods of all exposures for posterior

$$p(\mathbf{p}|\mathbf{I}_1,\mathbf{I}_2,\dots) \propto \prod_{i=1}^{n} \mathcal{L}(\mathbf{I}_i|\mathbf{p})\, p(\mathbf{p})$$

Investigate consistency by considering *n*->infinity,

noise-free

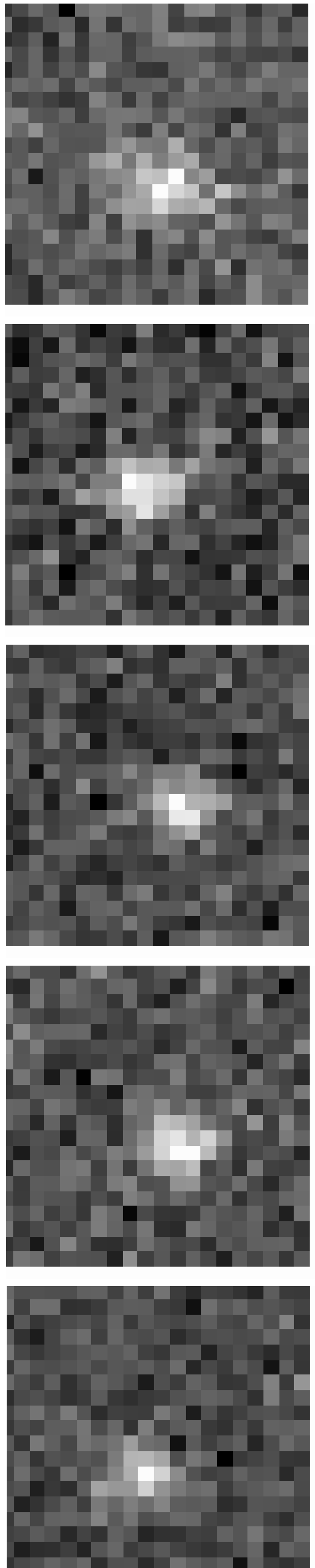
$$-2\ln p(\mathbf{p}|\mathbf{I}_1,\mathbf{I}_2,\dots) \simeq \mathrm{const} + n\|\mathbf{I}_{\mathrm{post}} - A\,\mathbf{L}\,\mathbf{f}_\rho\|^2_{\mathbf{N}}$$

Asymptotic posterior puts all probability mass at minimum of

$$(\mathbf{I}_{\mathrm{post}} - A\,\mathbf{L}\,\mathbf{f}_\rho)^{\mathrm{T}}\mathbf{N}^{-1}(\mathbf{I}_{\mathrm{post}} - A\,\mathbf{L}\,\mathbf{f}_\rho) = \min$$

1. Same as noise-free problem with metric $\mathbf{U} = \mathbf{N}^{-1}$
2. There is no noise bias: no effect for $\mathbf{N}^{-1} \rightarrow \lambda\,\mathbf{N}^{-1}$
3. Magnitude of bias depends on $\mathbf{LR}_{\mathrm{pre}}$ and $\mathbf{L}^{\mathrm{T}}\mathbf{N}^{-1}\mathbf{L}$

- Experiment 2: sample of *different* pre-seeing galaxies, same ε



Same ellipticity but different $\mathbf{q}_i = (A_i, t_i, \mathbf{x}_{0,i})$

Combine marginal posteriors of same ellipticity:

$$p(\epsilon|\mathbf{I}_1, \mathbf{I}_2, \ldots) \propto \prod_{i=1}^{n} p(\epsilon|\mathbf{I}_i)$$

$$= \int \underbrace{\mathrm{d}\mathbf{q}_1 \mathrm{d}\mathbf{q}_2 \ldots}_{\text{marginalize}} \prod_{i=1}^{n} \mathcal{L}(\mathbf{I}_i|\epsilon, \mathbf{q}_i)\, \underbrace{p(\mathbf{q}_i)\, p(\epsilon)}_{\text{prior}}$$

$$\simeq \prod_{i=1}^{n} p(\epsilon|\mathbf{I}_{\mathrm{post,i}})$$

noise-free

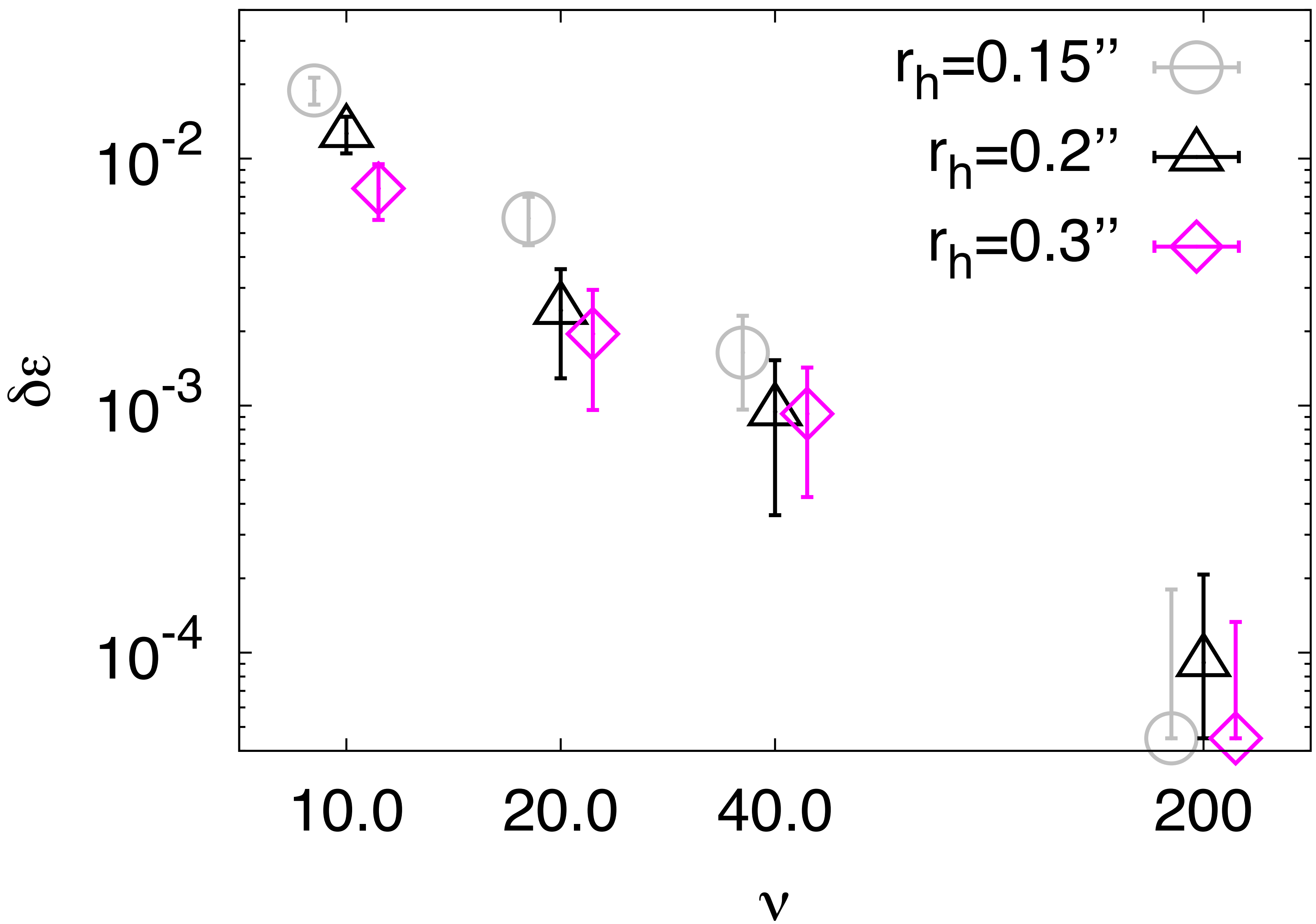
Now asymptotic consistency explicitly depends on prior density of nuisance parameters.

Prior does not become irrelevant here because complexity of model increases with every new image $\mathbf{I}_i$ in the sample.

- Bayesian (nuisance) priors can be *incorrect* (prior bias).

Correctly specified likelihood but incorrect prior:

bias



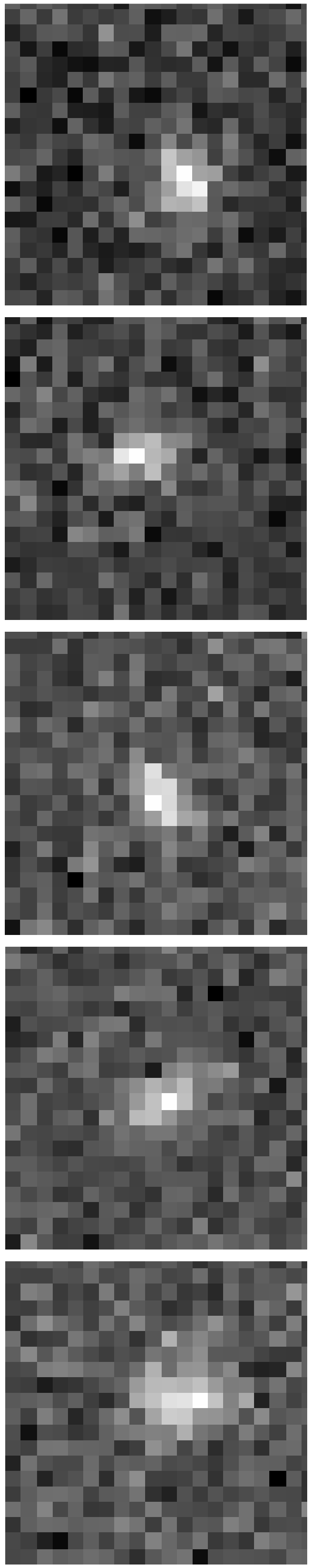
pixel 0.1''

PSF 0.1''

1. Prior bias introduces noise bias into Bayesian picture after all.
2. Becomes irrelevant if likelihood for *individual* images dominates.
3. Vanishes if prior equals actual distrb. of nuisance parameters.

Intermission

- Experiment 3: sample of *different* pre-seeing galaxies, same shear



Similar to Experiment 2 but now with shear posteriors:

prior for intrinsic shapes

$$p(g|\mathbf{I}) = p_{\rm g}(g)(1-|g|^2)^2 \int {\rm d}^2\epsilon \, \frac{p_{\rm s}\Big(\epsilon_{\rm s}(g,\epsilon)\Big)\, p(\epsilon|\mathbf{I})}{\mathcal{N}(\epsilon)|1-\epsilon g^*|^4}$$

prior *g*

where

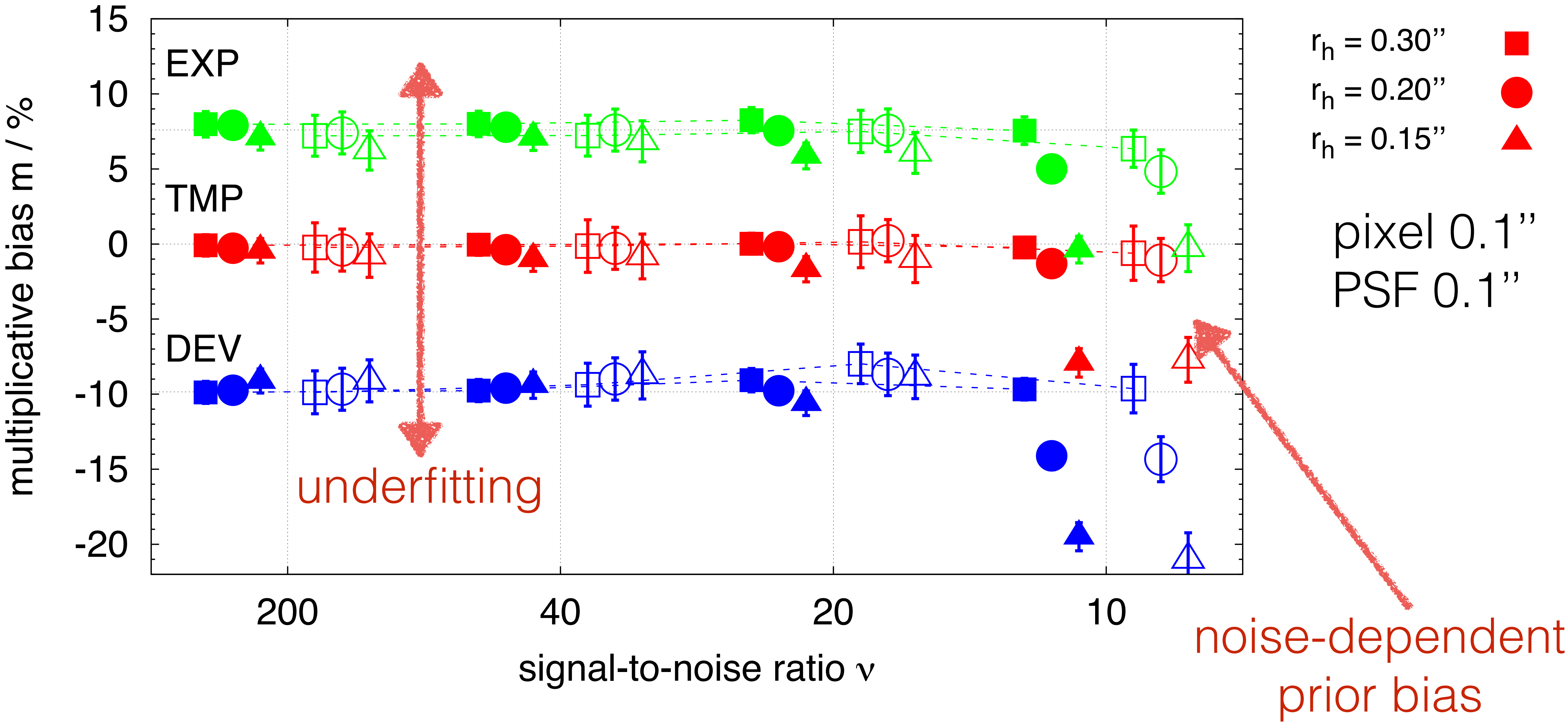
$$\mathcal{N}(\epsilon) := \int {\rm d}^2 g \, \frac{p_{\rm g}(g)\, p_{\rm s}\Big(\epsilon_{\rm s}(g,\epsilon)\Big)(1-|g|^2)^2}{|1-\epsilon g^*|^4} \; ; \; \epsilon_{\rm s}(g,\epsilon) := \frac{\epsilon - g}{1-\epsilon g^*}$$

Combine and marginalize to constrain reduced shear:

$$p(g|\mathbf{I}_1,\mathbf{I}_2,\dots) \propto \prod_{i=1}^{n} p(g|\mathbf{I}_i) \simeq \prod_{i=1}^{n} p(g|\mathbf{I}_{\rm post,i})$$

- Results for numerical analysis

$$g = (1+m)\, g_{\mathrm{true}} + c$$



open symbols: uniform prior for intrinsic shapes

filled symbols: correct prior for intrinsic shapes

- Conclusions for images with pixel noise

★ can construct a marginal posterior of GLAM ellipticity;

★ likelihood is misspecified for $\mathbf{L}\mathbf{R}_{\text{pre}} \neq 0$; could be fixed with prior density for $\mathbf{R}_{\text{pre}}$ (dedicated survey);

★ misspecified likelihood produces underfitting bias; depends on **L**, heterogeneity and correlation of noise but not overall S/N;

★ intrinsically different images $\mathbf{I}_{\text{pre}}$ can introduce noise-dependent prior bias if prior density of $\boldsymbol{q}=(A,t,\boldsymbol{x}_0)$ is not distribution of $\boldsymbol{q}$ in sample;

★ prior bias prominent when posterior of ellipticity is dominated by prior in individual image;

- backup slide: non-adaptive weighted moments

Assume fixed weight $w(\mathbf{x})$ and minimise w.r.t. to $\mathbf{p} = (\mathbf{x}_0, \epsilon, t, A)$

$$E(\mathbf{p}|I) = \langle w(\mathbf{x})\,[I(\mathbf{x}) - A\,\rho], I(\mathbf{x}) - A\,\rho\rangle$$

to find at the minimum the relations

$$Q_{ij} = \frac{\langle w(\mathbf{x}), (x_i - x_{0,i})(x_j - x_{0,j})\,I(\mathbf{x})\rangle}{\langle w(\mathbf{x}), I(\mathbf{x})\rangle}$$

$$\propto \begin{pmatrix} 1+\epsilon_1 & \epsilon_2 \\ \epsilon_2 & 1-\epsilon_1 \end{pmatrix}^2$$

$$\mathbf{x}_0 = \frac{\int \mathrm{d}^2x\; w(\mathbf{x})\,\mathbf{x}\,I(\mathbf{x})}{\int \mathrm{d}^2x\; w(\mathbf{x})\,I(\mathbf{x})}$$

Non-adaptive weighted moments are parameters of best-fit template with a metric $w(x)$ and $f(\rho) = \rho$.