Natural Image Statistics

Siwei Lyu

SUNY Albany

CIFAR NCAP Summer School 2009
August 7, 2009

Thanks to Eero Simoncelli for sharing some of the slides

big numbers

big numbers

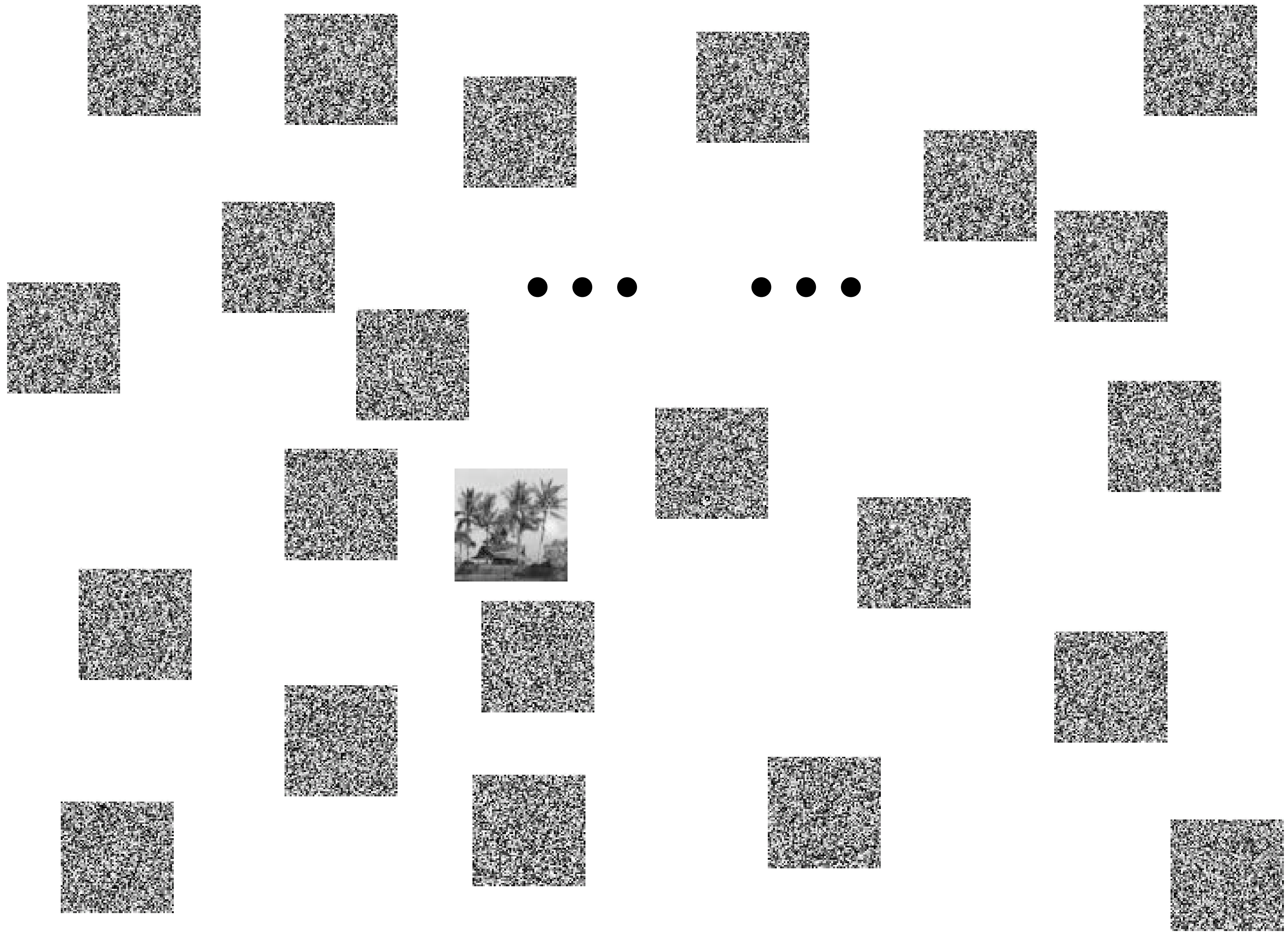
- seconds since big bang: ~ 10^{17}

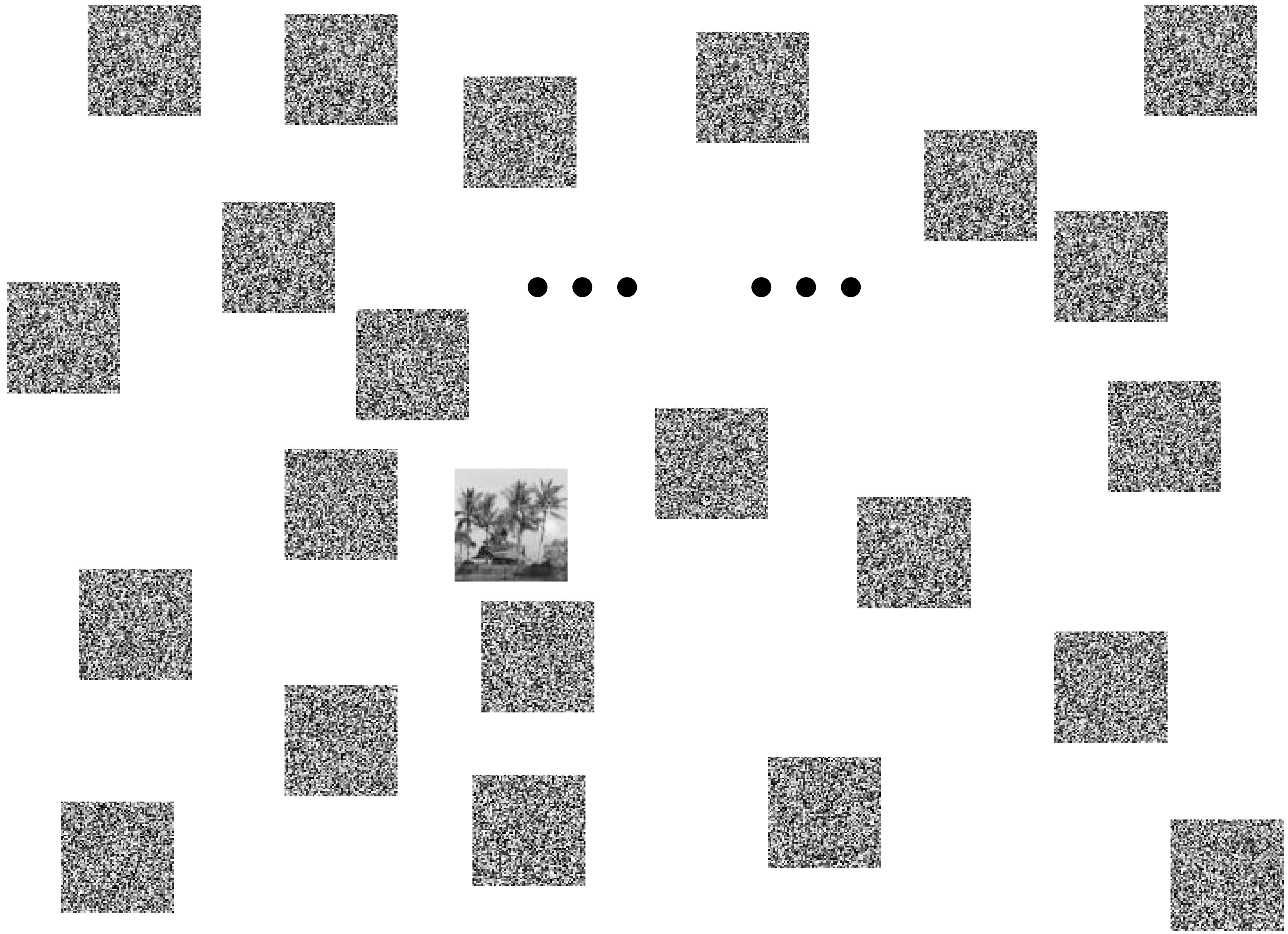
big numbers

- seconds since big bang: ~ 10^{17}
- atoms in the universe: ~10^{80}

big numbers

- seconds since big bang: ~ 10^{17}
- atoms in the universe: ~10^{80}
- 65×65 8-bit gray-scale images: ~10^{10000}





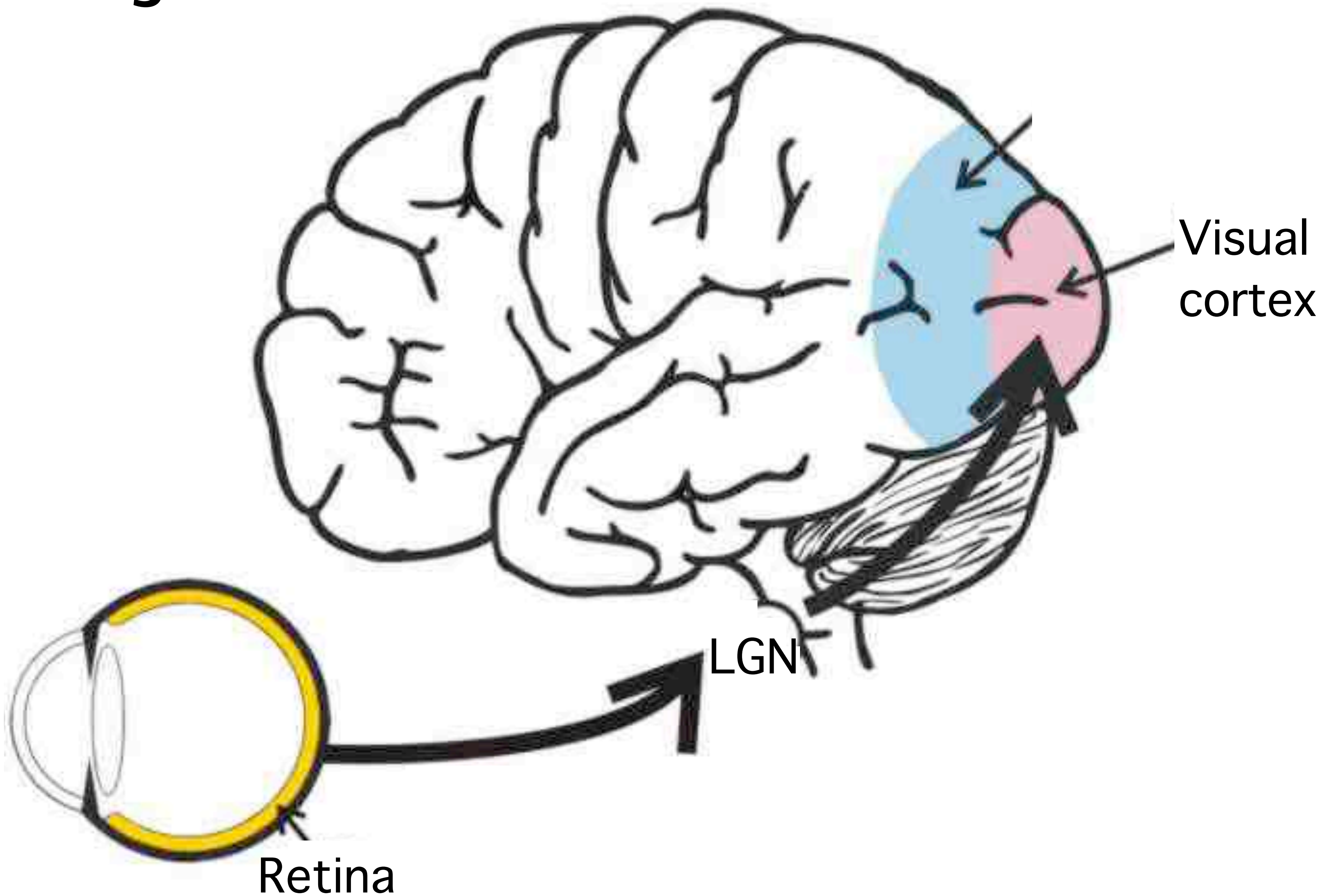
"The distribution of natural images is complicated. Perhaps it is something like ***beer foam***, which is mostly empty but contains a thin mesh-work of fluid which fills the space and occupies almost no volume. The fluid region represents those images which are natural in character."

[Ruderman 1996]

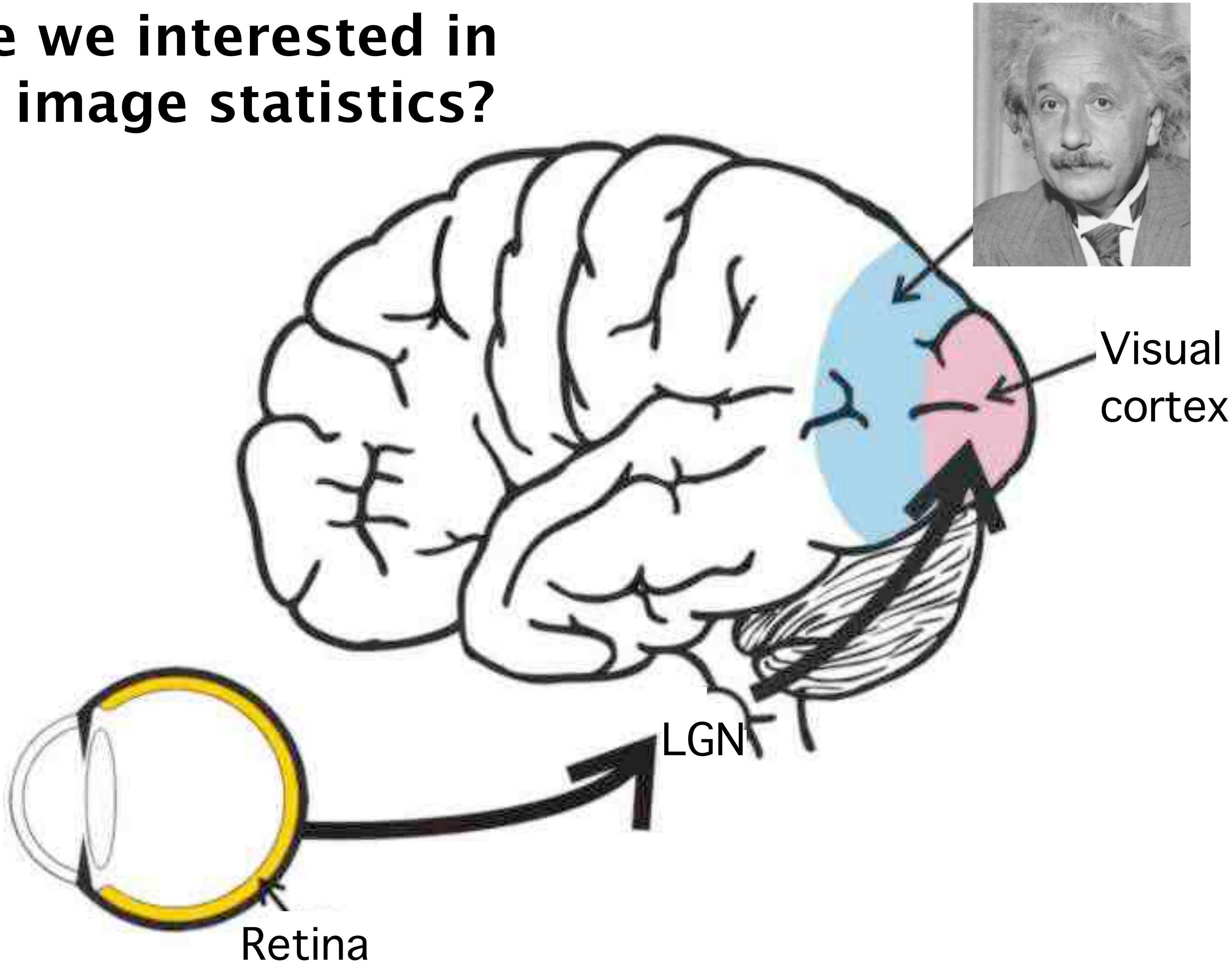
natural image statistics

- natural images are rare in image space
- they distinguish by nonrandom structures
- common statistical properties of natural images is the focal element in the study of natural image statistics

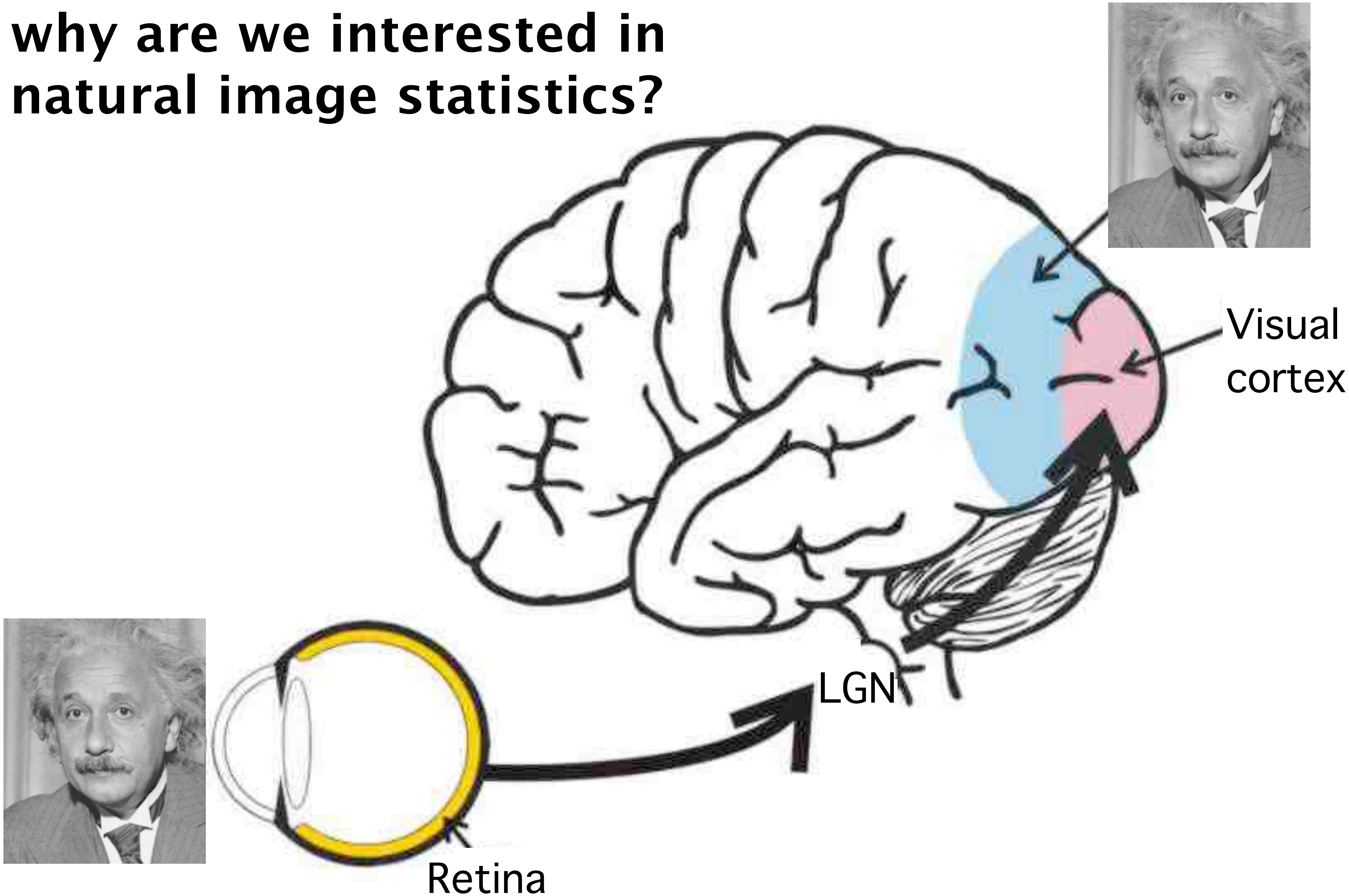
why are we interested in natural image statistics?



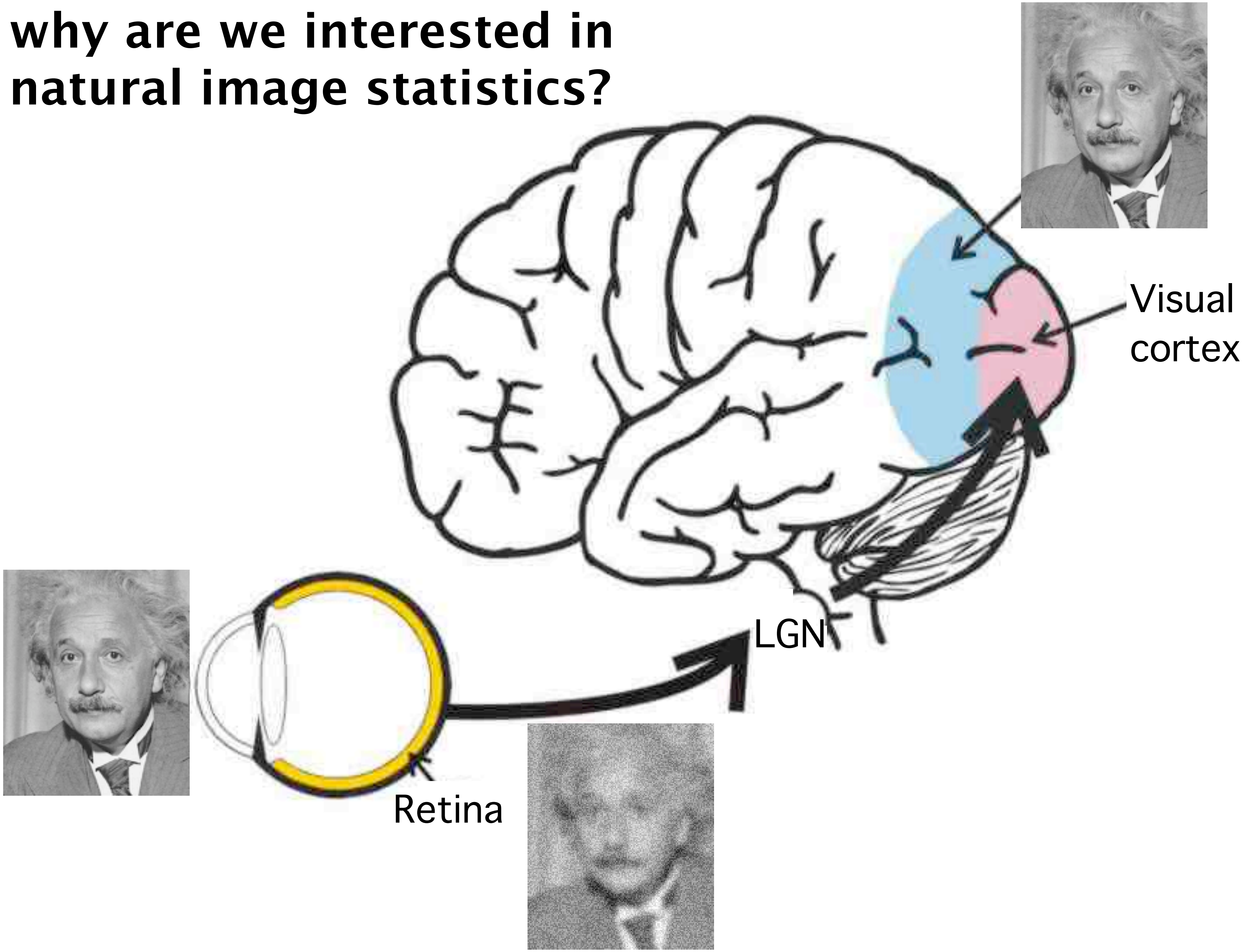
why are we interested in natural image statistics?

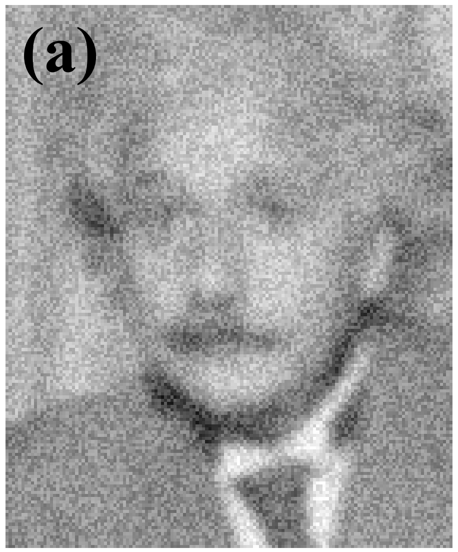


why are we interested in natural image statistics?

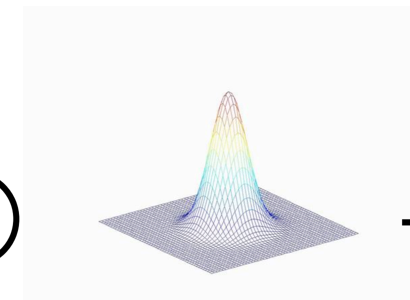


why are we interested in natural image statistics?

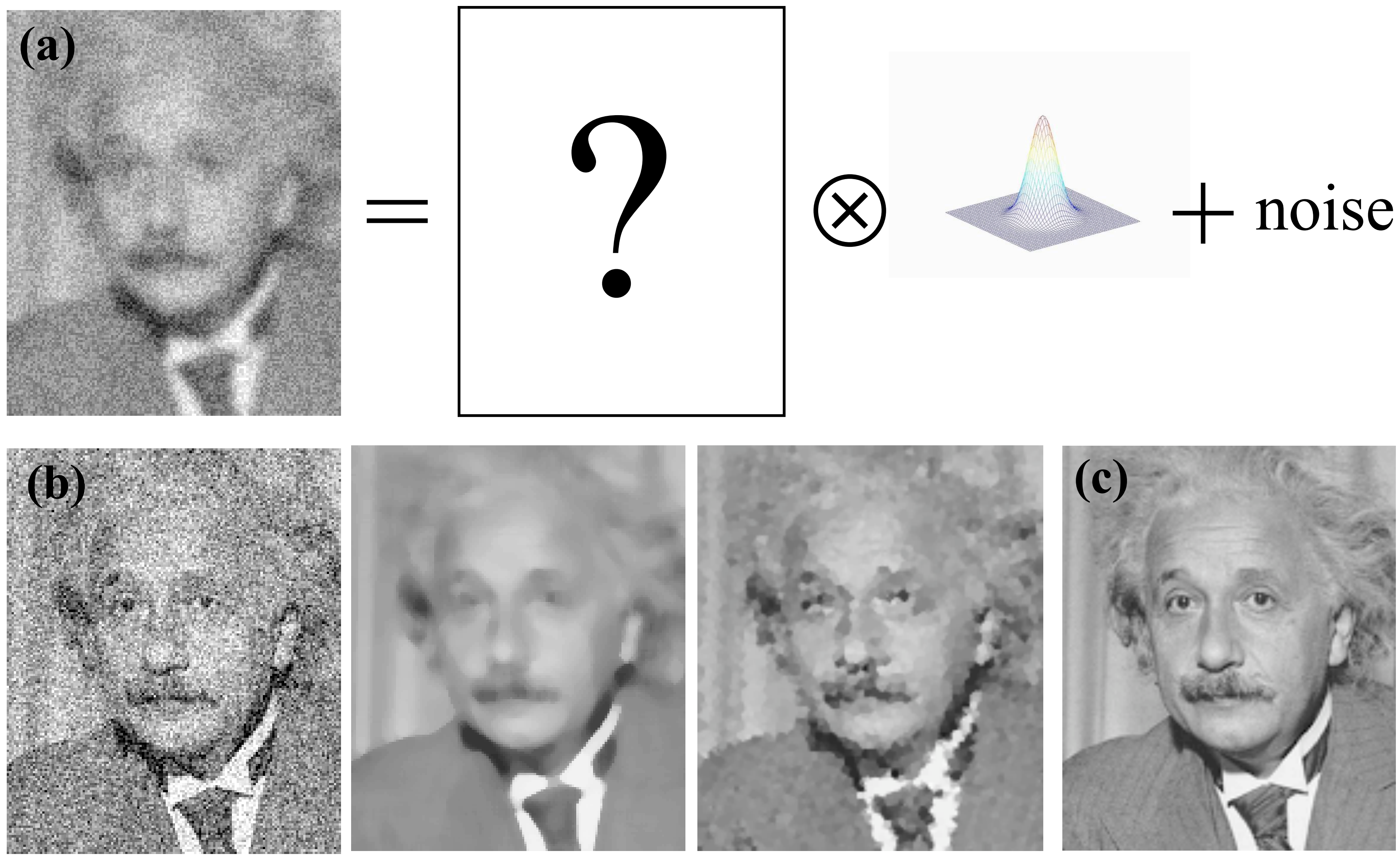




$$= \boxed{\;?\;} \otimes$$

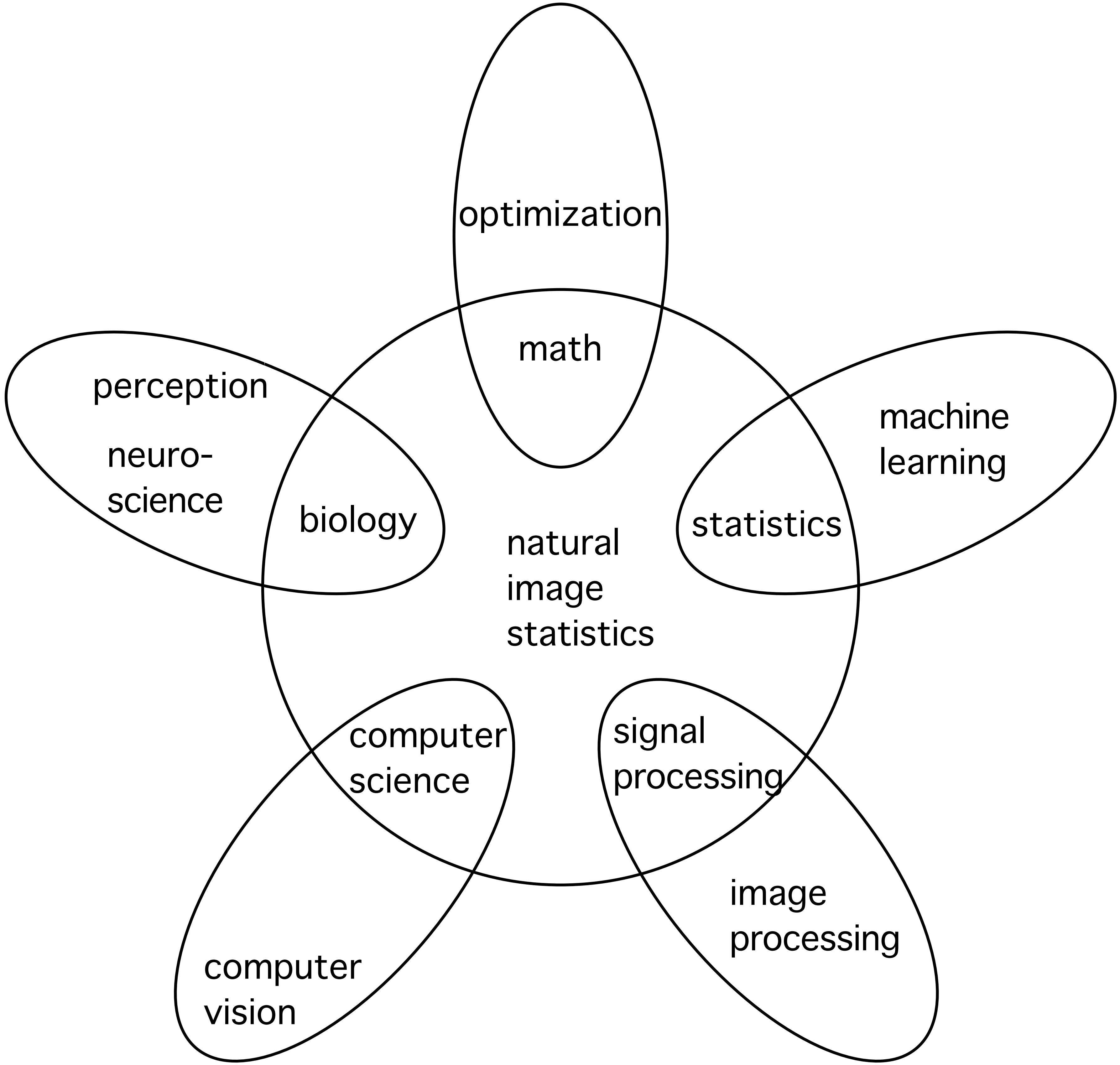


$+$ noise


(a)
=
?
⊗
+ noise
(b)
(c)

computer vision applications

- image restoration
  - de-noising, de-blurring and de-mosaicing, super-resolution and in-painting
- image compression
- texture synthesis
- image segmentation
- features for object detection and classification (SIFT, gist, “primal sketch”, saliency, etc)
- many others


optimization
math
perception
neuro-
science
biology
machine
learning
statistics
natural
image
statistics
computer
science
signal
processing
image
processing
computer
vision

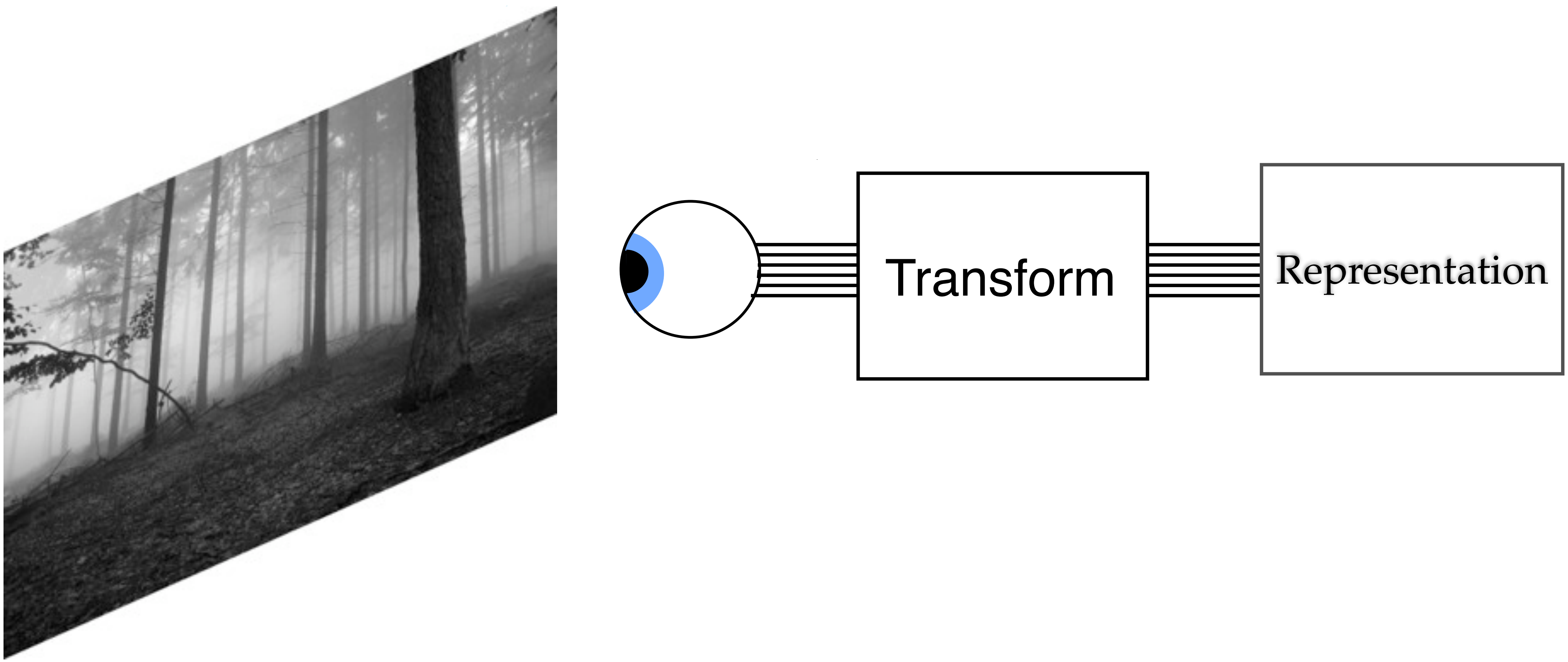
scope of this tutorial

- important developments following a general theme
- focusing on concepts
  - light on math or specific applications
- gray-scale intensity image, do not cover
  - color
  - time (video)
  - multi-image information (stereo)

main components

representation

representation



why representation matters?

- example (from David Marr)
- representation for numbers
  - Arabic: 123
  - Roman: MCXXIII
  - binary: 1111011
  - English: one hundred and twenty three

why representation matters?

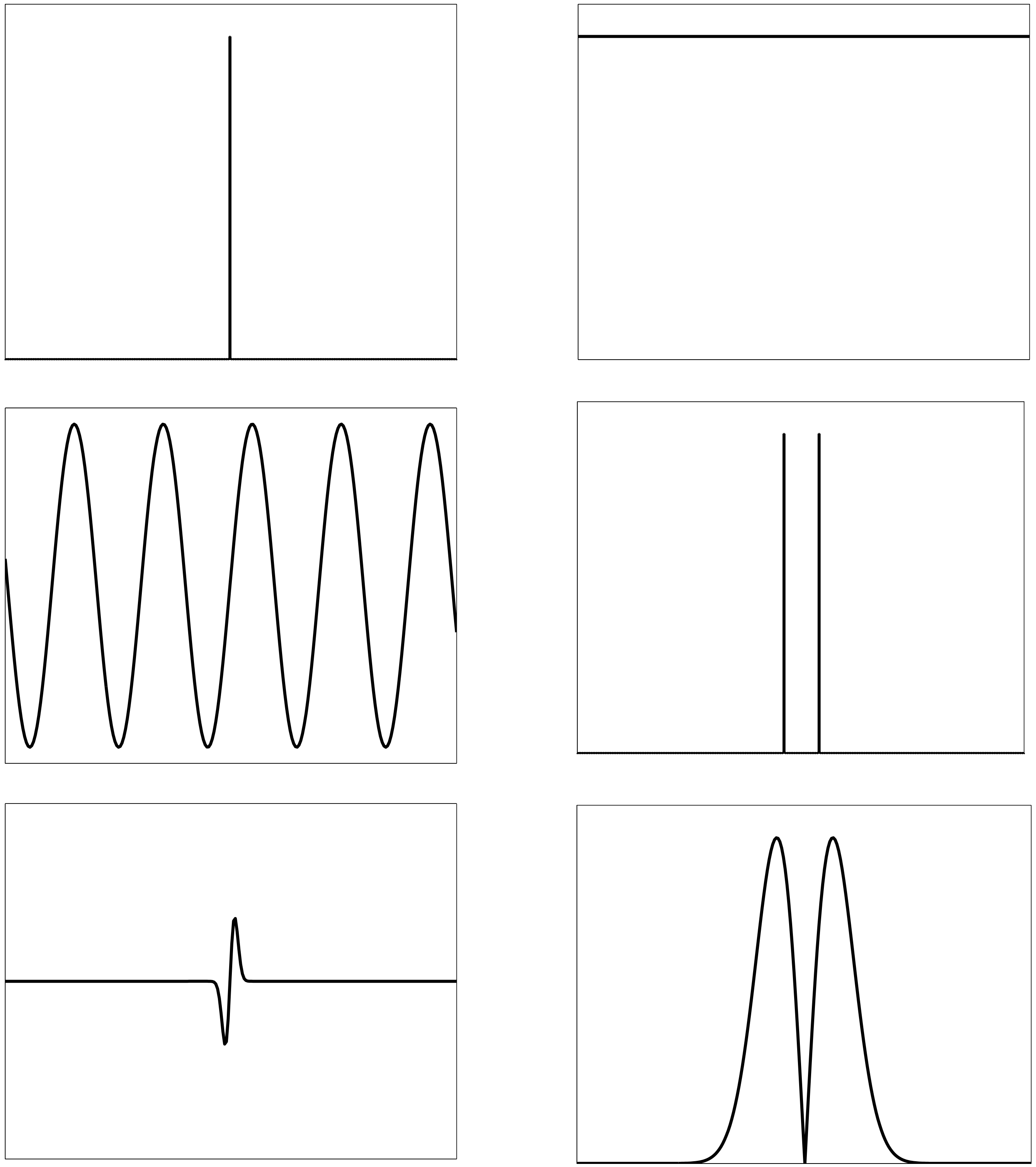
- example (from David Marr)
- representation for numbers
  - Arabic: 123 × 10
  - Roman: MCXXIII × X
  - binary: 1111011 × 110
  - English: one hundred and twenty three × ten

why representation matters?

- example (from David Marr)
- representation for numbers
  - Arabic: 123×4
  - Roman: MCXXIII $\times$ IV
  - binary: 1111011×100
  - English: one hundred and twenty three $\times$ four

linear representations

- pixel
- Fourier
- wavelet
  - localized
  - oriented



main components

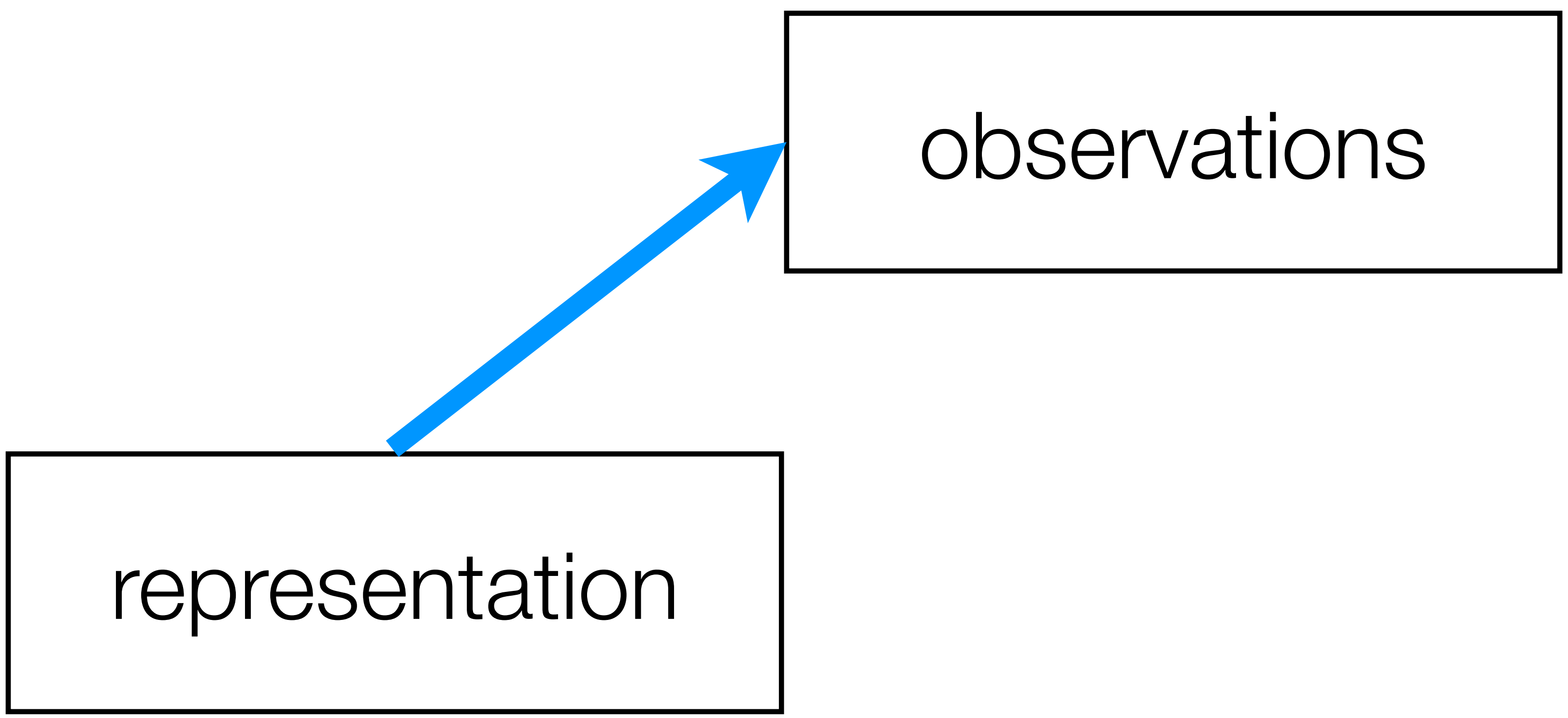


image data

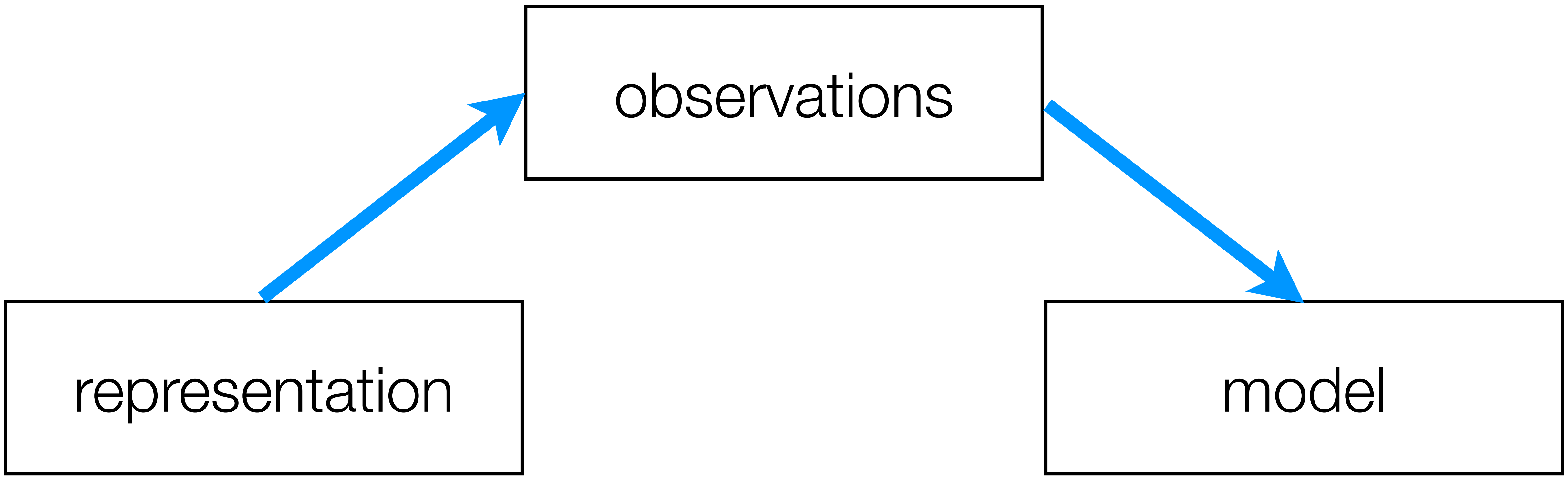
- calibrated - linearized response
- relatively large number

[van Hateren & van der Schaaf, 1998]

observations

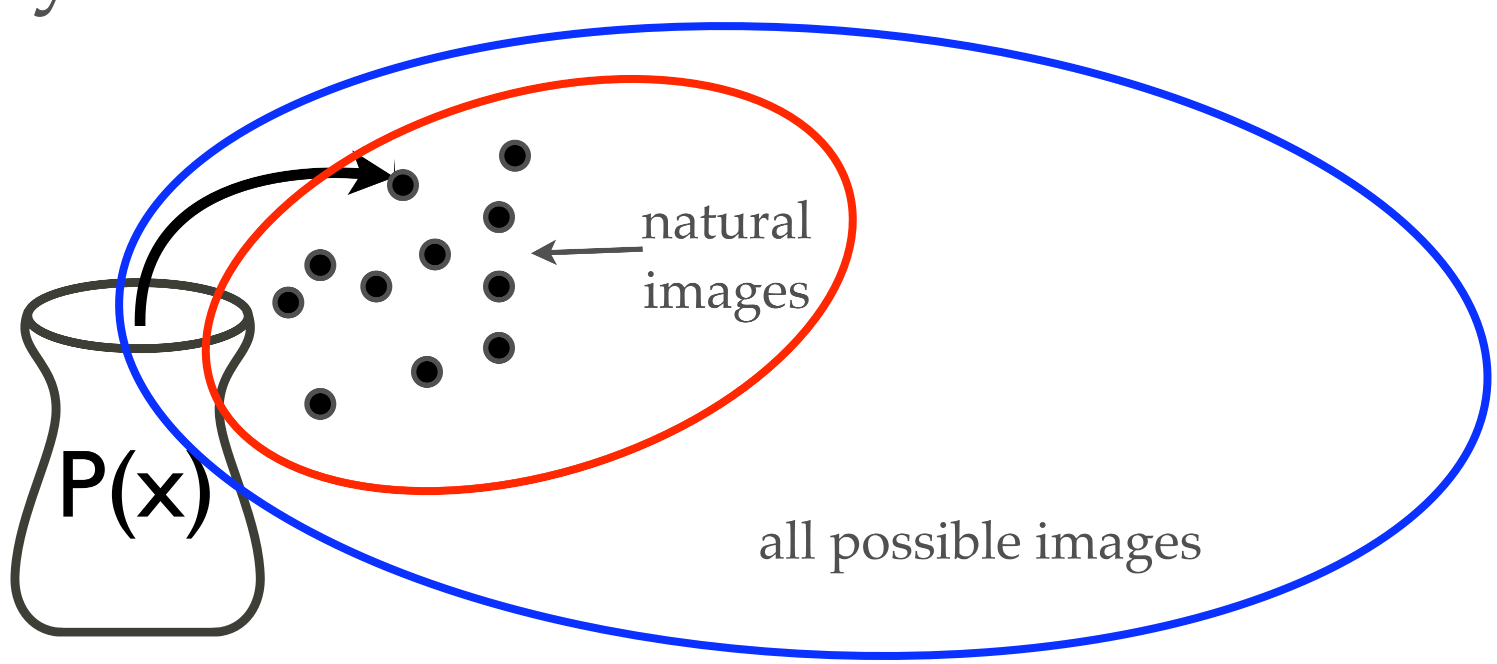
- second-order pixel correlations
- 1/f power law of frequency domain energy
- importance of phases
- heavy-tail non-Gaussian marginals in wavelet domain
- near elliptical shape of joint densities in wavelet domain
- decay of dependency in wavelet domain
- ………….

main components

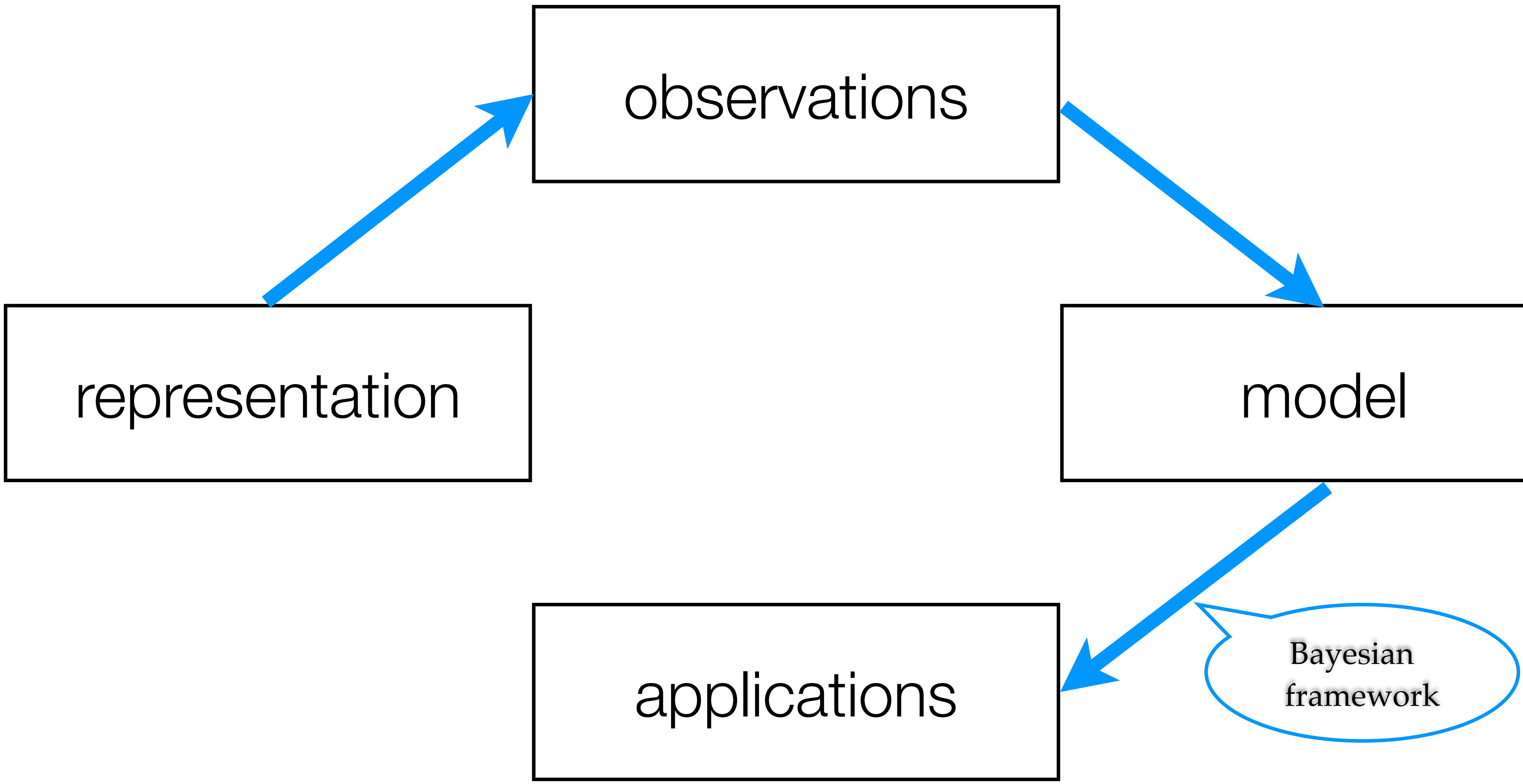


models

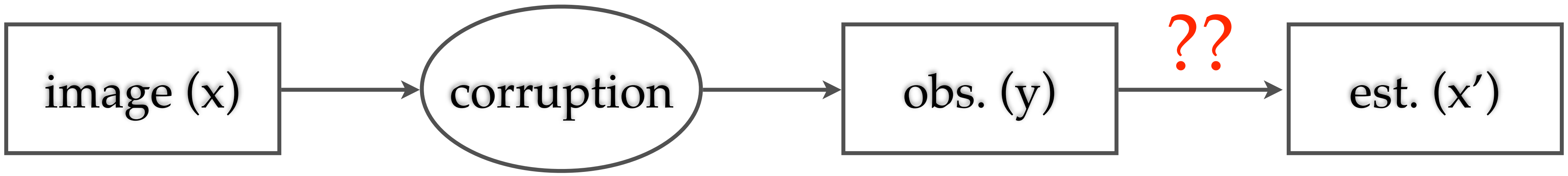
- physical imaging process (e.g., occlusion)
- nonlinear manifold of natural images
- non-parametric implicit model based on large set of images
- matching statistics of natural image signals with density models **<-- our focus**



main components



Bayesian framework



$$\min_{x'} \int_x L(x, x'(y))p(x|y)dx$$

$$\propto \quad \min_{x'} \int_x L(x, x'(y))p(y|x)p(x)dx$$

- x'(y): estimator
- L(x,x'(y)): loss functional
- p(x): prior model for natural images
- p(y | x): likelihood -- from corruption process

application: Bayesian denoising

- additive Gaussian noise

$$\begin{aligned} y &= x + w \\ p(y|x) &\propto \exp[-(y-x)^2/2\sigma_w^2] \end{aligned}$$

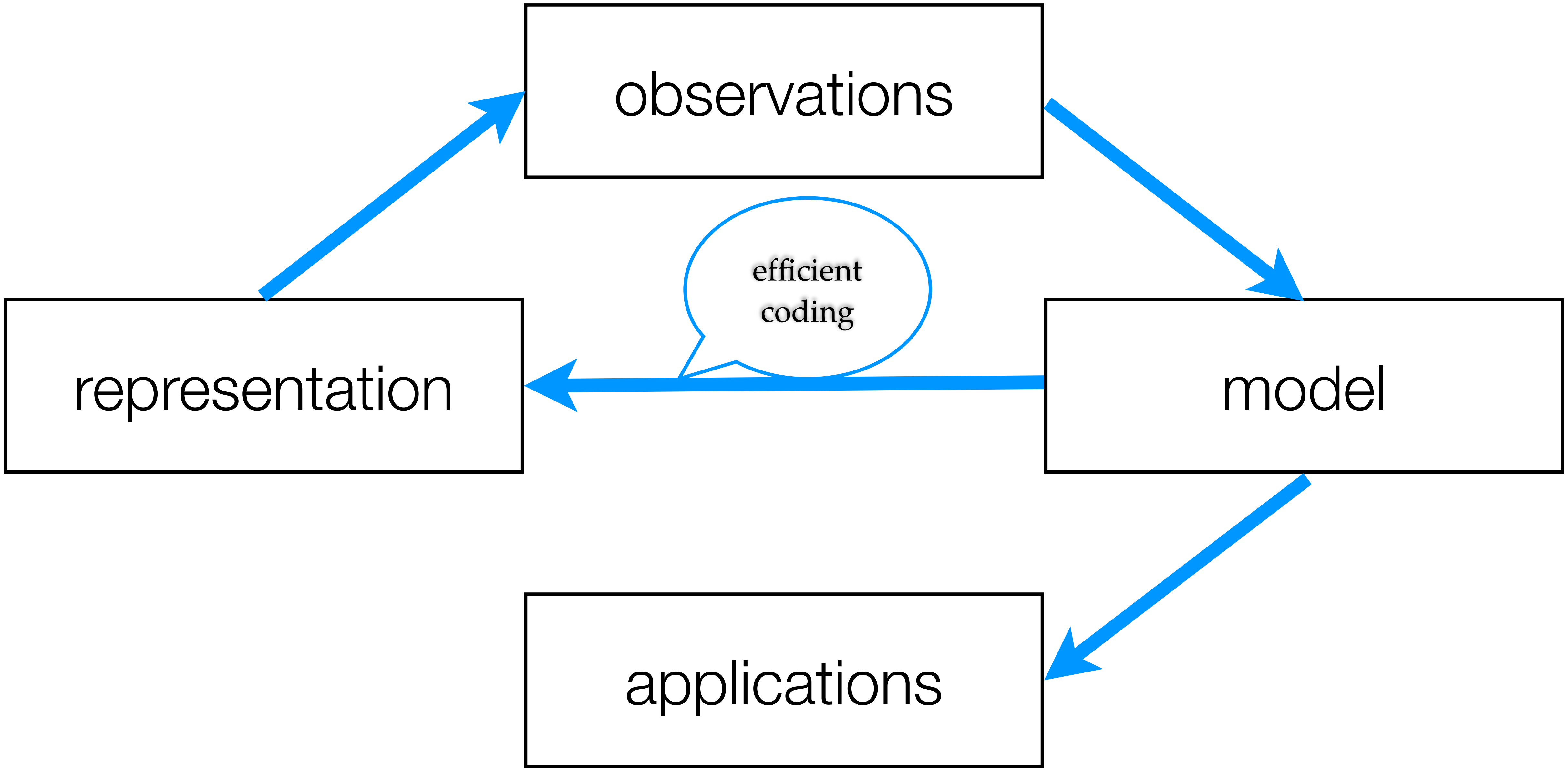
- maximum a posterior (MAP)

$$x_{\mathrm{MAP}} = \underset{x}{\operatorname{argmax}}\, p(x|y) = \underset{x}{\operatorname{argmax}}\, p(y|x)p(x)$$

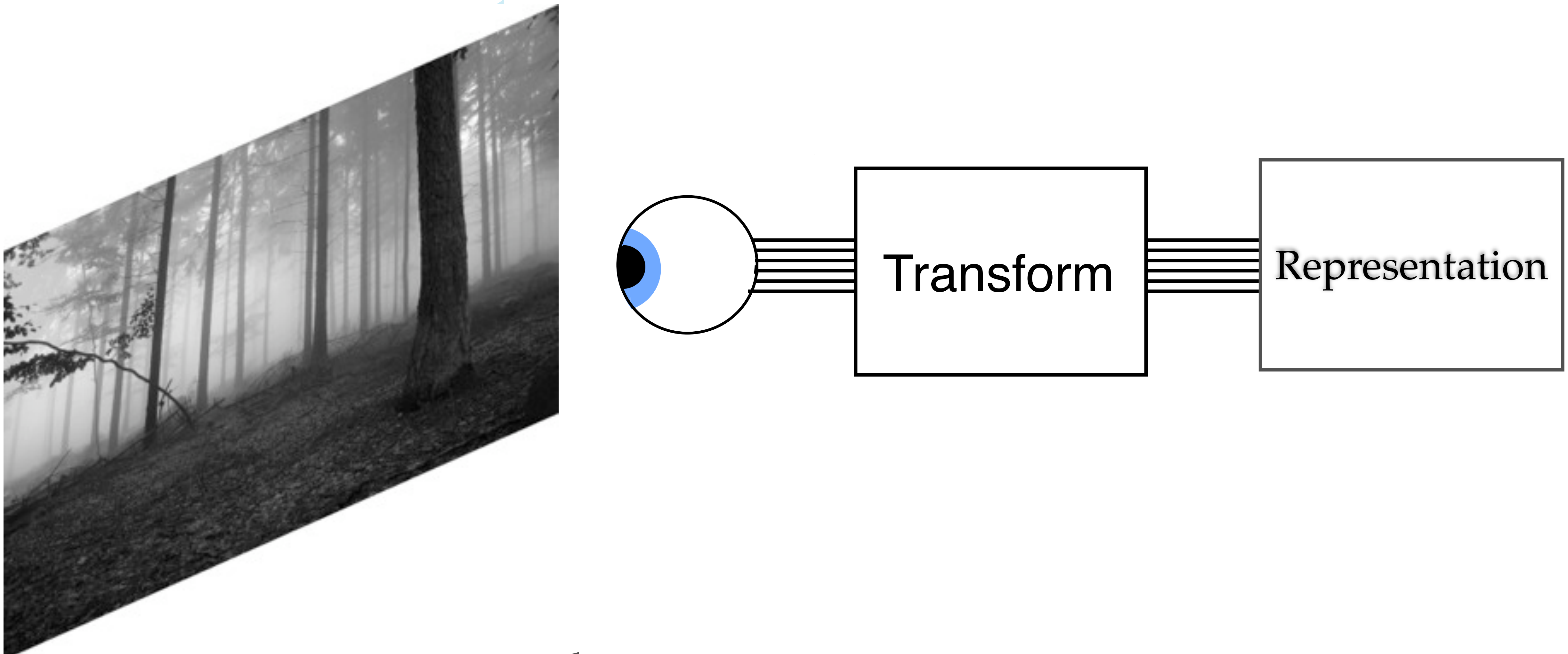
- minimum mean squares error (MMSE)

$$\begin{aligned} x_{\mathrm{MMES}} &= \underset{x'}{\operatorname{argmin}} \int_x \|x - x'\|^2 p(x|y)dx \\ &= \frac{\int_x xp(y|x)p(x)dx}{\int_x p(y|x)p(x)dx} = E(x|y) \end{aligned}$$

main components



representation

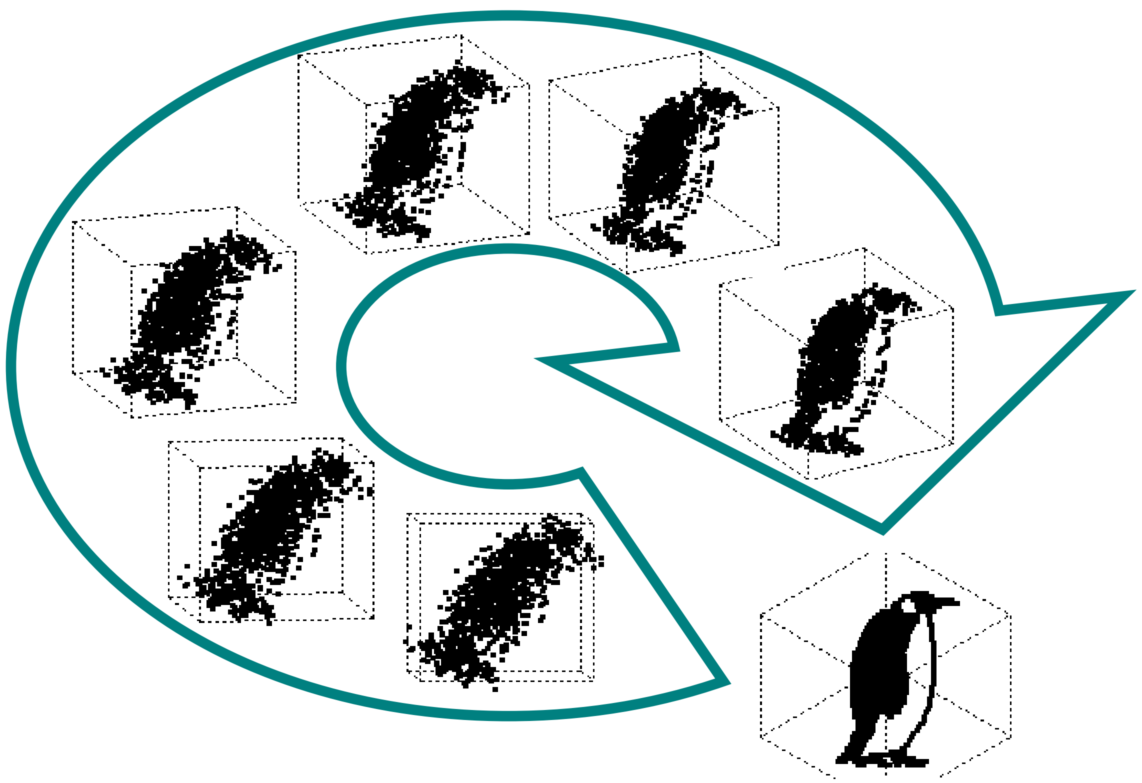


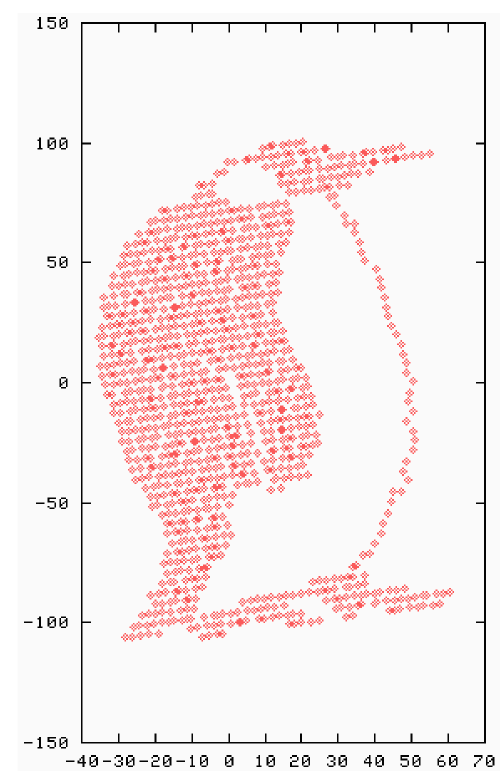
- unsupervised learning
- specify desired properties of the transform outputs

what are such properties?

what makes a good representation?

- intuitively, transformed signal should be "simpler"
  - reduced dimensionality





what makes a good representation?

- intuitively, transformed signal should be "simpler"
  - reduced dependency

r has less dependency than x

  - optimum: r is independent, $p(r) = \prod_{i=1}^{d} p(r_i)$
- reducing dependency is a general approach to relieve the curse of dimensionality
- are there dependency in natural images?

redundancy in natural images

- structure = predictability = redundancy

[Kersten, 1987]

measure of statistical dependency

multi-information (MI):

$$
\begin{aligned}
I(\vec{x}) &= D_{\mathrm{KL}}\left(p(\vec{x}) \,\middle\|\, \prod_k p(x_k)\right) \\
&= \int_{\vec{x}} p(\vec{x}) \log \frac{p(\vec{x})}{\prod_k p(x_k)} d\vec{x} \\
&= \sum_{i=1}^{d} H(x_k) - H(\vec{x})
\end{aligned}
$$

[Studeny and Vejnarova, 1998]

efficient coding

[Attneave ’54; Barlow ’61; Laughlin ’81; Atick ’90; Bialek etal ‘91]

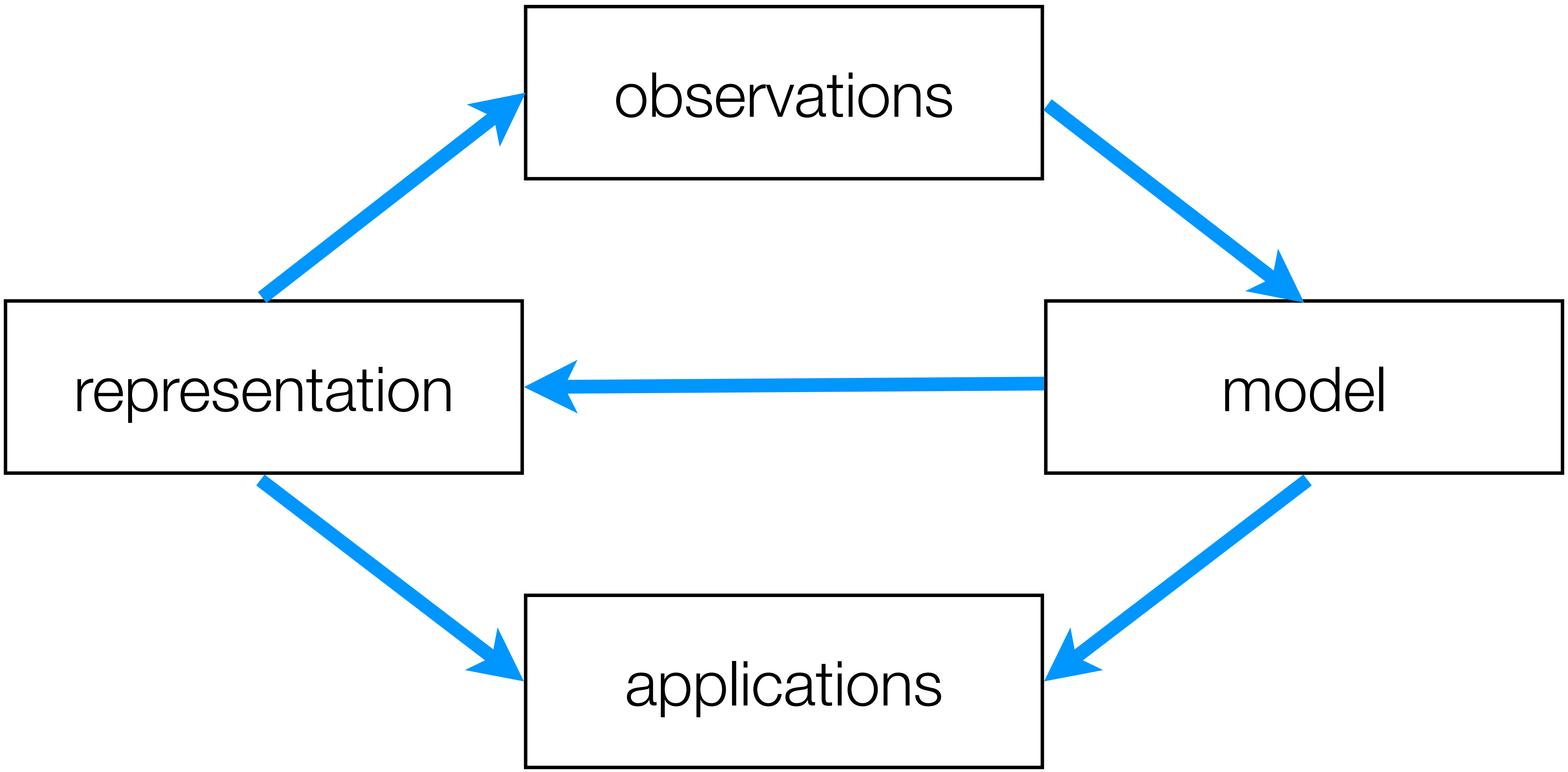
$$I(r,x) = H(r) - H(r|x)$$

- maximize mutual information of stimulus & response, subject to constraints (e.g. metabolic)
- noiseless case => redundancy reduction:

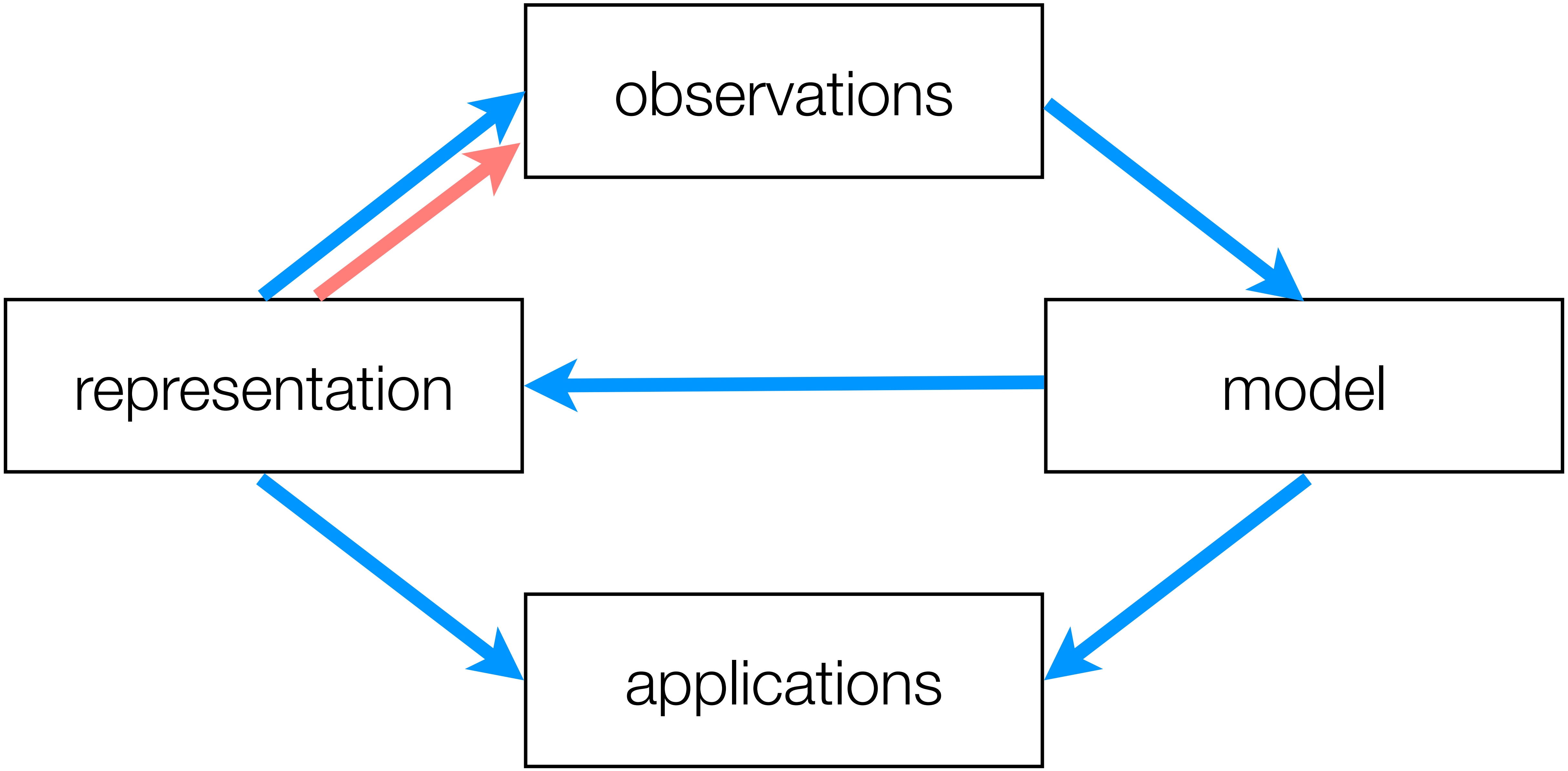
$$H(r|x) = 0 \Rightarrow I(r,x) = H(r) = \sum_{i=1}^{d} H(r_i) - I(r)$$

  - independent components
  - efficient (maxEnt) marginals

main components

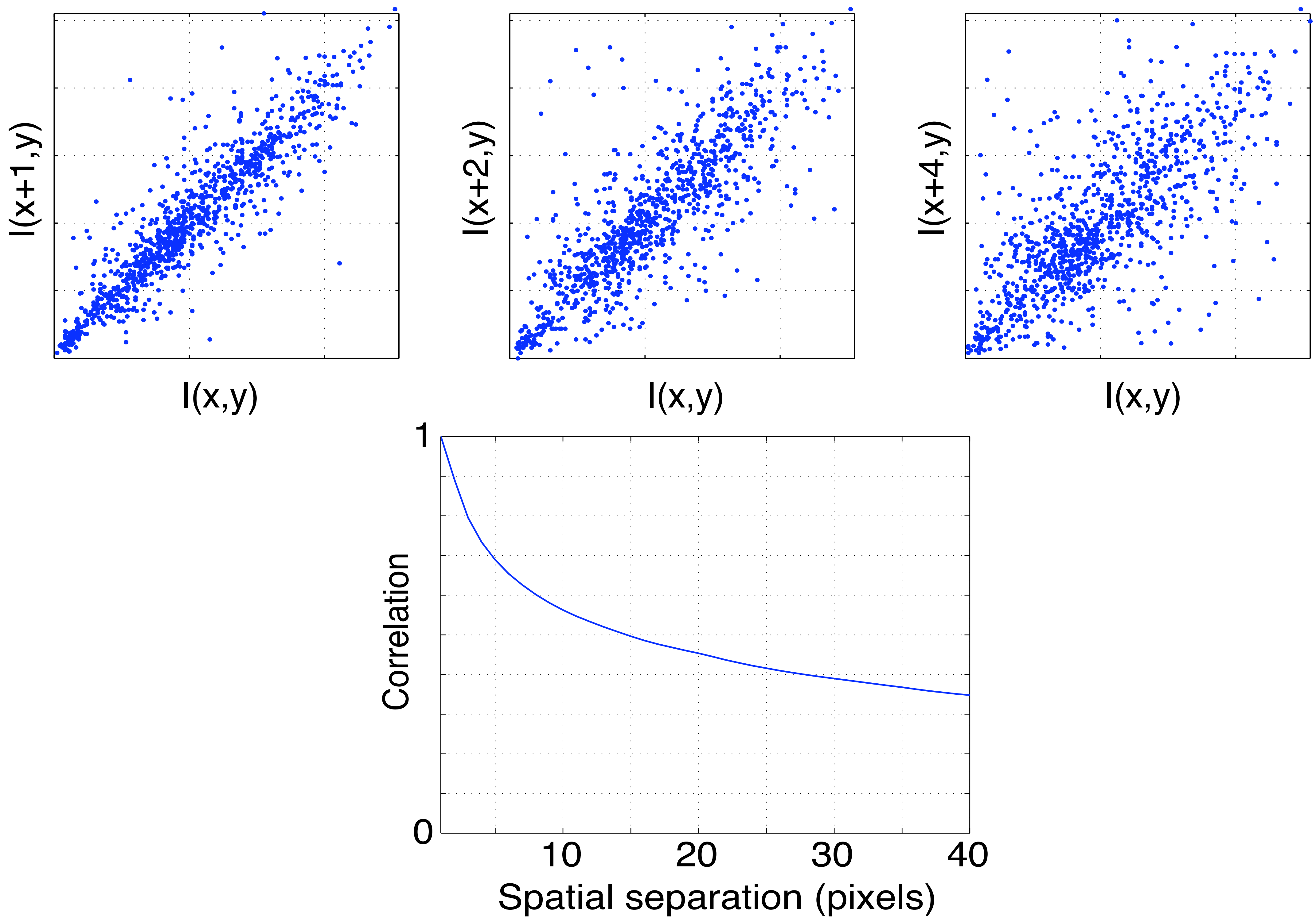


closed loop



pixel domain

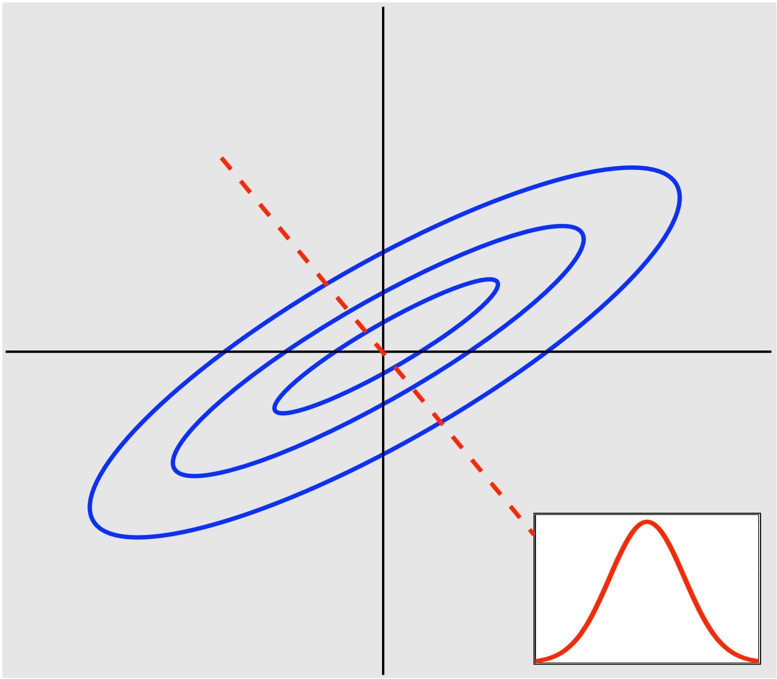
observation



model

- maximum entropy density [Jaynes 54]
  - assume zero mean
  - $\Sigma = E(\vec{x}\vec{x}^T)$: consistent w/ second order statistics
  - find $p(\vec{x})$ with maximum entropy
  - solution:

$$p(\vec{x}) \propto \exp\left(-\frac{1}{2}\vec{x}^T\Sigma^{-1}\vec{x}\right)$$



Gaussian model for Bayesian denoising

- additive Gaussian noise

$$\begin{aligned} \vec{y} &= \vec{x} + \vec{w} \\ p(\vec{y}|\vec{x}) &\propto \exp[-\|\vec{y} - \vec{x}\|^2 / 2\sigma_w^2] \end{aligned}$$

- Gaussian model

$$p(\vec{x}) \propto \exp\left(-\frac{1}{2}\vec{x}^T\Sigma^{-1}\vec{x}\right)$$

- posterior density (another Gaussian)

$$p(\vec{x}|\vec{x}) \propto \exp\left(-\frac{1}{2}\vec{x}^T\Sigma^{-1}\vec{x} - \frac{\|\vec{x} - \vec{y}\|^2}{2\sigma_w^2}\right)$$

- inference (Wiener filter)

$$\vec{x}_{\mathrm{MAP}} = \vec{x}_{\mathrm{MMSE}} = \Sigma(\Sigma + \sigma_w^2 I)^{-1}\vec{y}$$

efficient coding transform

- for Gaussian p(x)

$$I(\vec{x}) \propto \sum_{i=1}^{d} \log(\Sigma)_{ii} - \log\det(\Sigma)$$

- minimum (independent) when Σ is diagonal
- a transform that ***diagonalizes*** Σ can eliminate all dependencies (second-order)

PCA

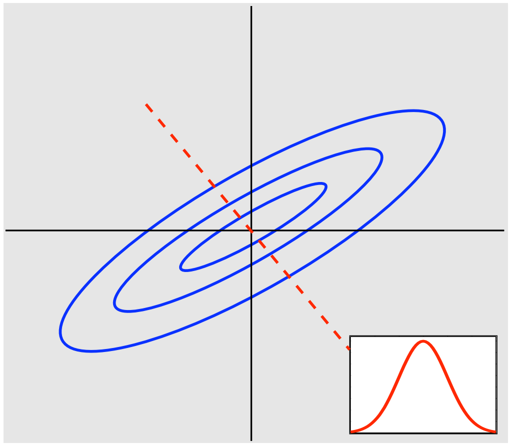
- eigen-decomposition of Σ: $\Sigma = U \Lambda U^T$
  - U: orthonormal matrix (rotation)
    $U^T U = U U^T = I$
  - Λ: diagonal matrix, $\Lambda_{ii} \geq 0$ -- eigenvalue

$$\begin{aligned} E\{U^T \vec{x}(U^T \vec{x})^T\} &= U^T E\{\vec{x}\vec{x}^T\} U \\ &= U^T U \Lambda U^T U = \Lambda \end{aligned}$$

- $s = U^T x$, or $x = Us$, s is independent Gaussian

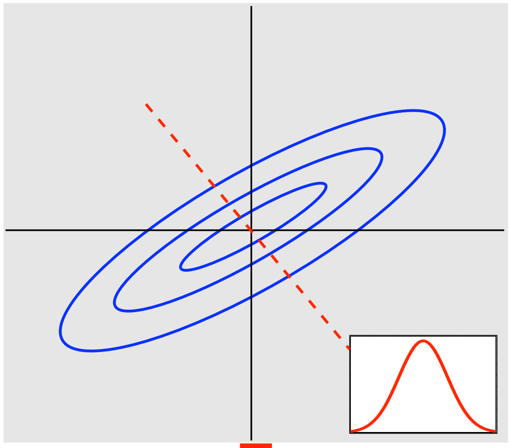
- principal component analysis (PCA)
  - Karhunen Loeve transform

PCA

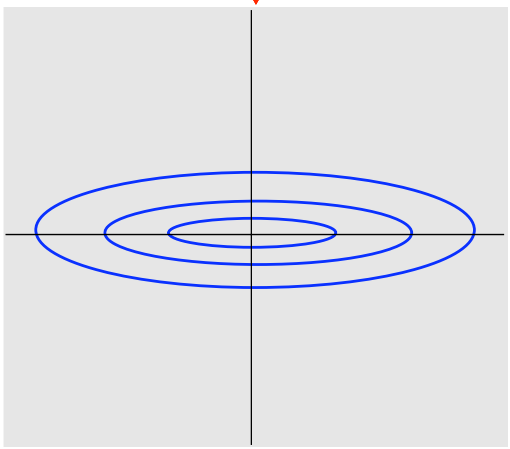


$\vec{x}$

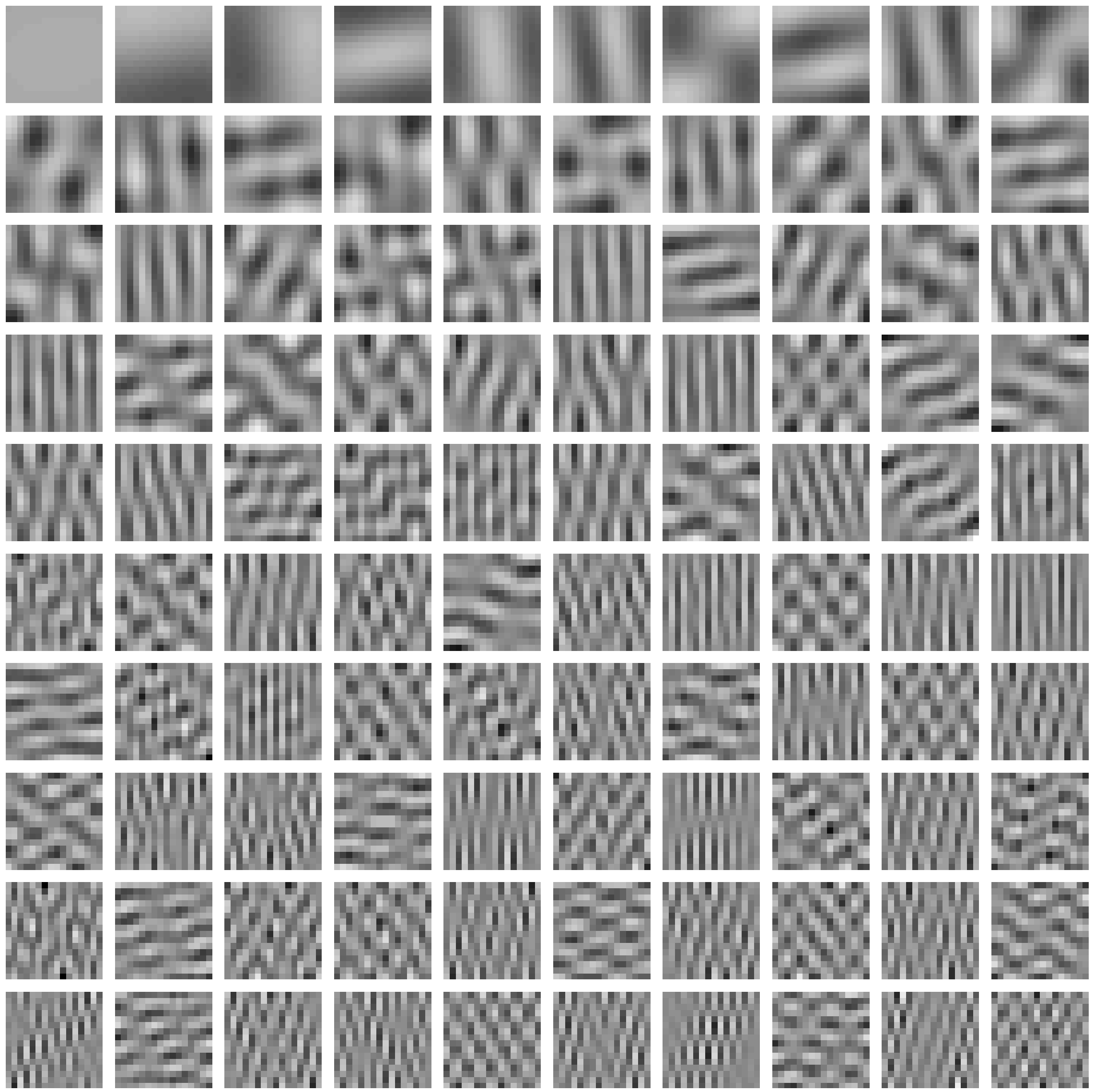
PCA



$\vec{x}$



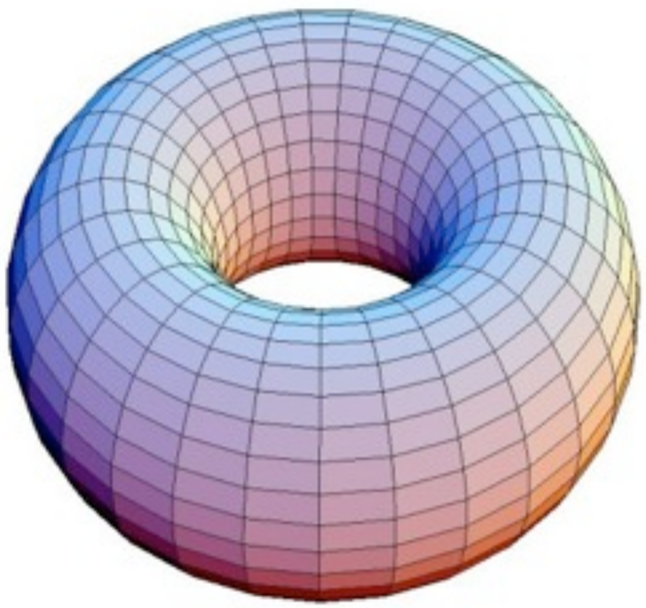
$$\vec{x}_{\mathrm{PCA}} = U^T \vec{x}$$



PCA bases learned from natural images (U)

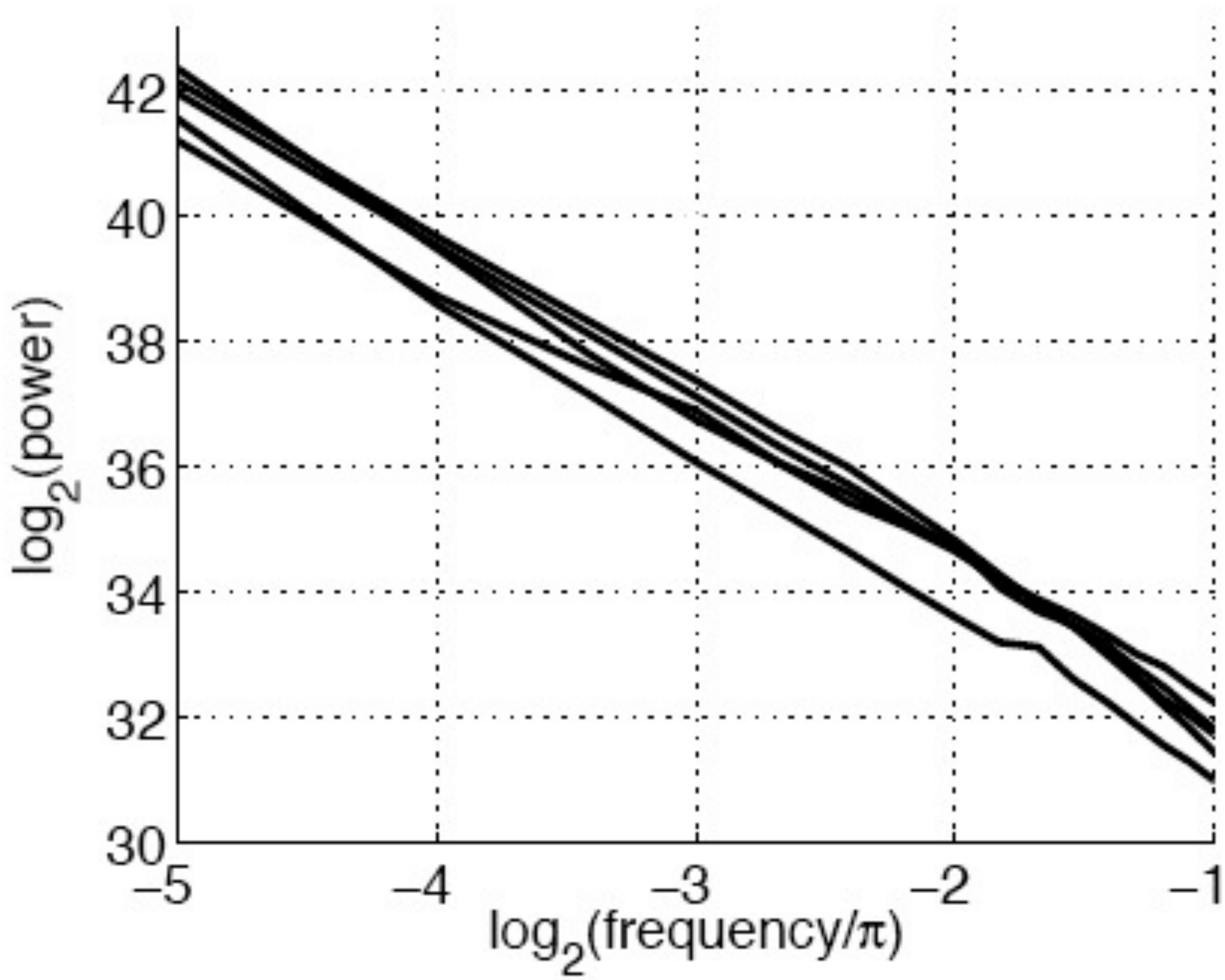
representation

- PCA is for local patches
  - data dependent
  - expensive for large images
- assume translation invariance
  cyclic boundary handling
  - image lattice on a torus
    - covariance matrix is block circulant
      - eigenvectors are complex exponential
        - diagonalized (decorrelated) with DFT
          - PCA => Fourier representation



observations

- spectral power



[Ritterman 52; DeRiugin 56; Field 87; Tolhurst 92; Ruderman/Bialek 94; ...]

figure from [Simoncelli 05]

model

- power law

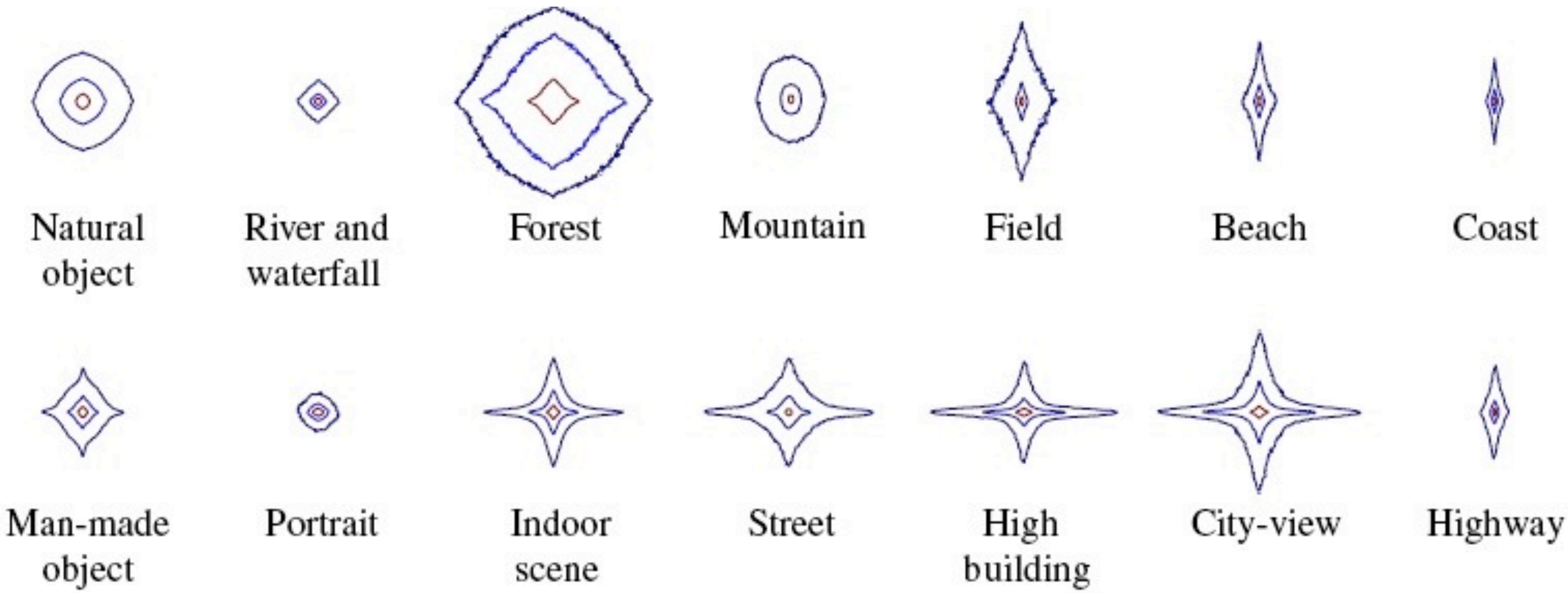
$$F(\omega) = \frac{A}{\omega^\gamma}$$

- scale invariance $F(s\omega) = s^p F(\omega)$

- denoising (Wiener filter in frequency domain)

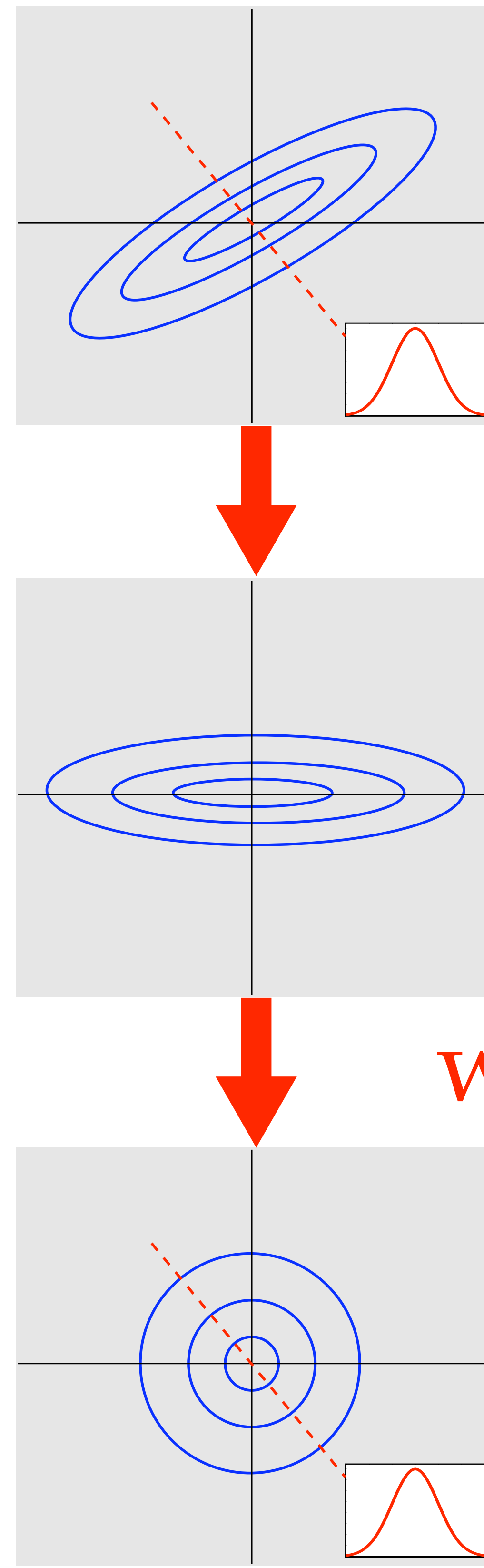
$$\hat{X}(\omega) = \frac{A/\omega^\gamma}{A/\omega^\gamma + \sigma^2} \cdot Y(\omega)$$

further observations



[Torralba and Oliva, 2003]

PCA



$\vec{x}$

$$\Sigma = U \Lambda U^T$$

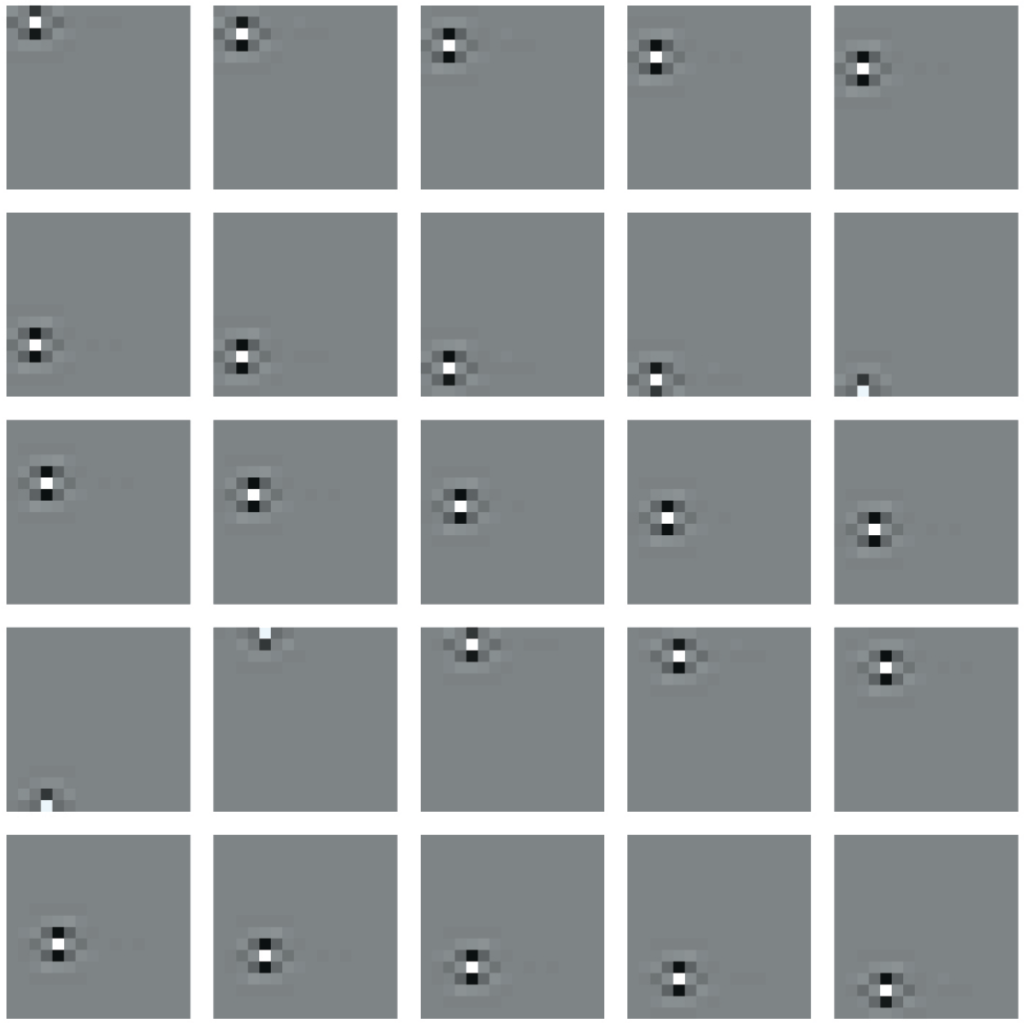
$$U^T \vec{x}$$

whitening

$$\Lambda^{-\frac{1}{2}} U^T \vec{x}$$

not unique! $V \Lambda^{-\frac{1}{2}} U^T \vec{x}$

zero-phase (symmetric) whitening (ZCA)



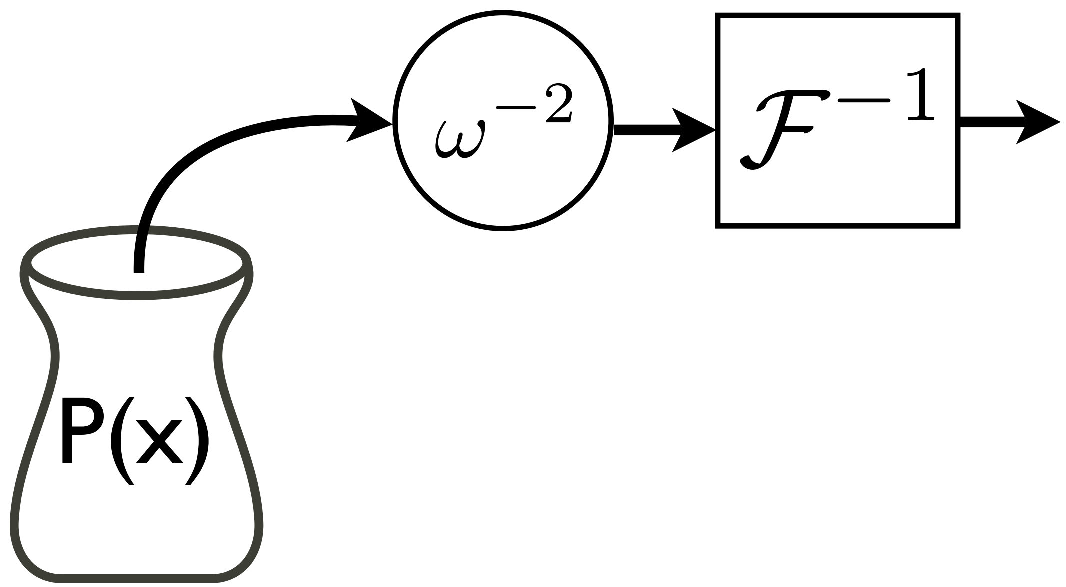
$$A^{-1} = U\Lambda^{-\frac{1}{2}}U^T$$

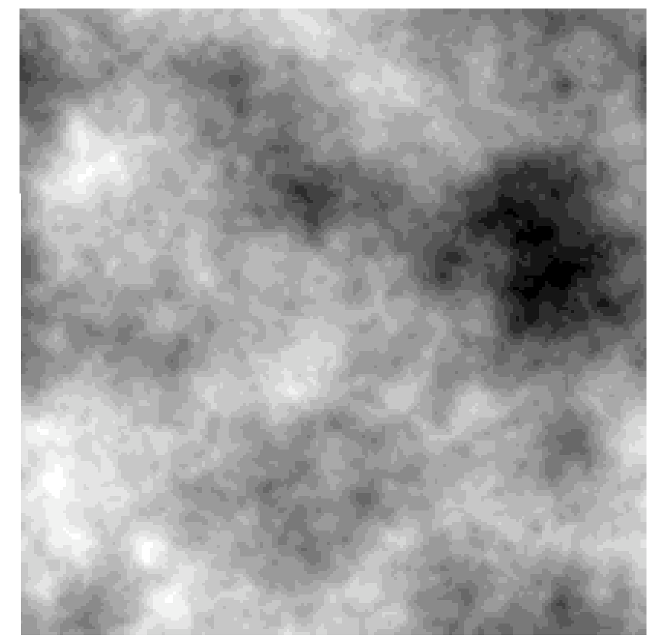
minimum wiring length

receptive fields of retina neurons [Atick & Redlich, 92]

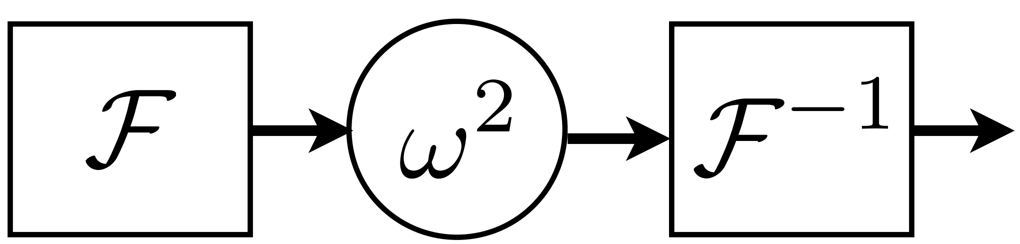
second-order constraints are weak

$1/\varpi^{\gamma}$ Gaussian sample





whitened natural image



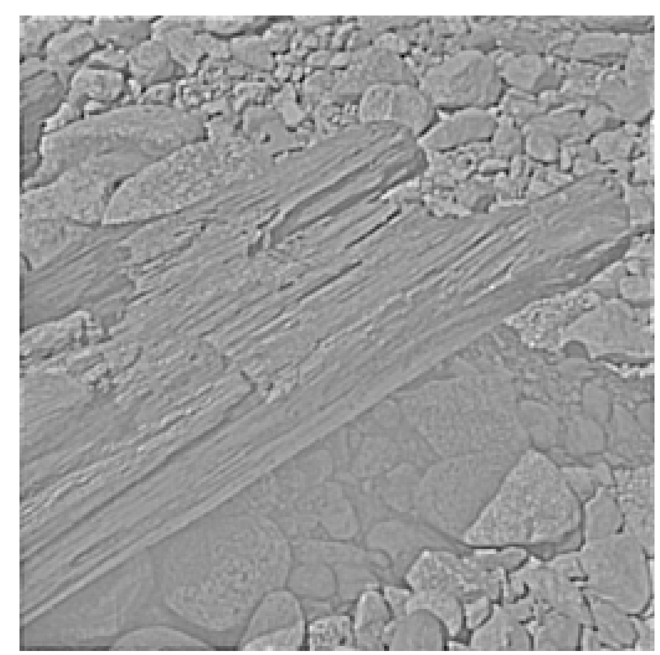
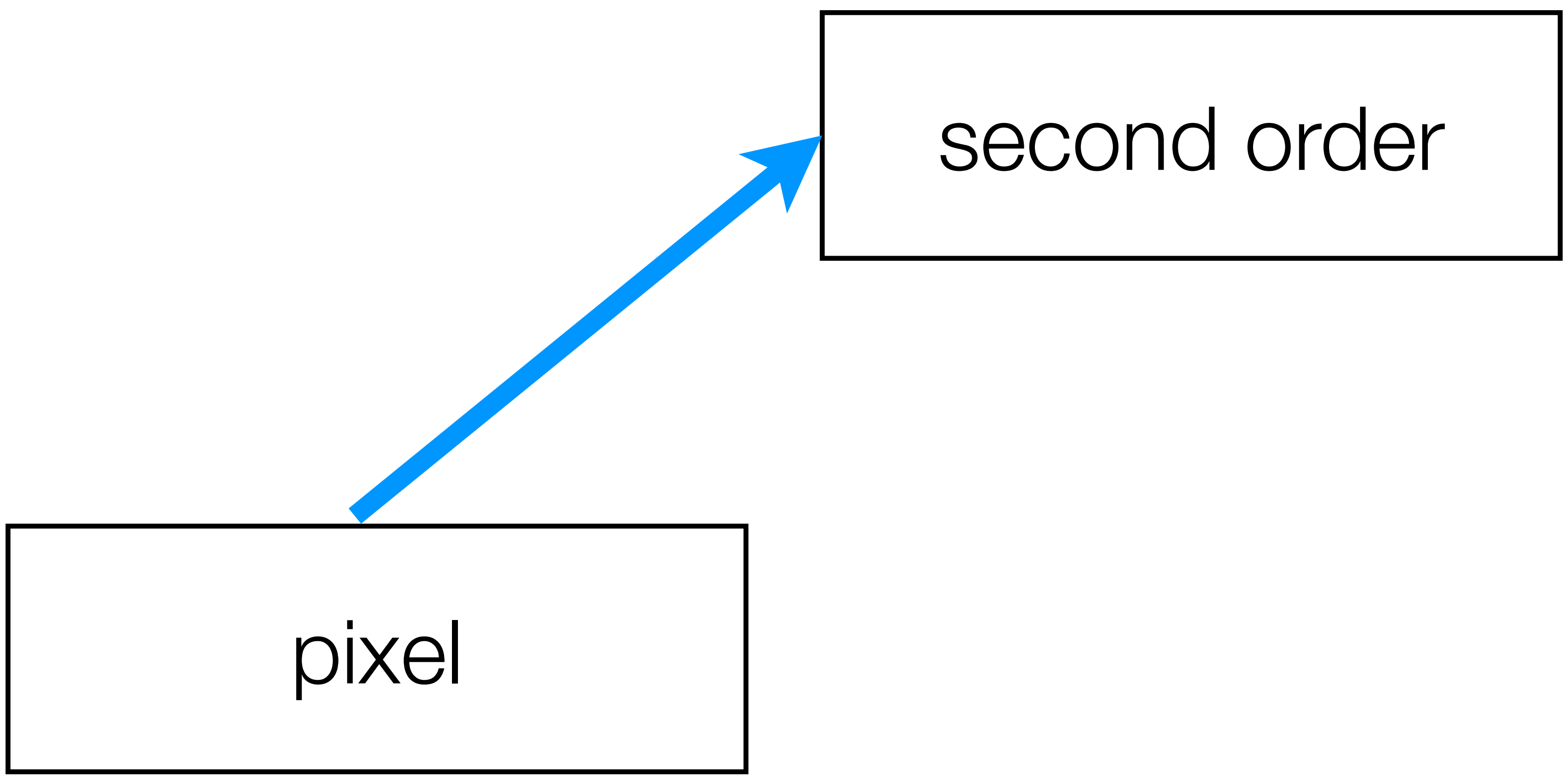


figure courtesy of Eero Simoncelli

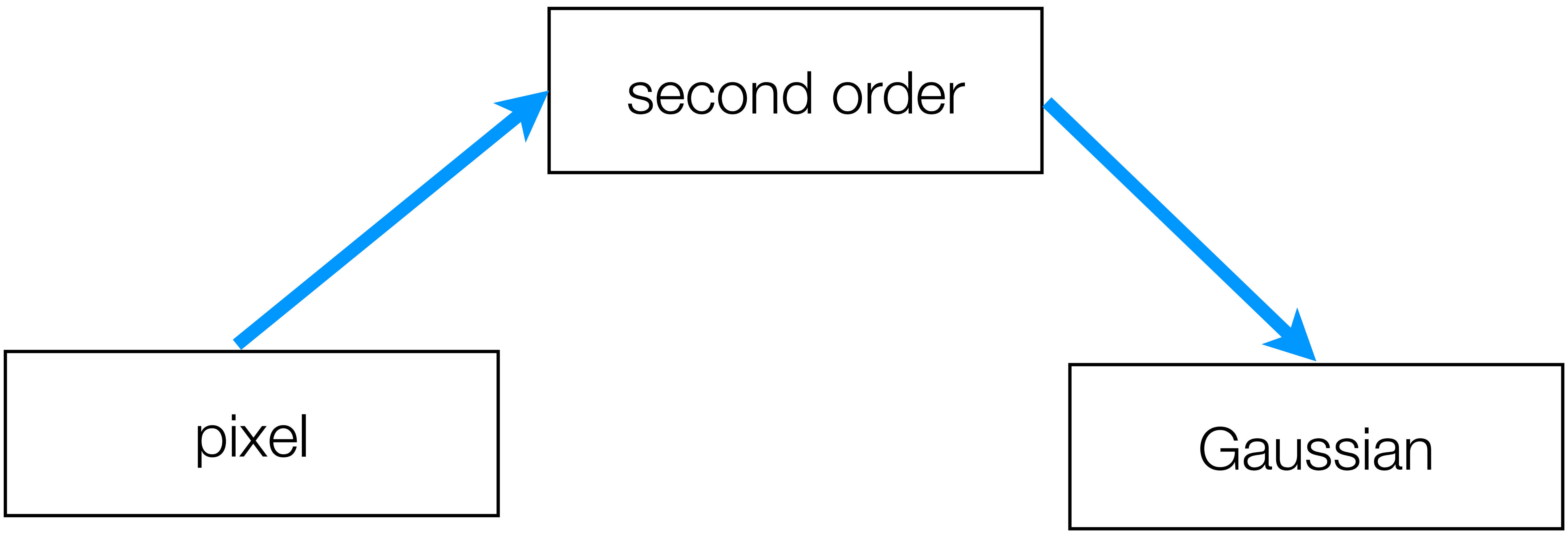
summary

pixel

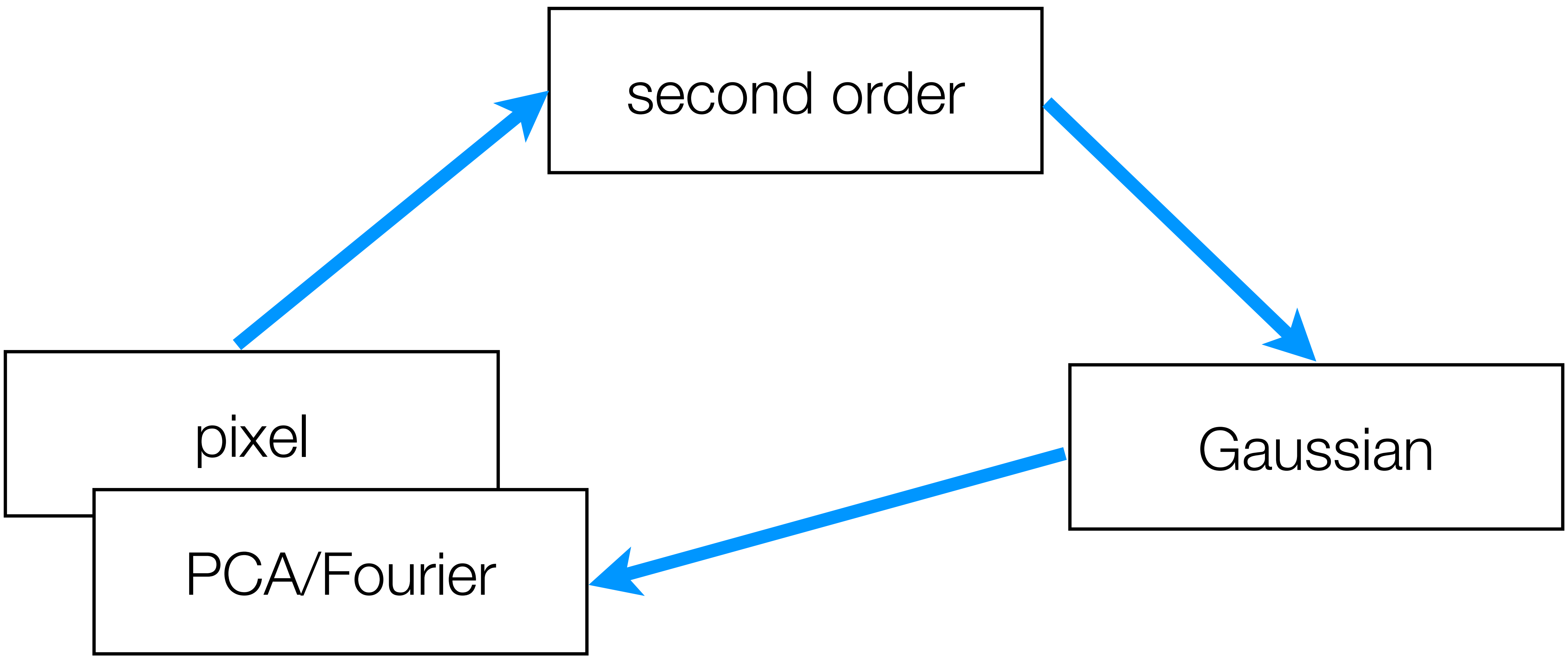
summary



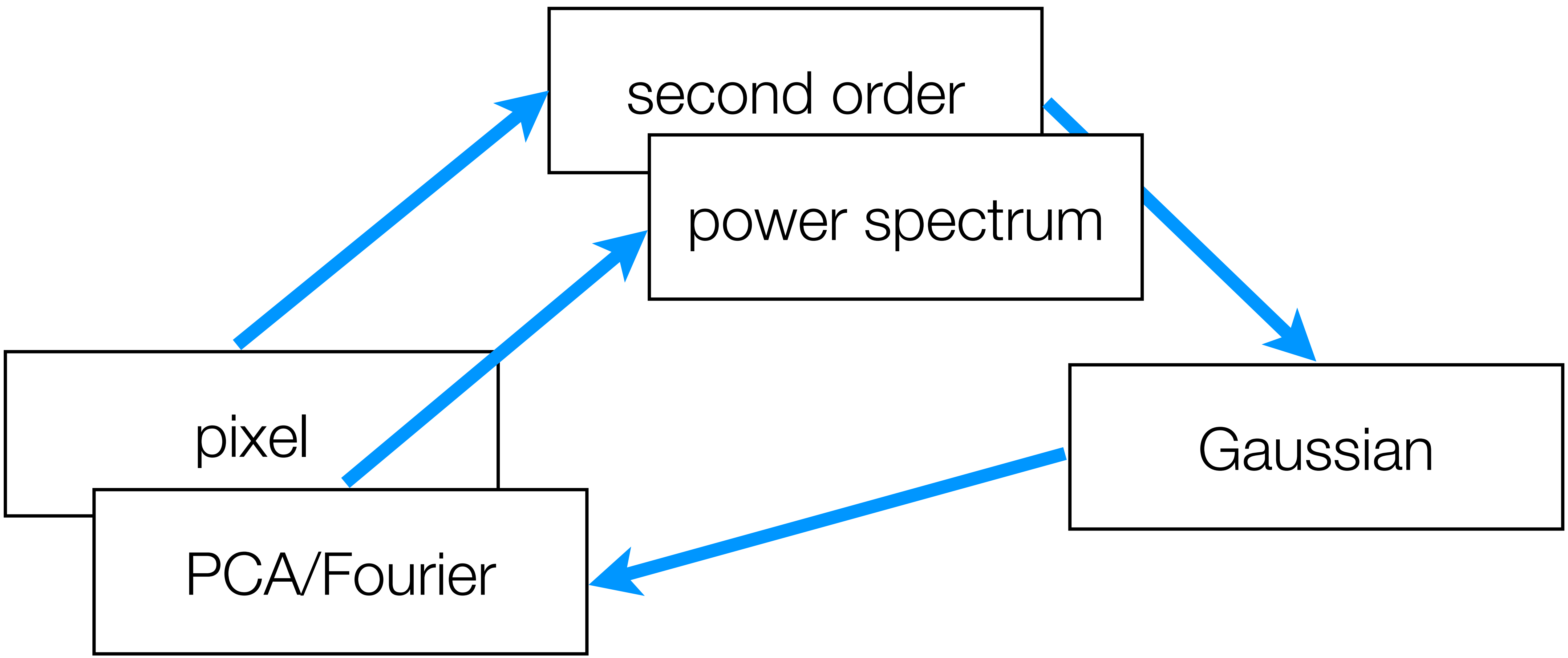
summary



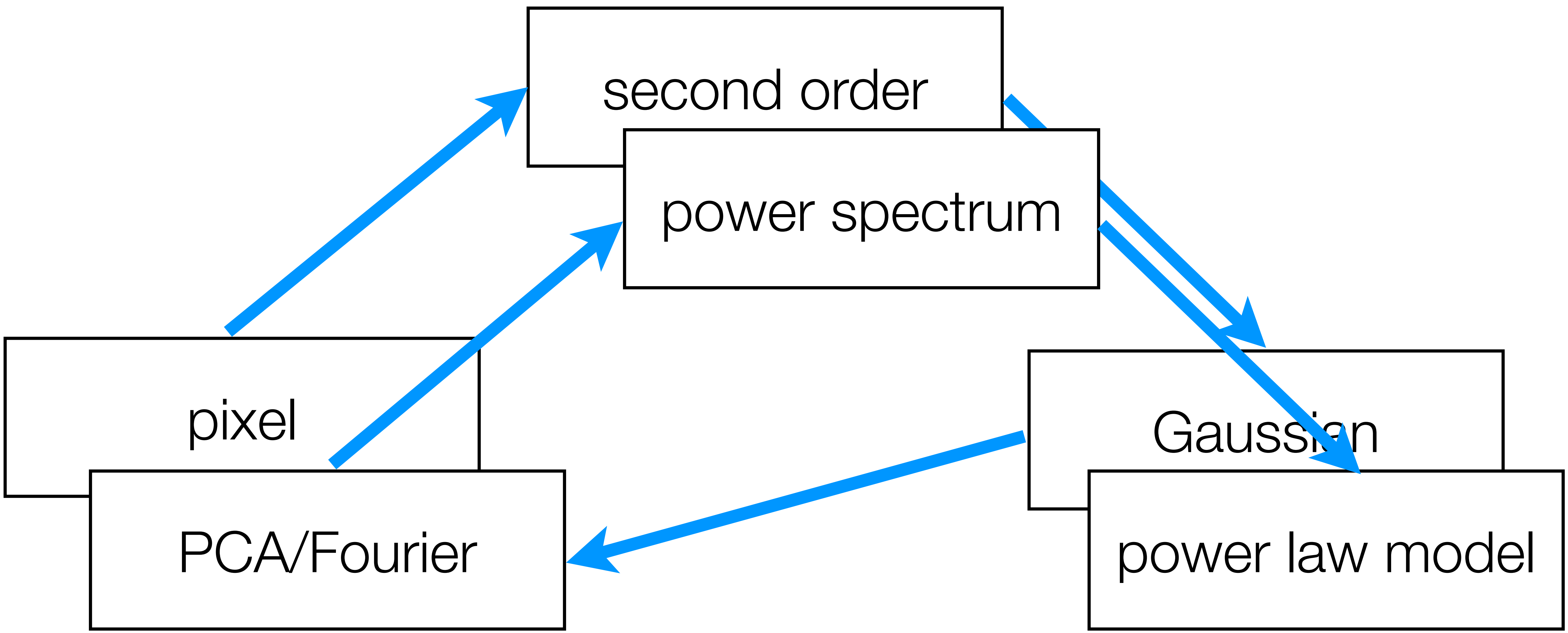
summary



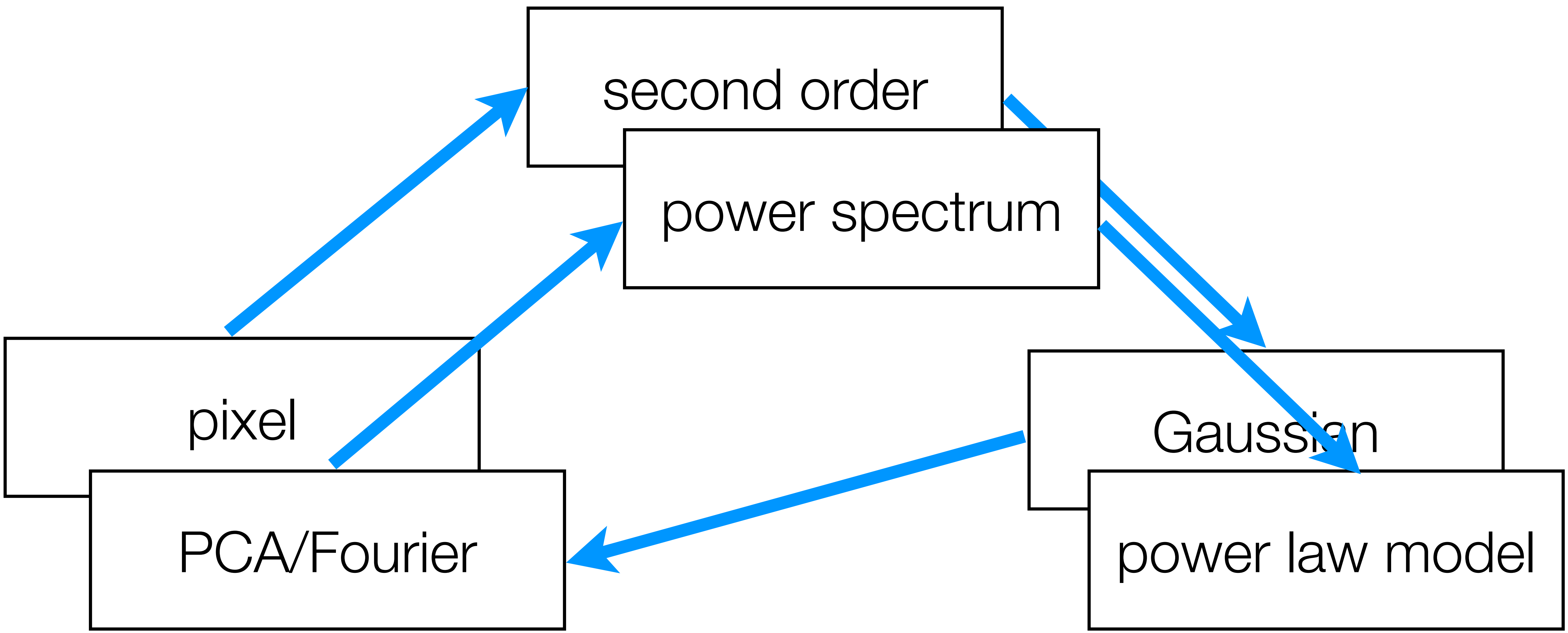
summary



summary


second order
power spectrum
pixel
PCA/Fourier
Gaussian
power law model

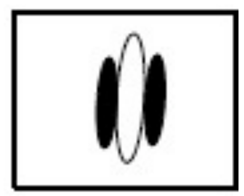
summary



Not enough!

bandpass filter domain

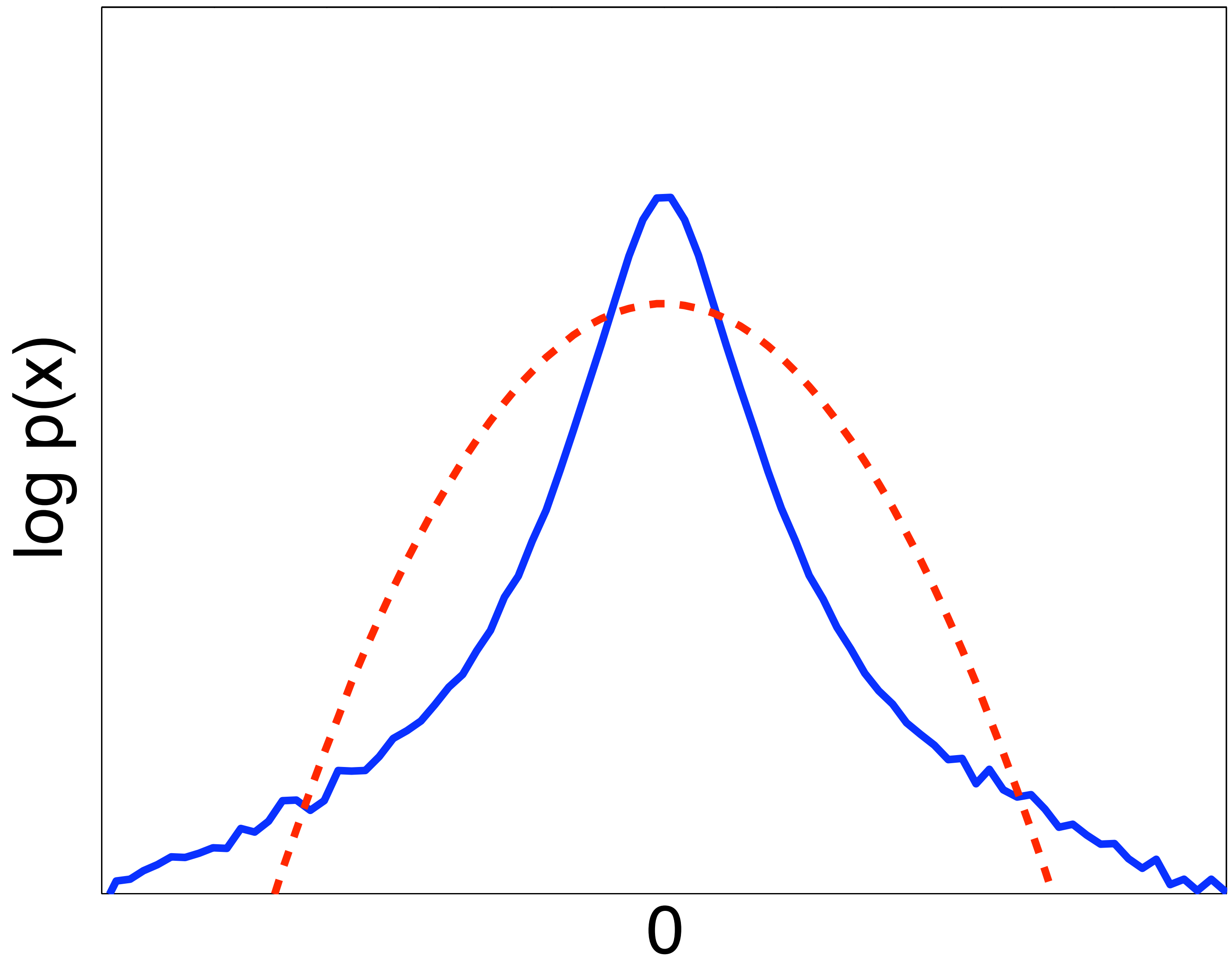
$\otimes$



$=$

observation

- sparseness



[Burt&Adelson 82; Field 87; Mallat 89; Daugman 89, ...]

model

- if we only enforce consistency on 1D marginal densities, i.e., $p(x_i) = q_i(x_i)$
  - maximum entropic density is the *factorial* density $p(\vec{x}) = \prod_{i=1}^{d} q_i(x_i)$

  - multi-information is non-negative, and achieves minimum (zero) when x_is are independent $H(\vec{x}) = \sum_i H(x_i) - I(\vec{x})$
- there are second order dependencies, so derived model is a ***linearly transformed factorial*** (LTF) model

model

- linearly transformed factorial (LTF)
  - independent sources: $p(\vec{s}) = \prod_{i=1}^{d} p(s_i)$
  - A: invertible linear transform (basis)

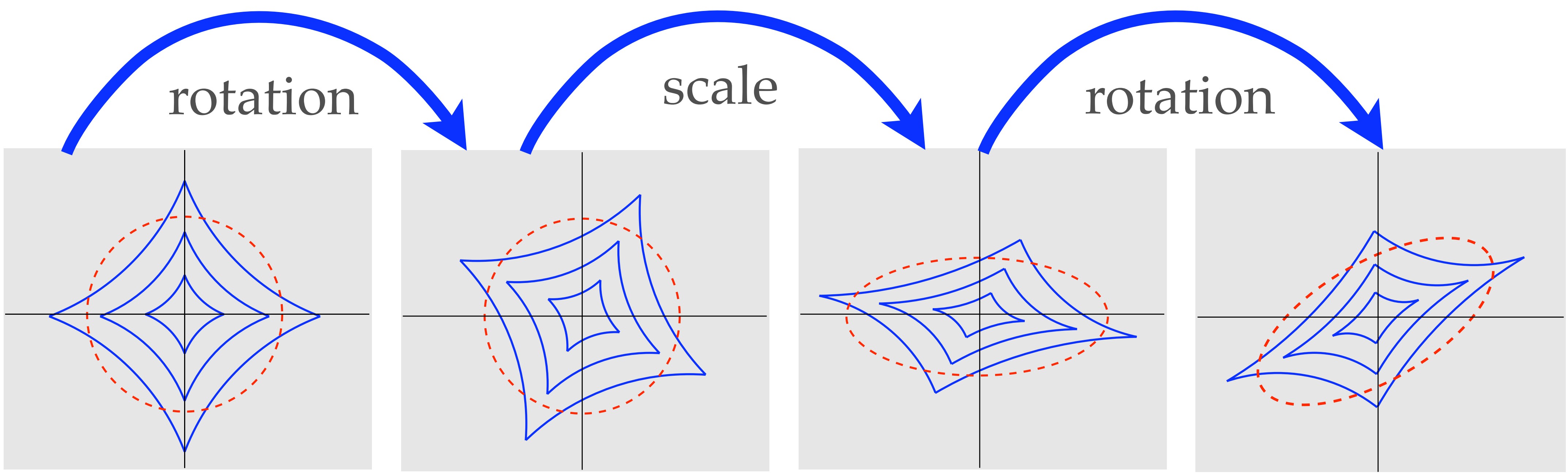
$$
\begin{aligned}
\vec{x} &= A\vec{s} = \begin{pmatrix} | & \cdots & | \\ \vec{a}_1 & \cdots & \vec{a}_d \\ | & \cdots & | \end{pmatrix} \begin{pmatrix} s_1 \\ \vdots \\ s_d \end{pmatrix} \\
&= s_1\vec{a}_1 + \cdots + s_d\vec{a}_d
\end{aligned}
$$

  - A^{-1}: filters for analysis

$$\vec{s} = A^{-1}\vec{x}$$

LTF model

- SVD of matrix A: $A = U\Lambda^{1/2}V^T$
  - U,V: orthonormal matrices (rotation)
    $U^TU = UU^T = I$ and $V^TV = VV^T = I$
  - Λ: diagonal matrix
    $(\Lambda_{ii})^{1/2} \geq 0$ -- singular value



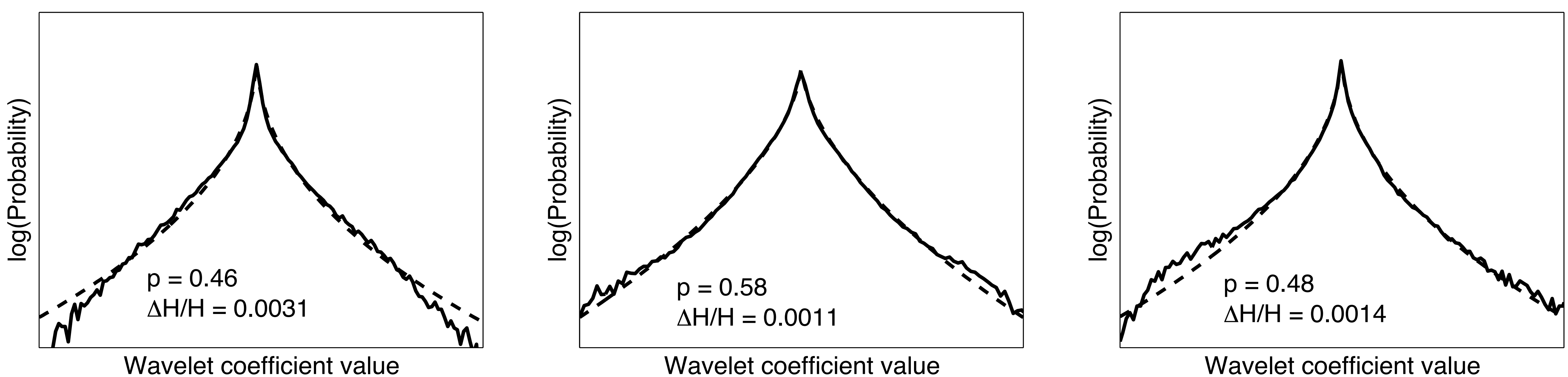
s | V^Ts | $\Lambda^{1/2}V^Ts$ | $x = U\,\Lambda^{1/2}V^Ts$

marginal model

- well fit with generalized Gaussian

$$p(s) \propto \exp\left(-\frac{|s|^p}{\sigma}\right)$$

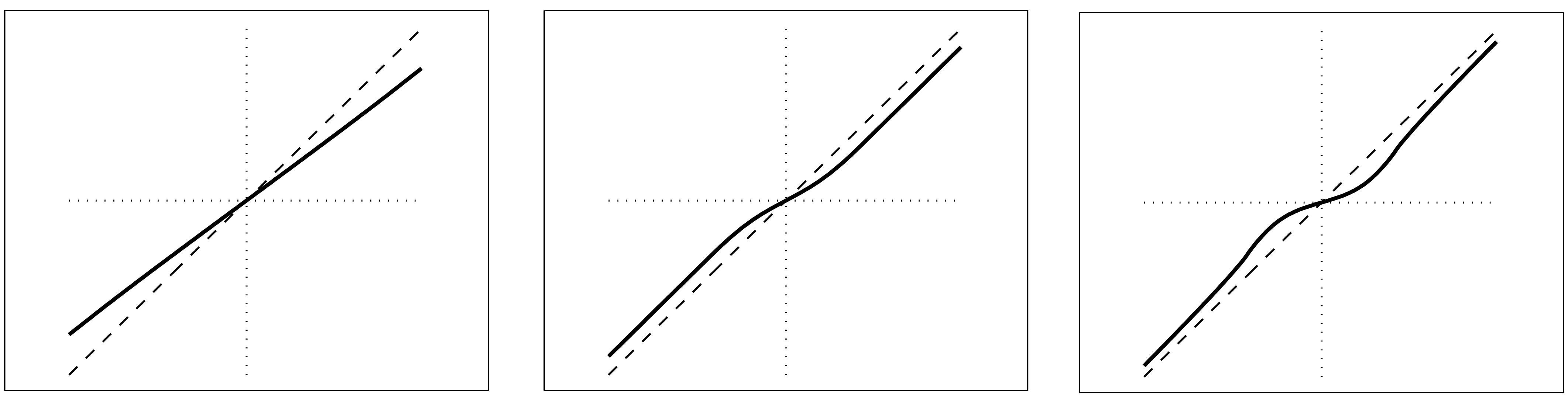
[Mallat 89; Simoncelli&Adelson 96; Moulin&Liu 99; …]



Bayesian denoising

$$\hat{x}(y) = \int dx\ \mathcal{P}_{x|y}(x|y)\ x = \frac{\int dx\ \mathcal{P}_{y|x}(y|x)\ \mathcal{P}_x(x)\ x}{\int dx\ \mathcal{P}_{y|x}(y|x)\ \mathcal{P}_x(x)}$$

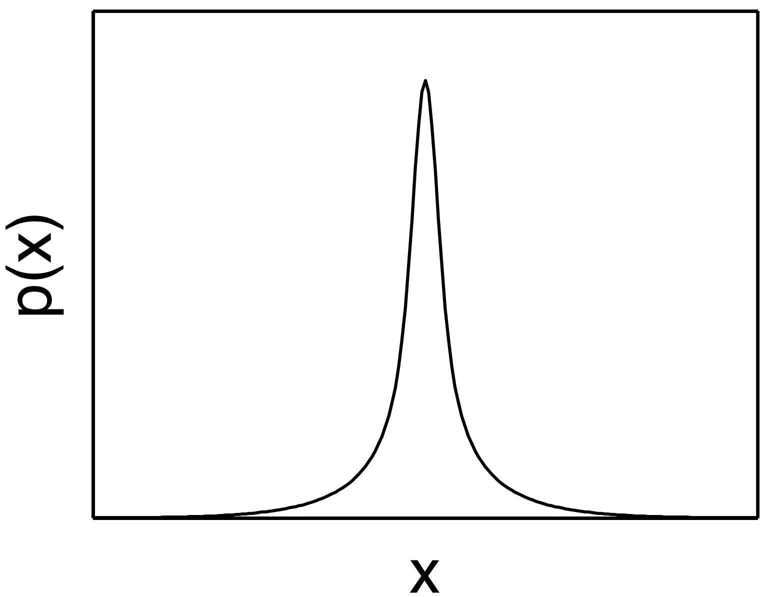
$$= \frac{\int dx\ \mathcal{P}_n(y-x)\ \mathcal{P}_x(x)\ x}{\int dx\ \mathcal{P}_n(y-x)\ \mathcal{P}_x(x)},$$



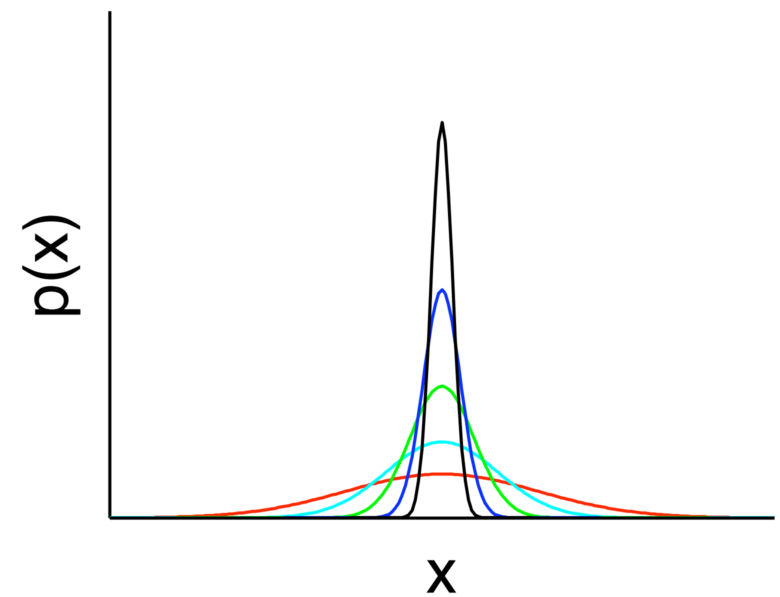
$p = 2.0$ $p = 1.0$ $p = 0.5$

[Simoncelli & Adelson, '96]

scale mixture of Gaussians (GSM)



=



[Andrews & Mallows 74, Wainwright & Simoncelli, 99]

$$x = u\sqrt{z}$$

- u: zero mean Gaussian with unit variance
- z: positive random variable
- special cases (different p(z))

generalized Gaussian, Student's t, Bessel's K, Cauchy, α-stable, etc

efficient coding transform

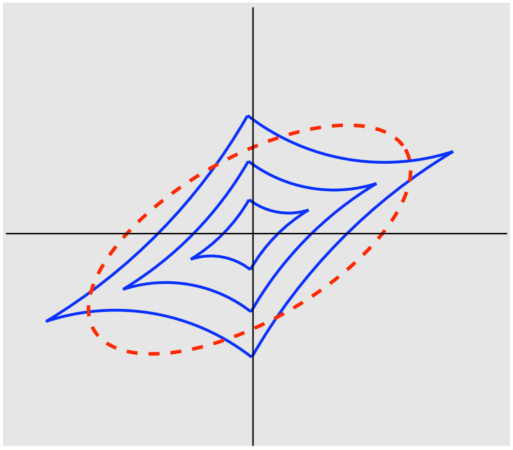
- LTF model => independent component analysis (ICA)
  [Comon 94; Cardoso 96; Bell/Sejnowski 97; …]
  - many different implementations (JADE, InfoMax, FastICA, etc.)
  - interpretation using SVD

$$\vec{s} = A^{-1}\vec{x} = V\Lambda^{-1/2}U^T\vec{x}$$

  - where to get U

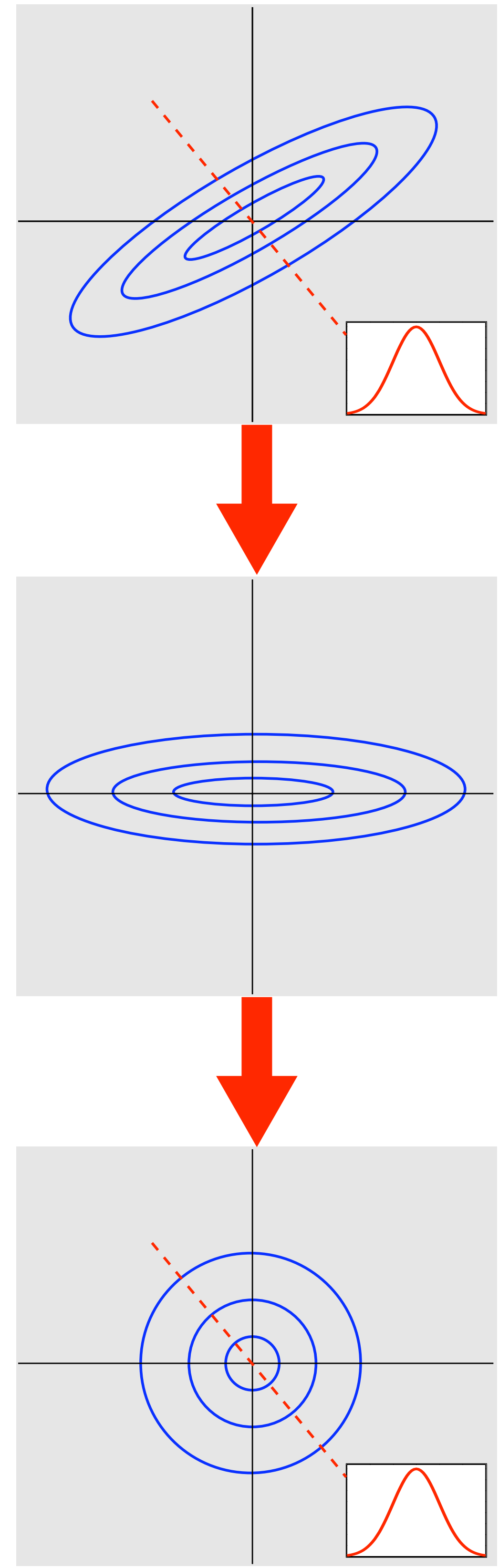
$$\begin{aligned} E\{\vec{x}\vec{x}^T\} &= AE\{\vec{s}\vec{s}^T\}A^T \\ &= U\Lambda^{1/2}V^T I V\Lambda^{1/2}U^T \\ &= U\Lambda U^T \end{aligned}$$

ICA

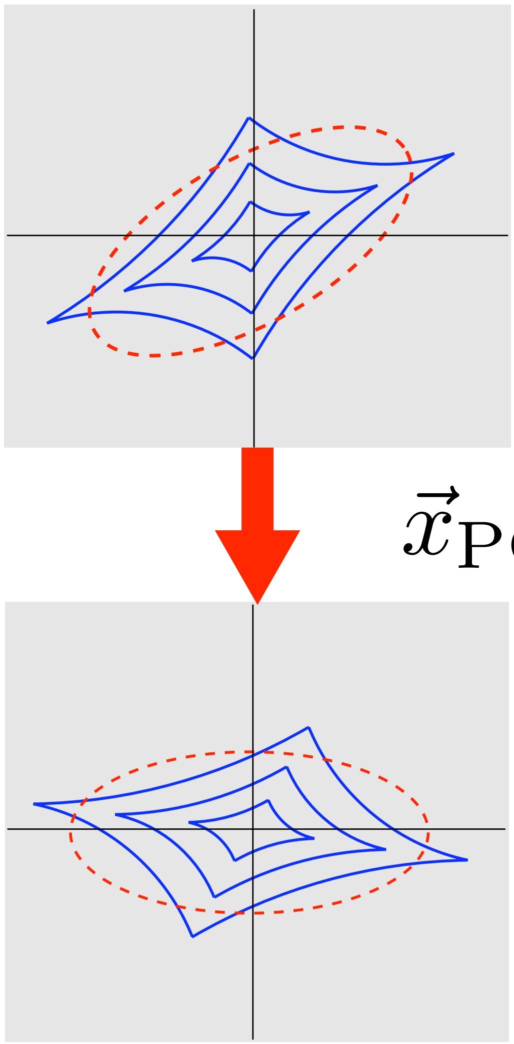


$\vec{x}$

PCA



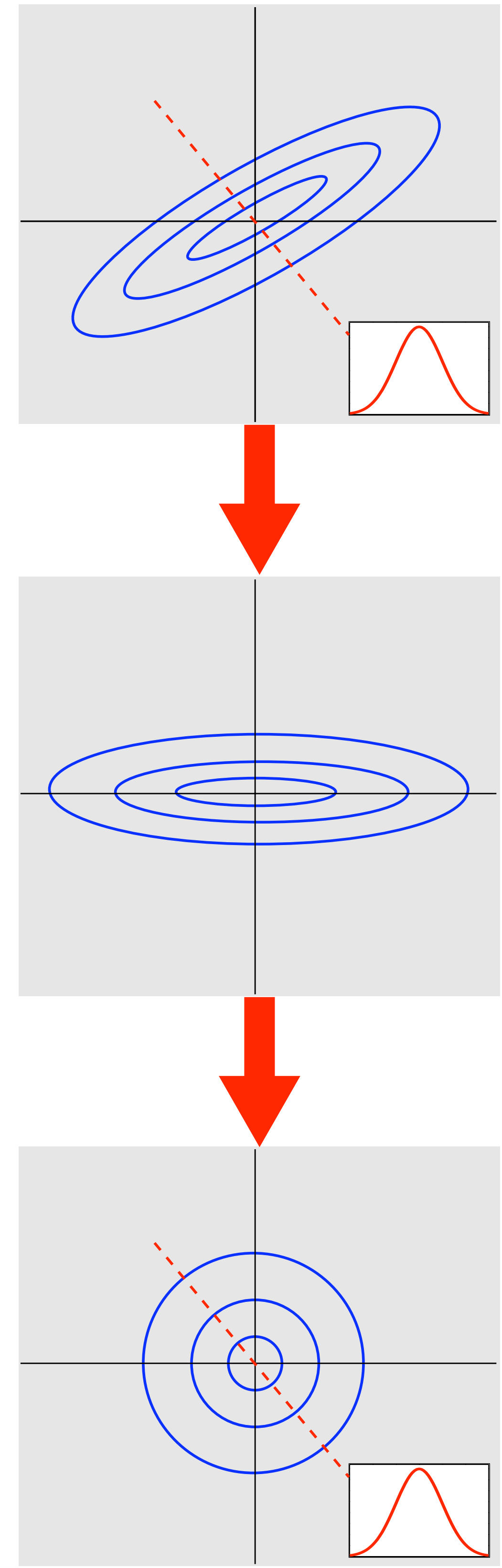
ICA



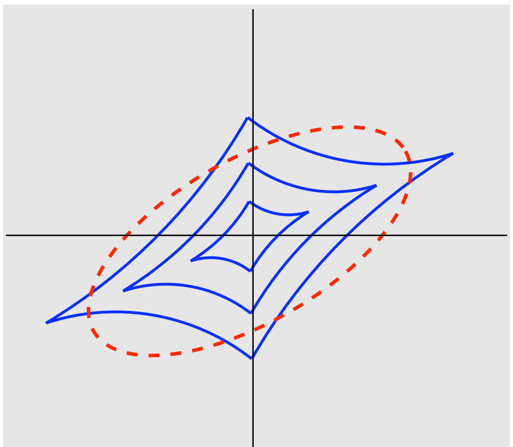
$$\vec{x}$$

$$\vec{x}_{\mathrm{PCA}} = U^T \vec{x}$$

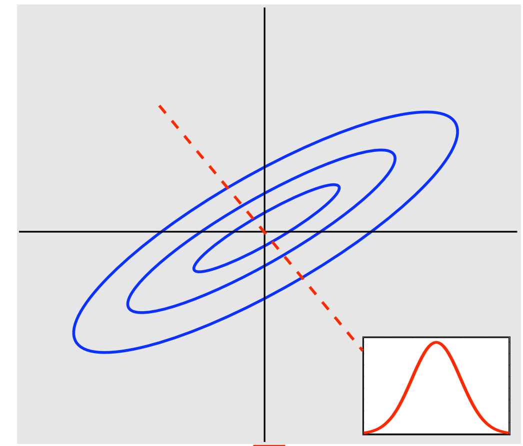
PCA



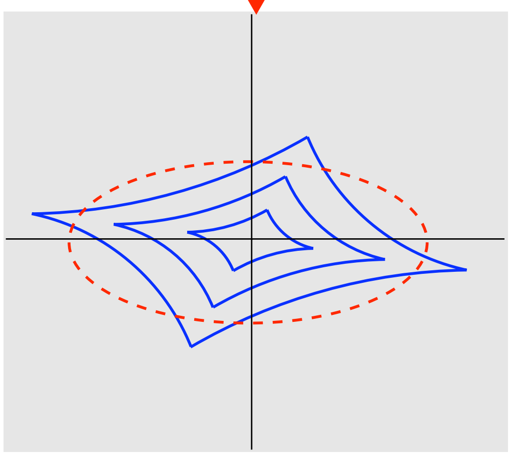
ICA PCA

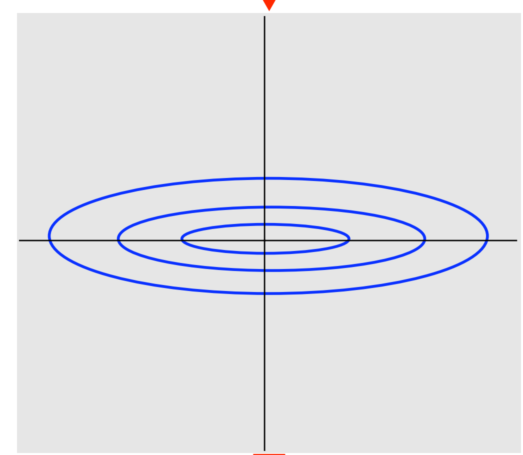


$\vec{x}$

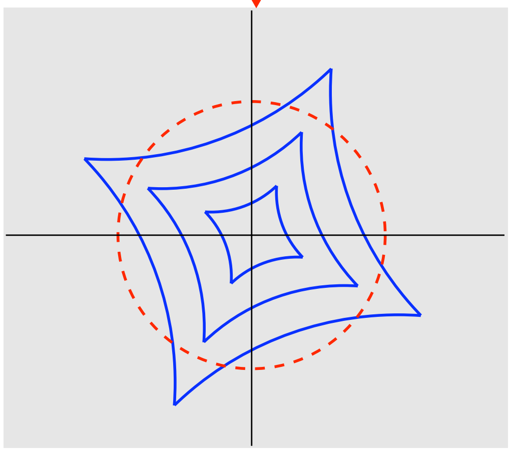


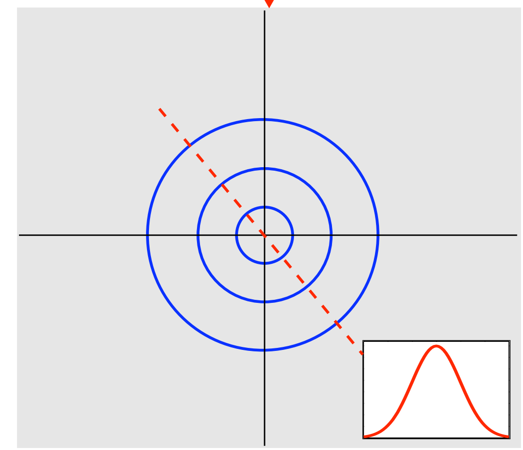
$$\vec{x}_{\mathrm{PCA}} = U^T \vec{x}$$

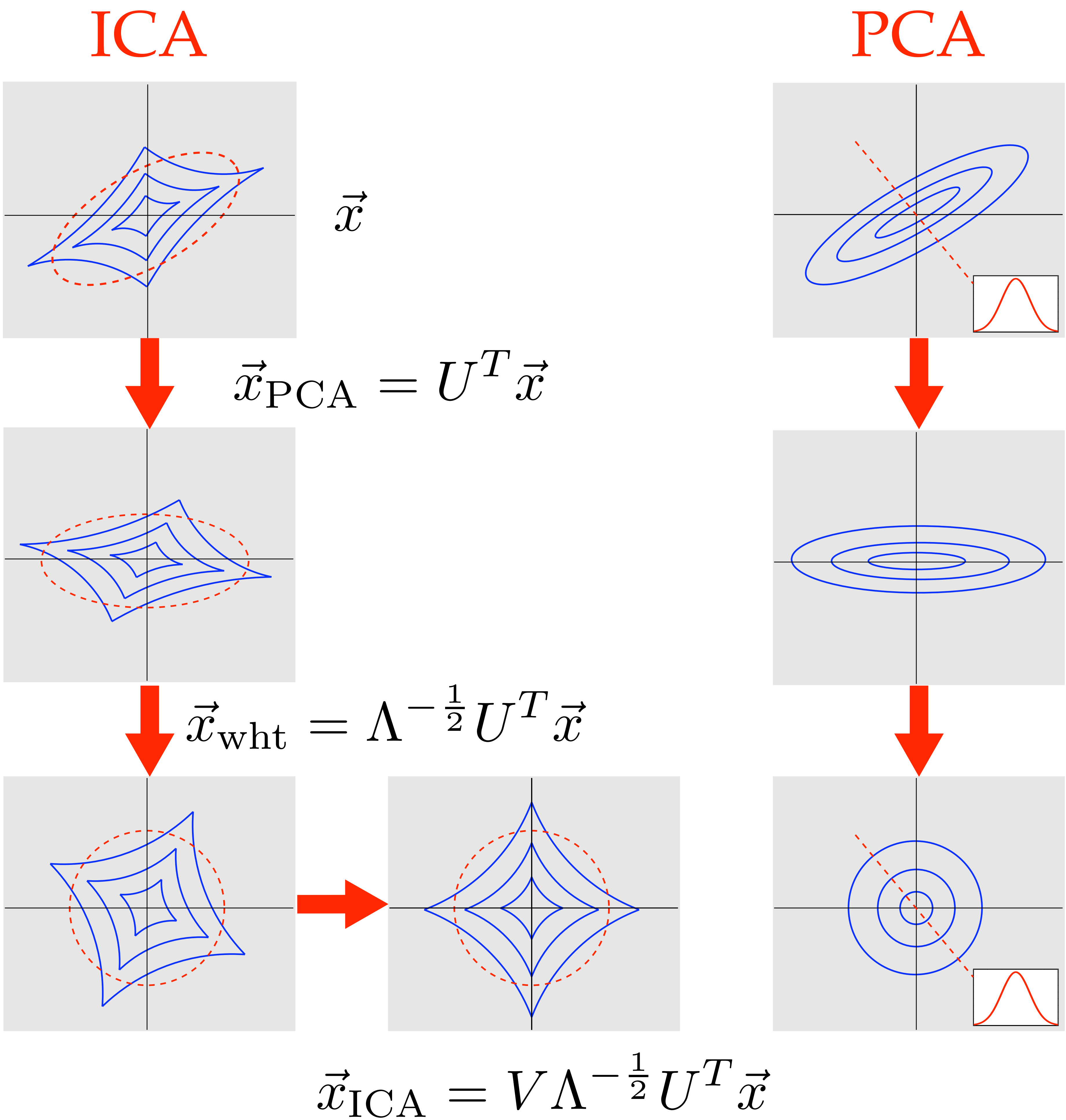




$$\vec{x}_{\mathrm{wht}} = \Lambda^{-\frac{1}{2}} U^T \vec{x}$$

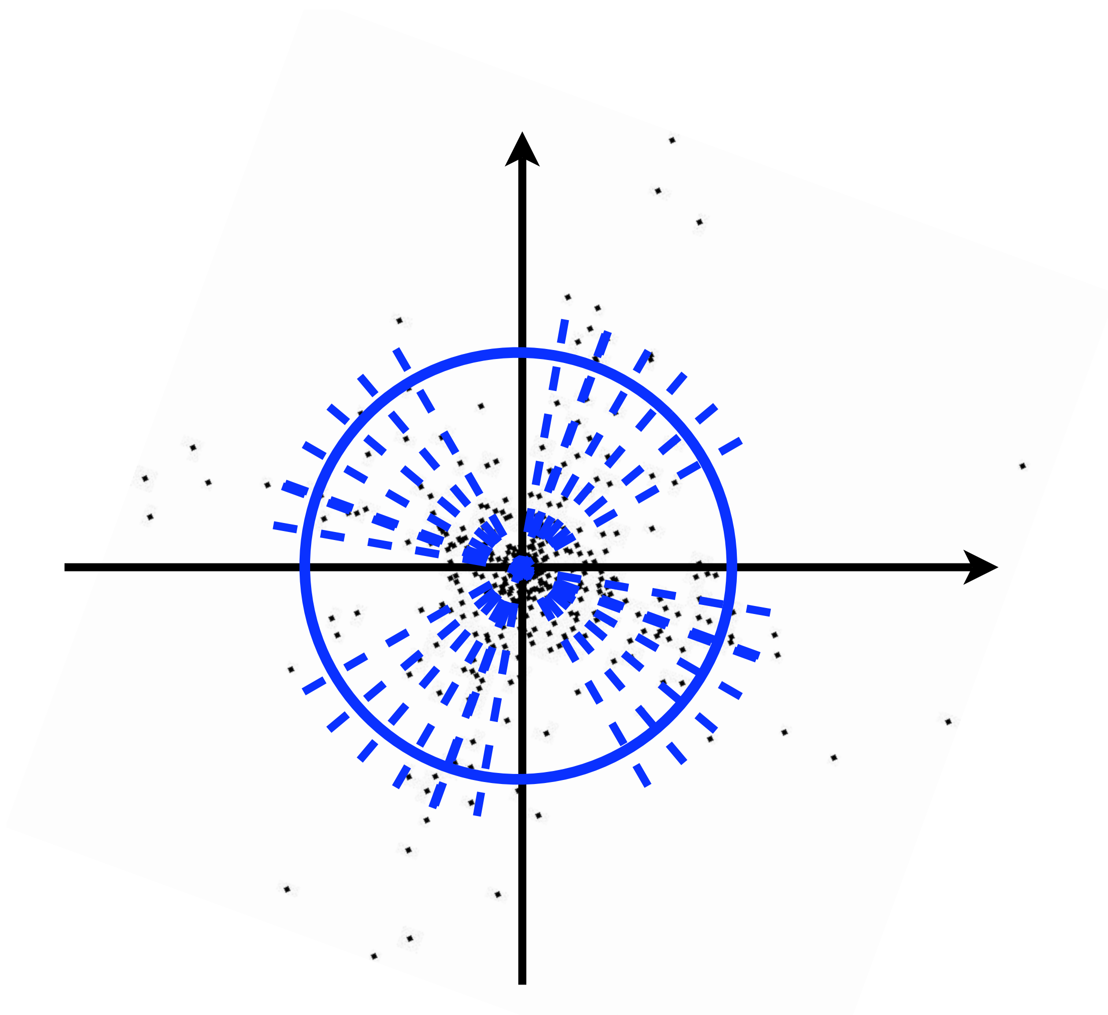





ICA
PCA
$\vec{x}$
$\vec{x}_{\mathrm{PCA}} = U^T \vec{x}$
$\vec{x}_{\mathrm{wht}} = \Lambda^{-\frac{1}{2}} U^T \vec{x}$
$\vec{x}_{\mathrm{ICA}} = V \Lambda^{-\frac{1}{2}} U^T \vec{x}$

finding V

- find final rotation that maximizes non-Gaussianity
  - linear mixing makes more Gaussian (CLT)
  - equivalent to maximize sparseness



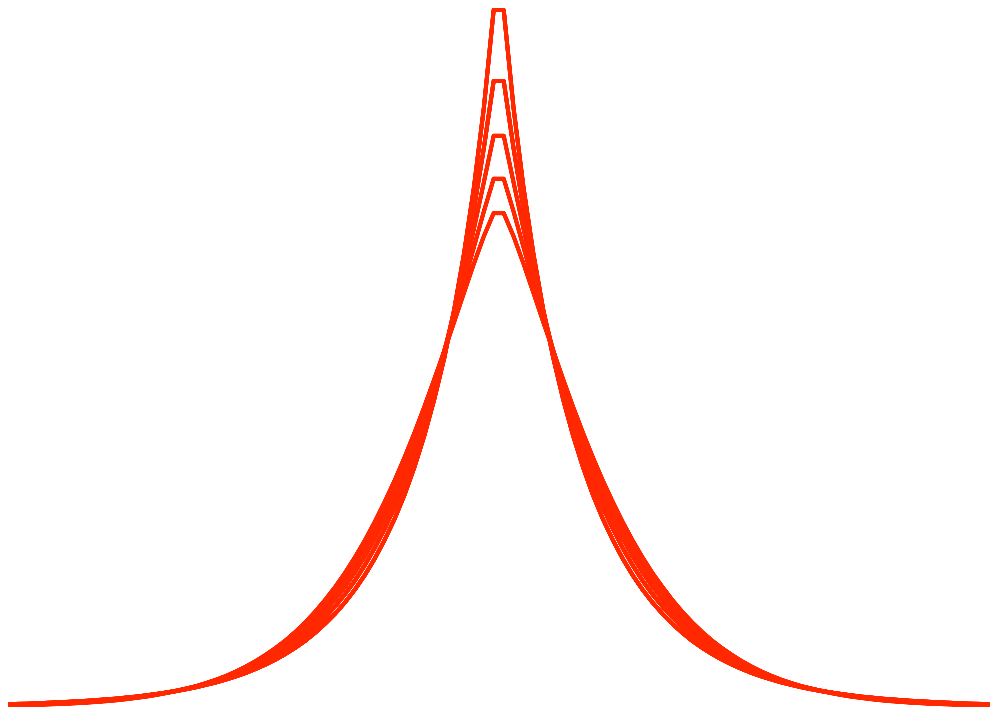
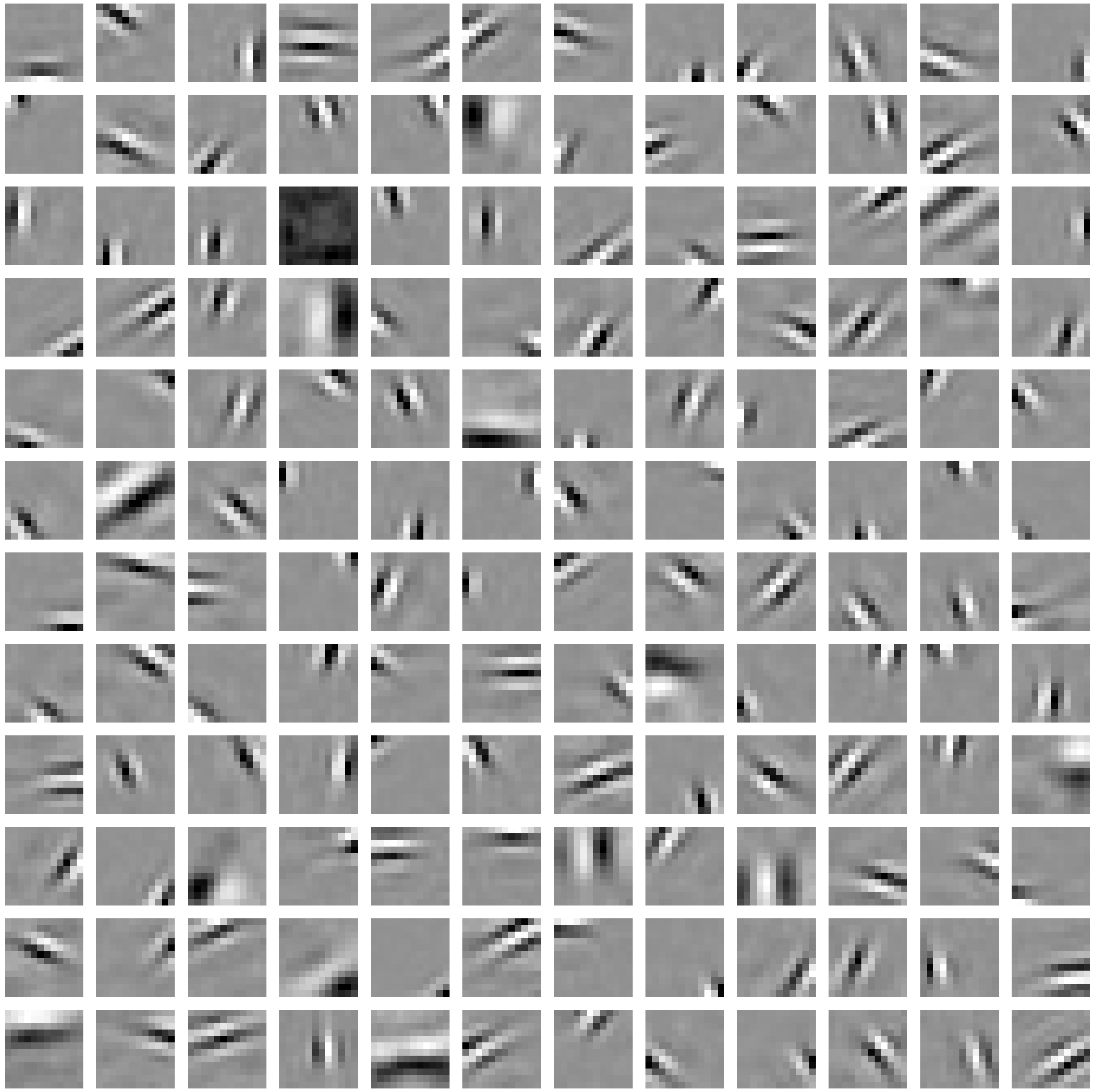


figure from [Bethge 2008]



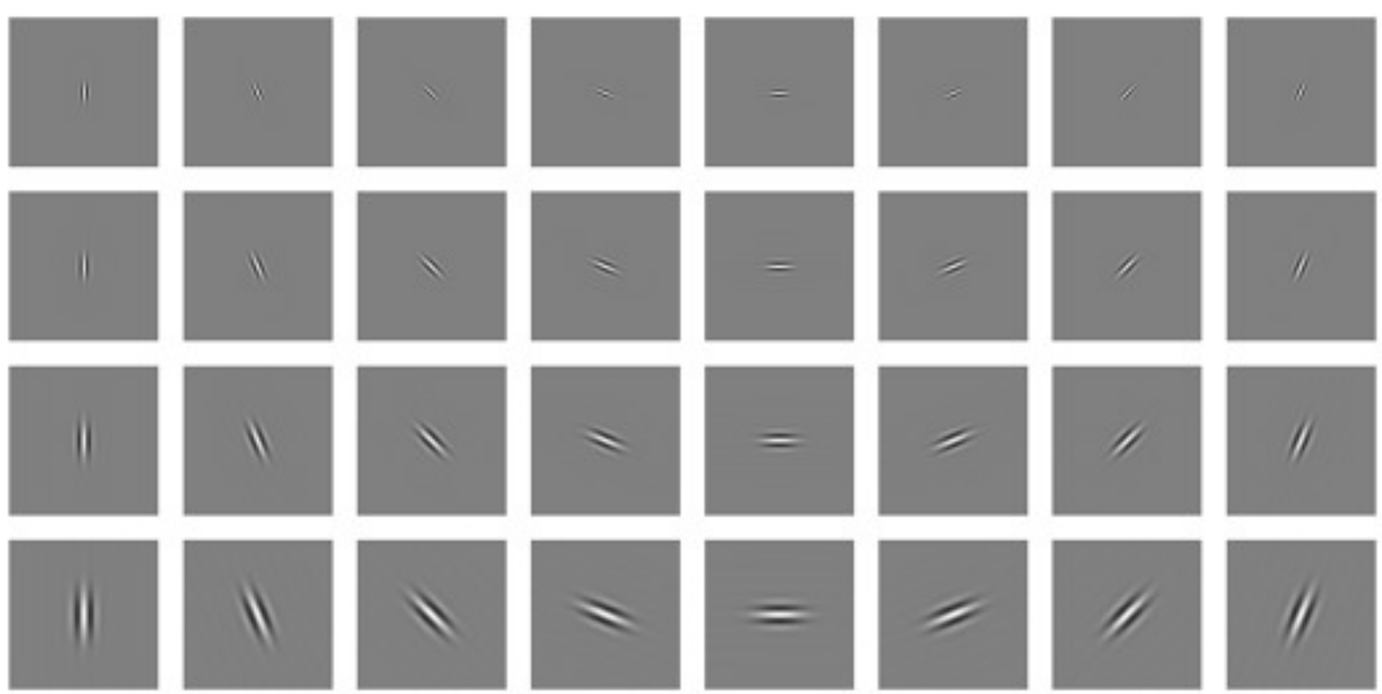
ICA bases (squared columns of A) learned from natural images
- similar shape to receptive field of V1 simple cells
  [Olshausen & Field 1996, Bell & Sejnowski 1997]

break

representation

- ICA basis resemble wavelet and other multi-scale oriented linear representations - localized in spatial location, frequency band and local orientation
- ICA basis are learned from data, while wavelet basis are fixed

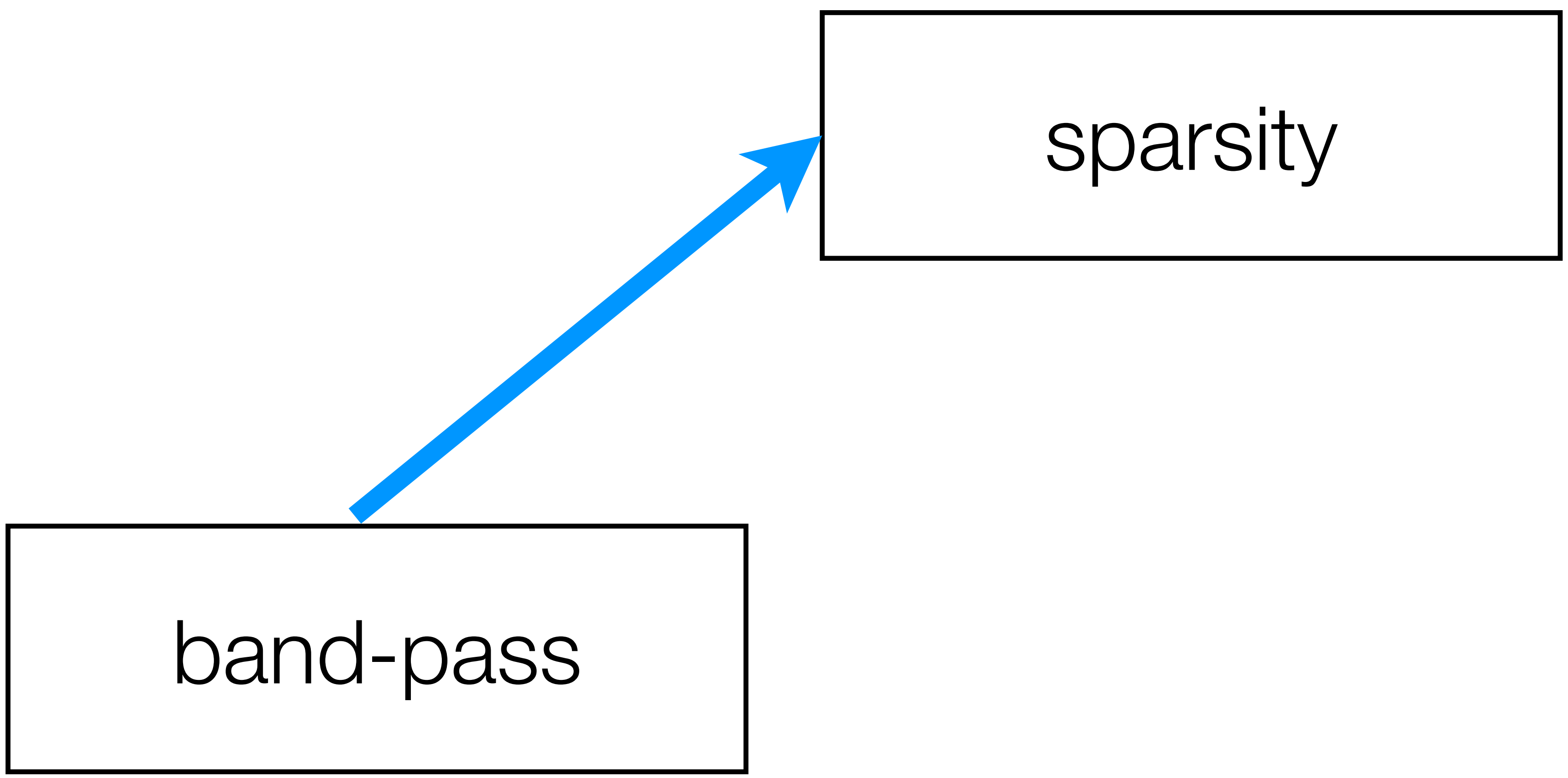
Gabor wavelet basis



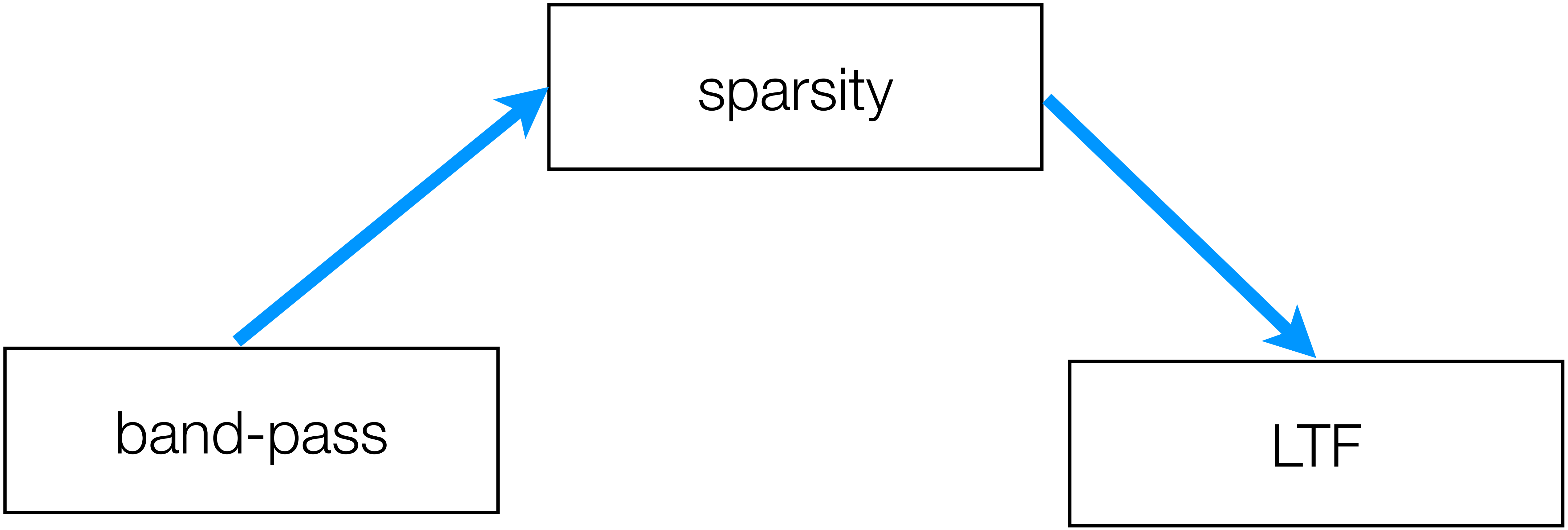
summary

band-pass

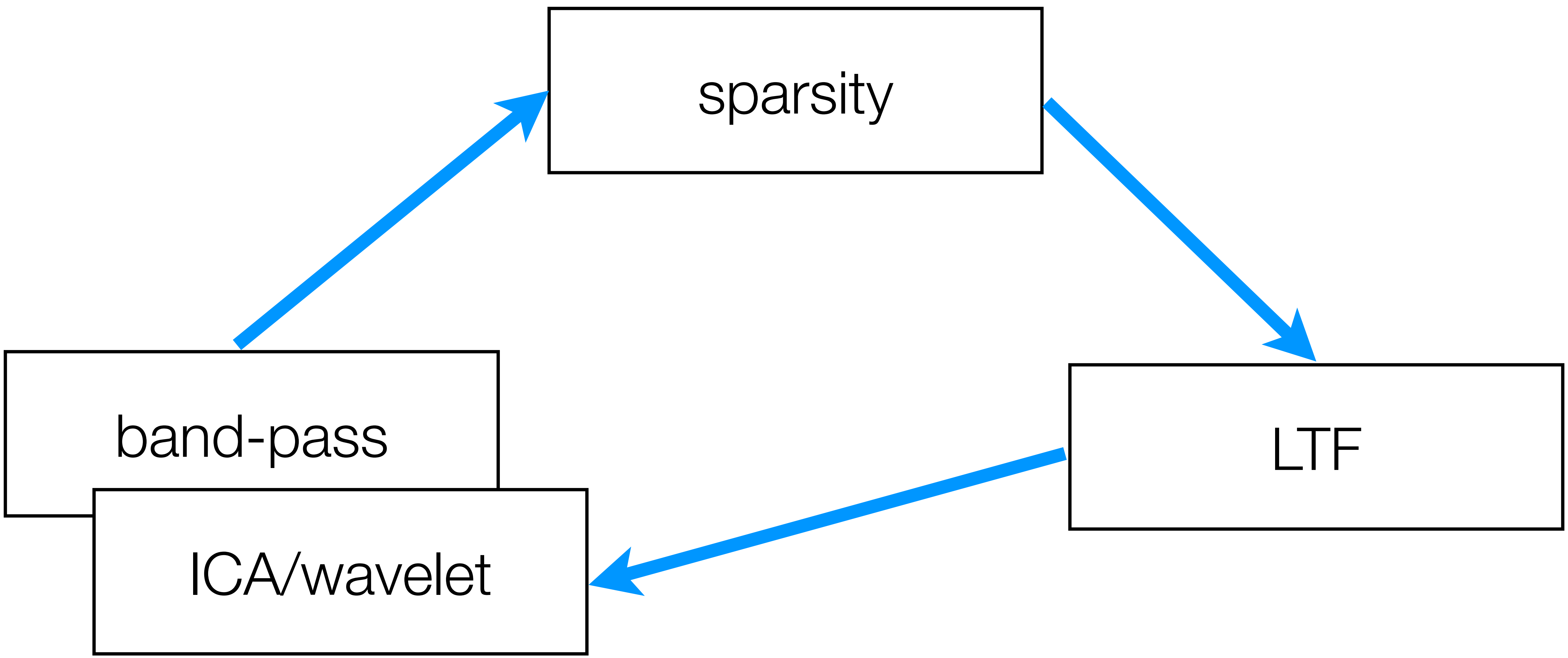
summary



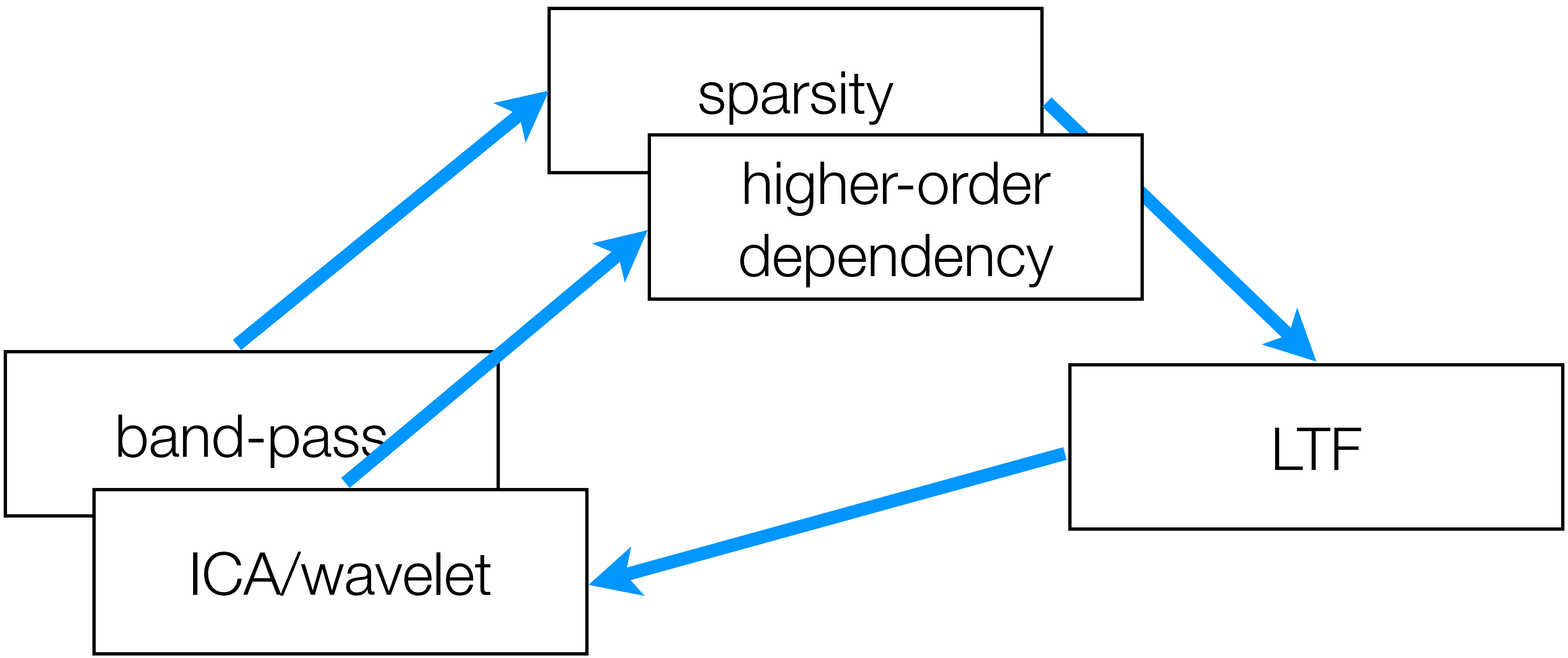
summary



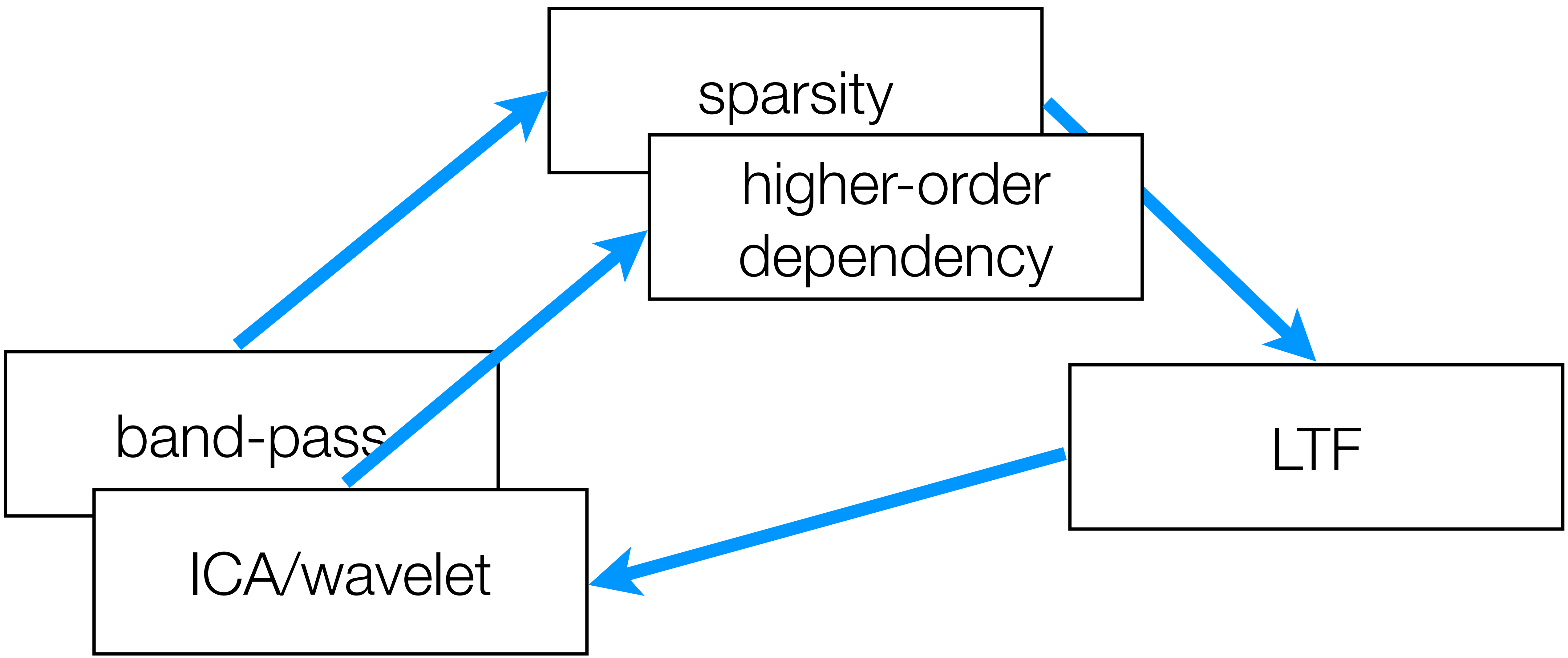
summary



summary


sparsity
higher-order
dependency
band-pass
ICA/wavelet
LTF

summary

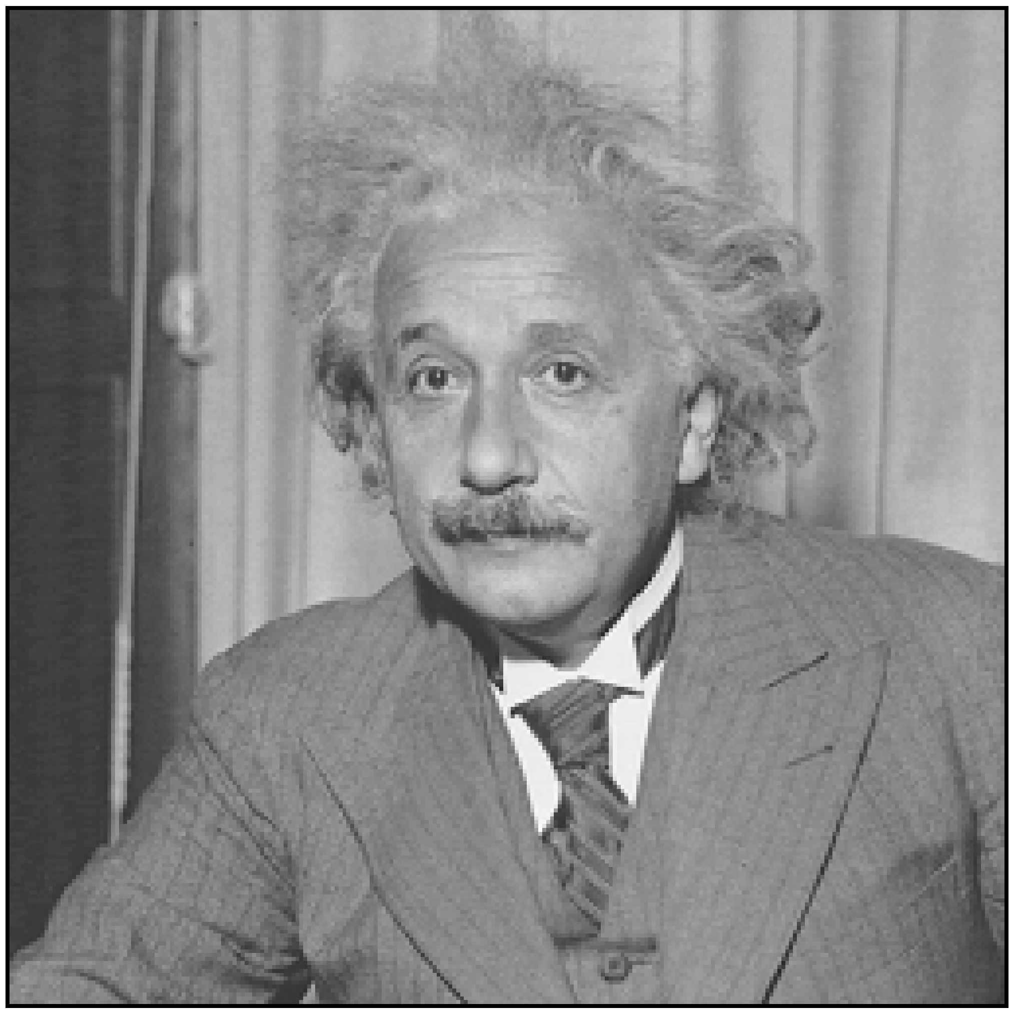


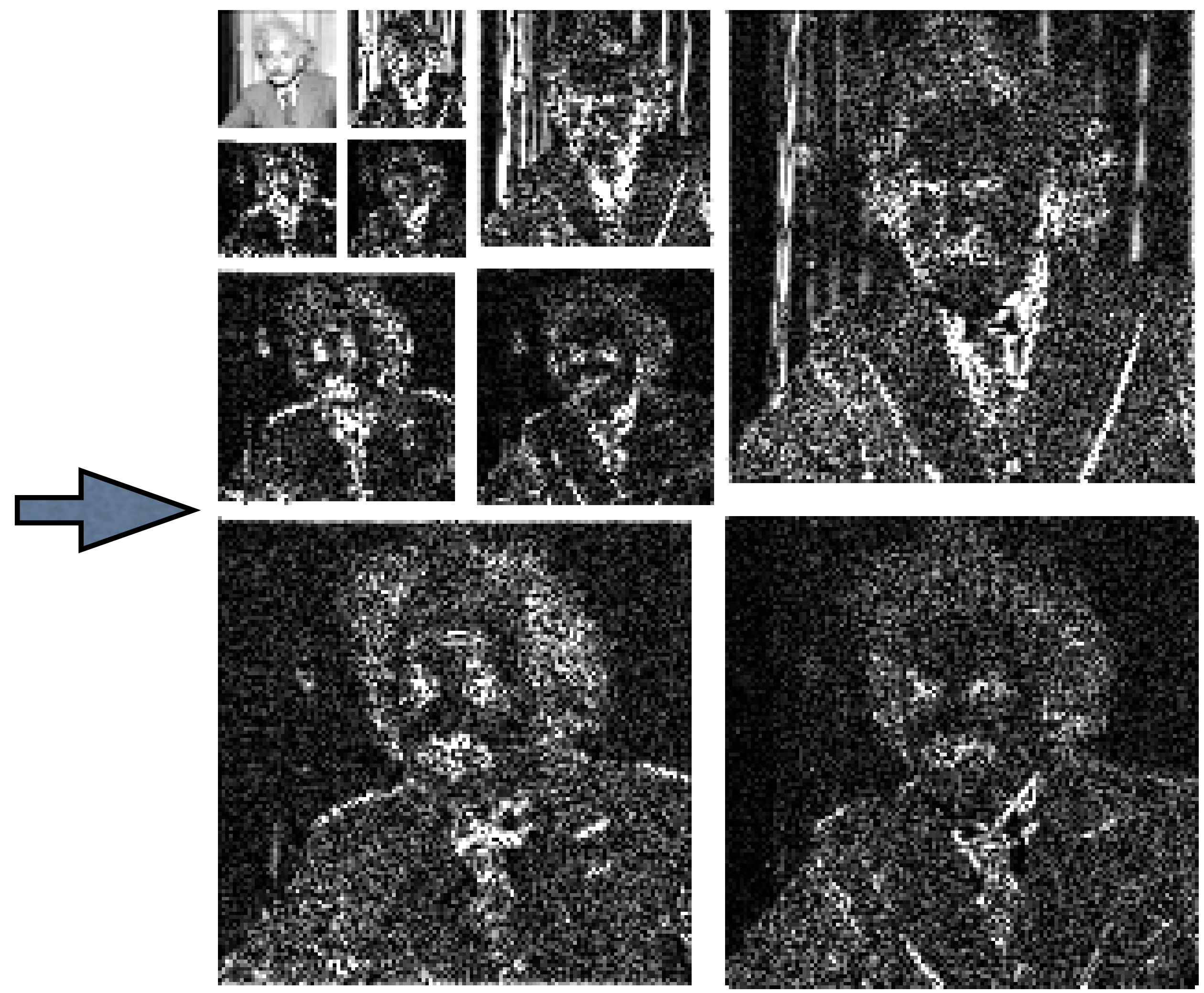
Not enough!

problems with LTF

- any band-pass or high-pass filter will lead to heavy tail marginals (even random ones)
- if natural images are truly linear mixture of independent non-Gaussian sources, random projection (filtering) should look like Gaussian
  - central limit theorem

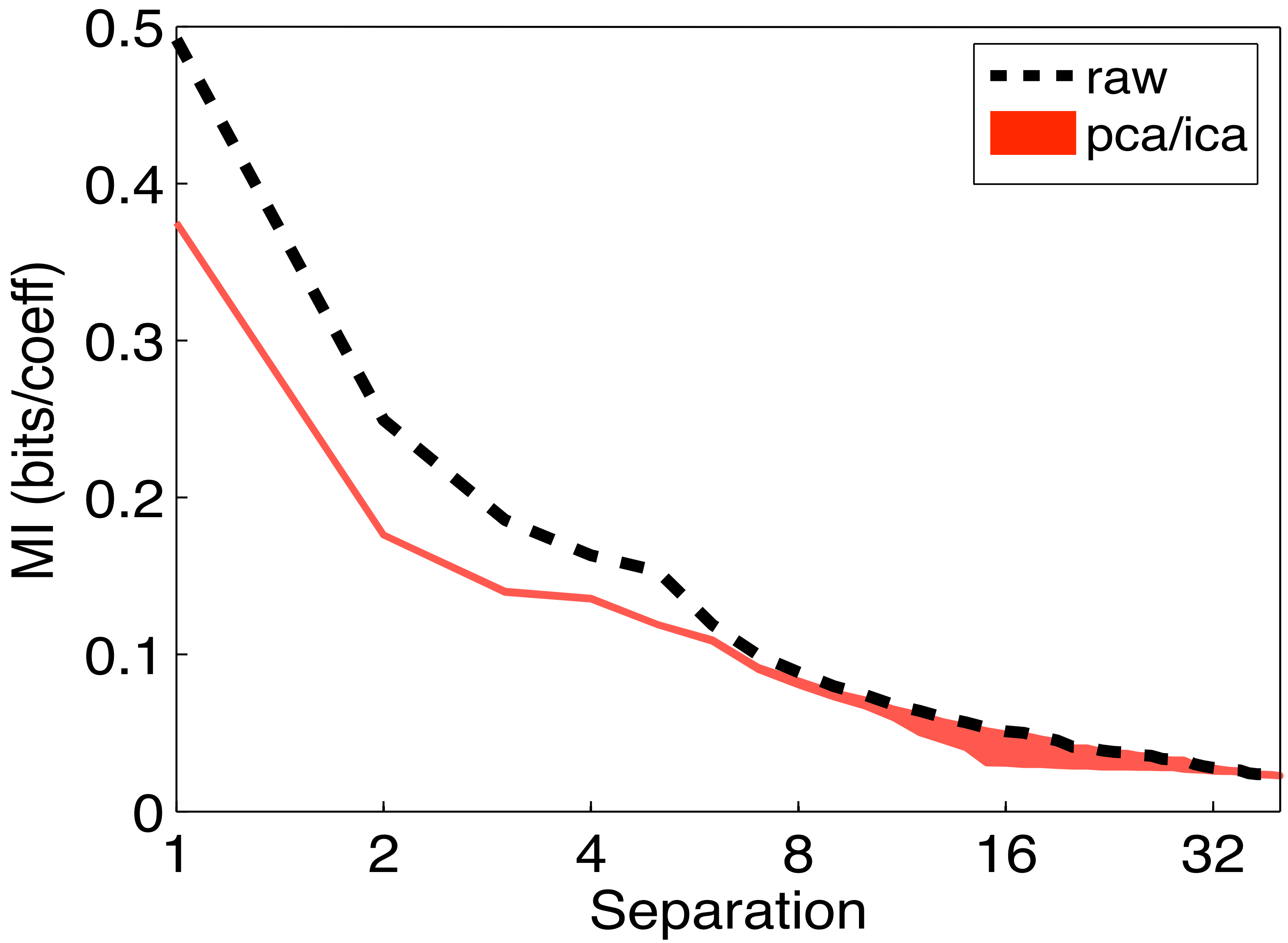
problems with LTF



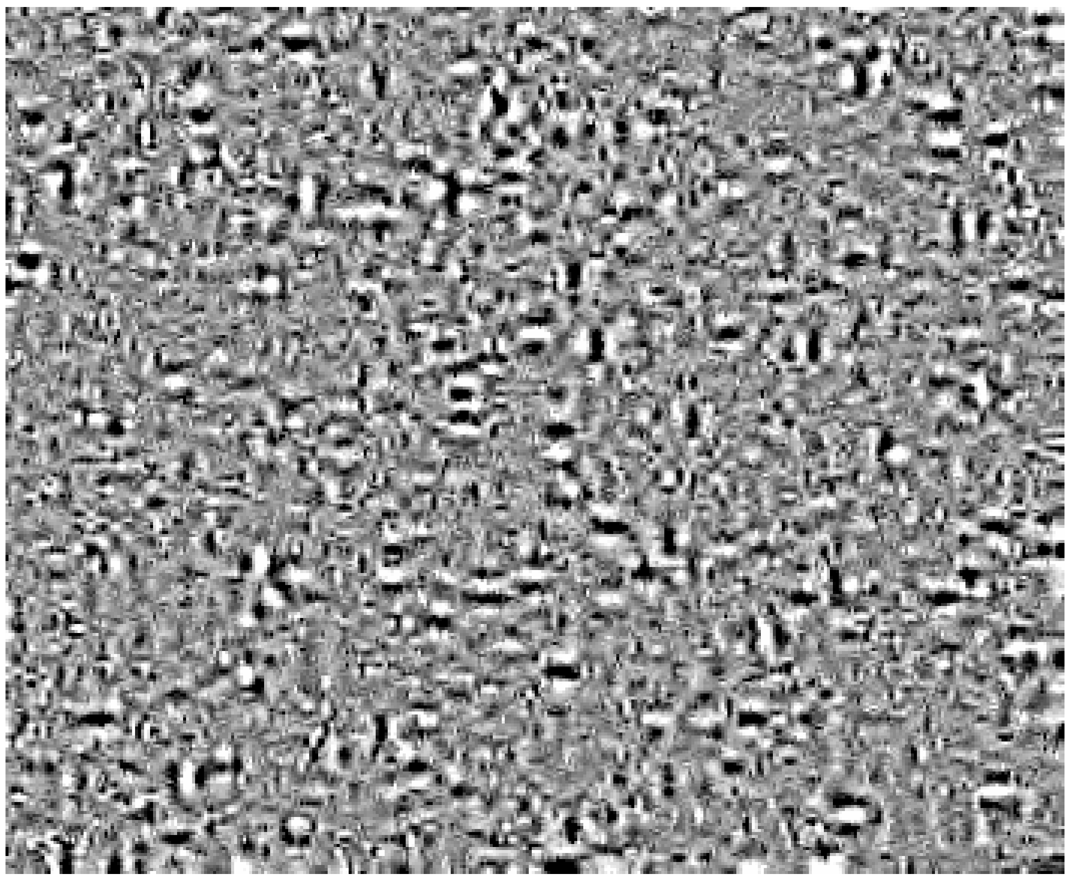


[Simoncelli '97; Buccigrossi &Simoncelli '99]

- Large-magnitude subband coefficients are found at neighboring positions, orientations, and scales.



ICA achieves very little improvement over PCA in terms of dependency reduction [Bethge 06, Lyu & Simoncelli 08] sample from LTF



natural images after ICA filtering

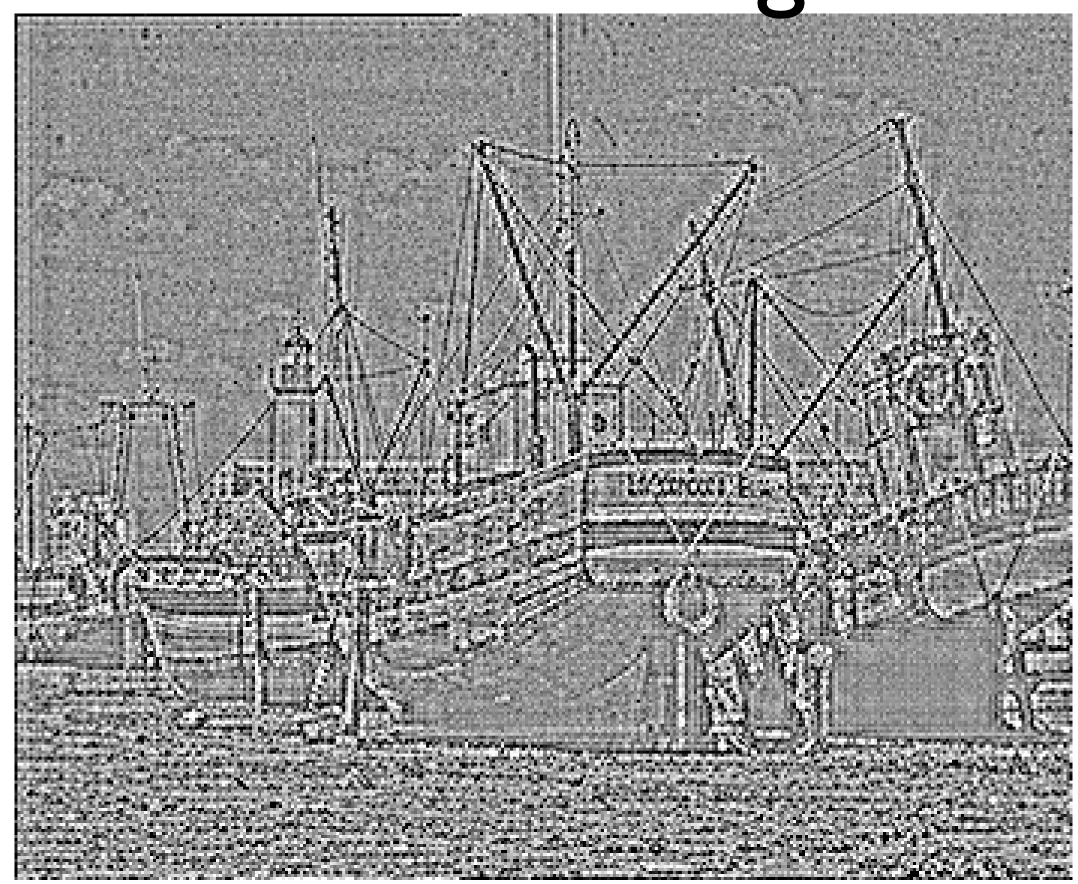


figure courtesy of Eero Simoncelli

remedy

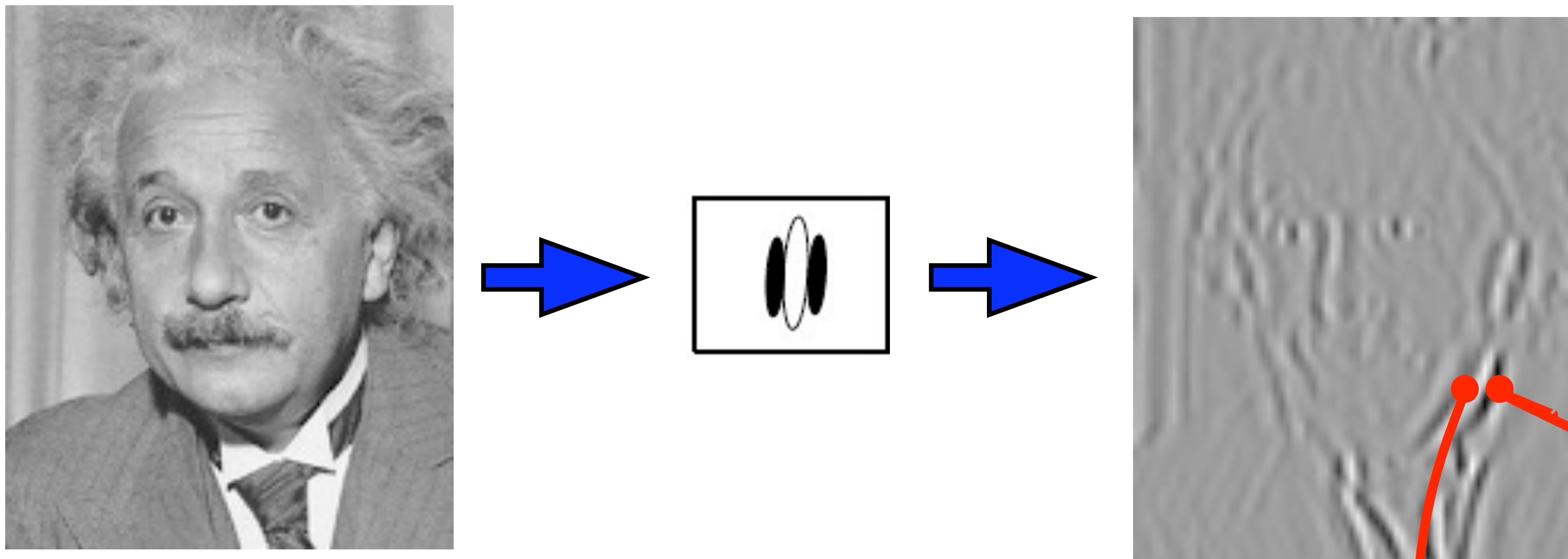
- assumptions in LTF model and ICA
  - factorial marginals for filter outputs
  - linear combination
  - invertible

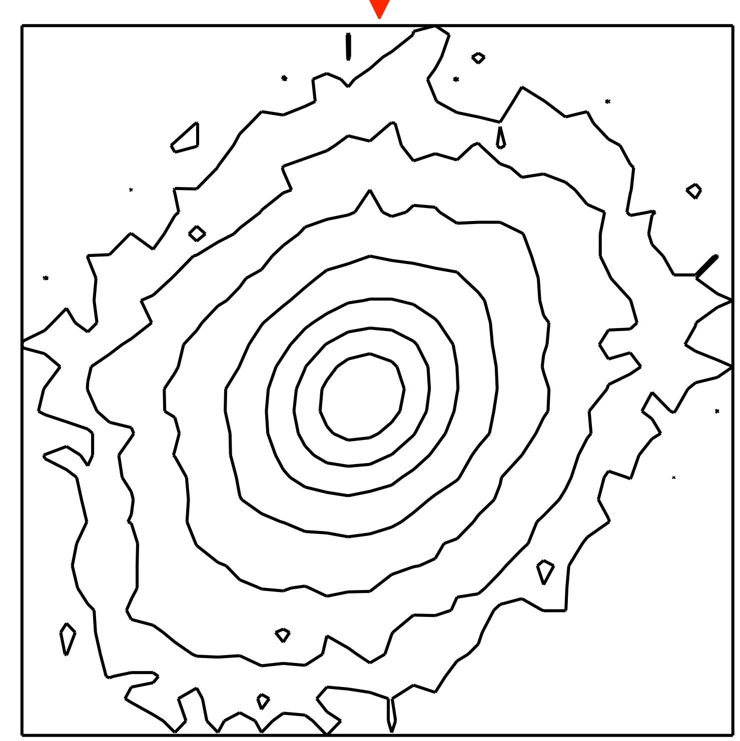
remedy

- assumptions in LTF model and ICA
  - factorial marginals for filter outputs
  - linear combination
  - ~~invertible~~
- model => [Zhu, Wu & Mumford 1997; Portilla & Simoncelli 2000]
  MaxEnt joint density with constraints on filter output
- representation => sparse coding [Olshausen & Field 1996]
  - find filters giving optimum sparsity
  - compressed sensing [Candes & Donoho 2003]

remedy

- assumptions in LTF model and ICA
  - factorial marginals for filter outputs
  - ~~linear combination~~ nonlinear
  - invertible

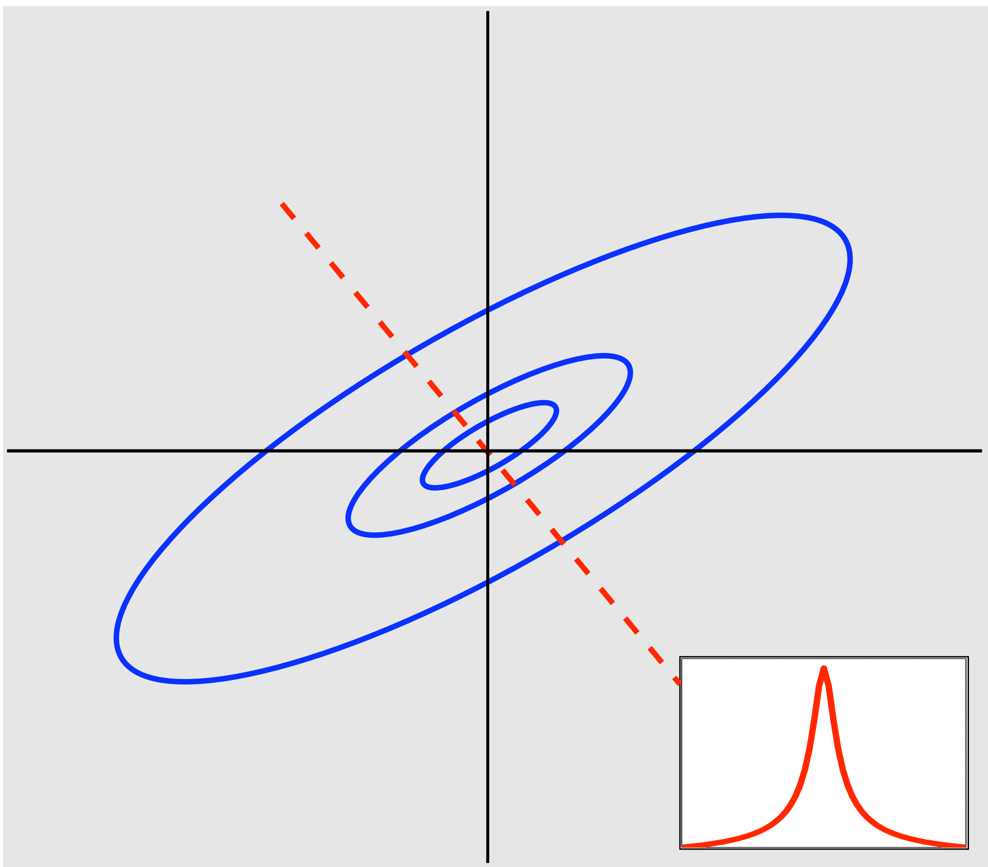




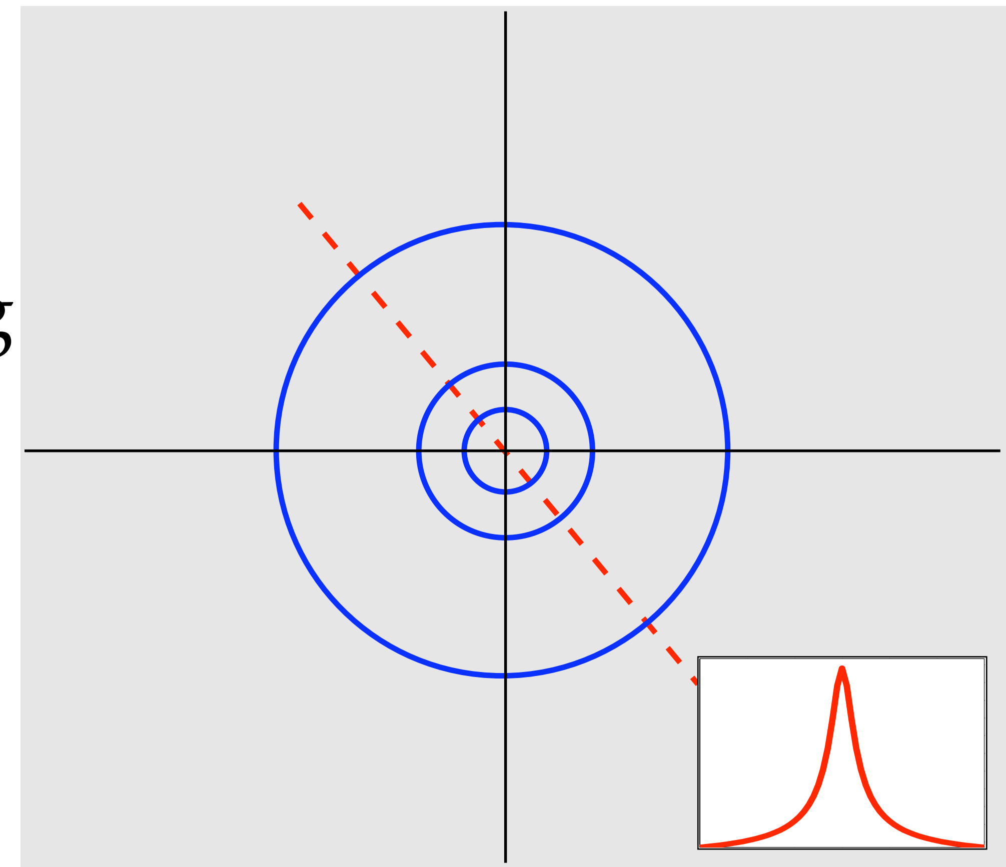
joint density of natural image
band-pass filter responses
with separation of 2 pixels

elliptically symmetric density

spherically symmetric density



whitening

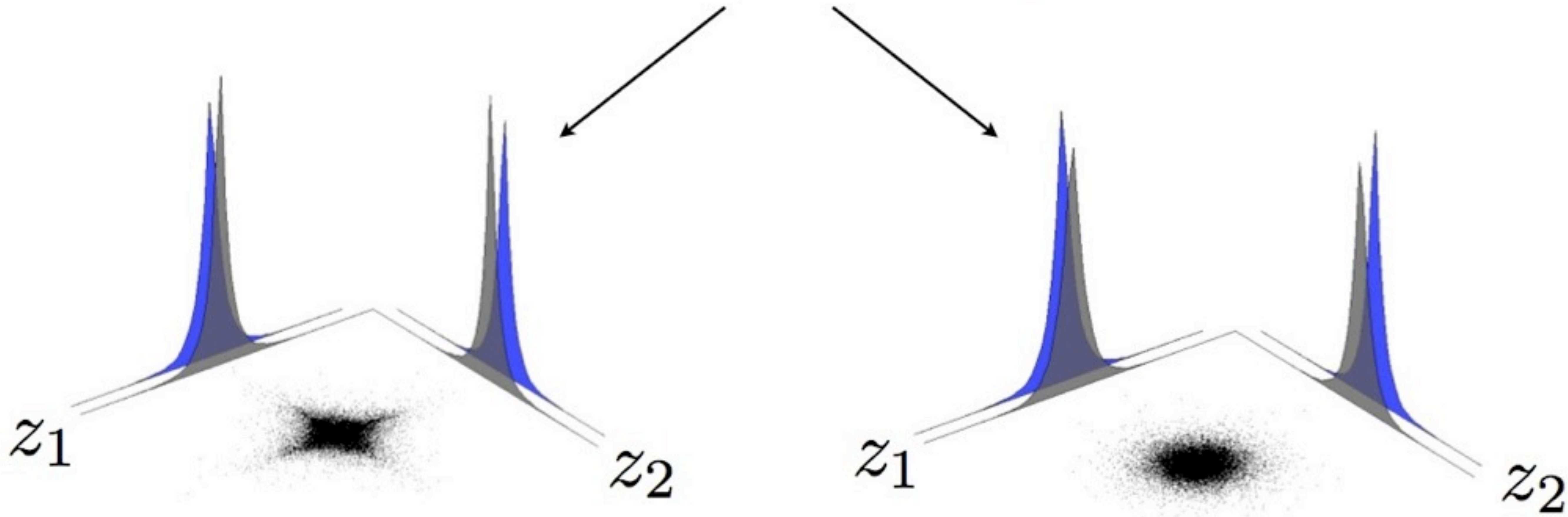


$$p_{\mathrm{esd}}(\vec{x}) = \frac{1}{\alpha|\Sigma|^{\frac{1}{2}}} f\left(-\frac{1}{2}\vec{x}^T\Sigma^{-1}\vec{x}\right)$$

$$p_{\mathrm{ssd}}(\vec{x}) = \frac{1}{\alpha} f\left(-\frac{1}{2}\vec{x}^T\vec{x}\right)$$

(Fang et.al. 1990)

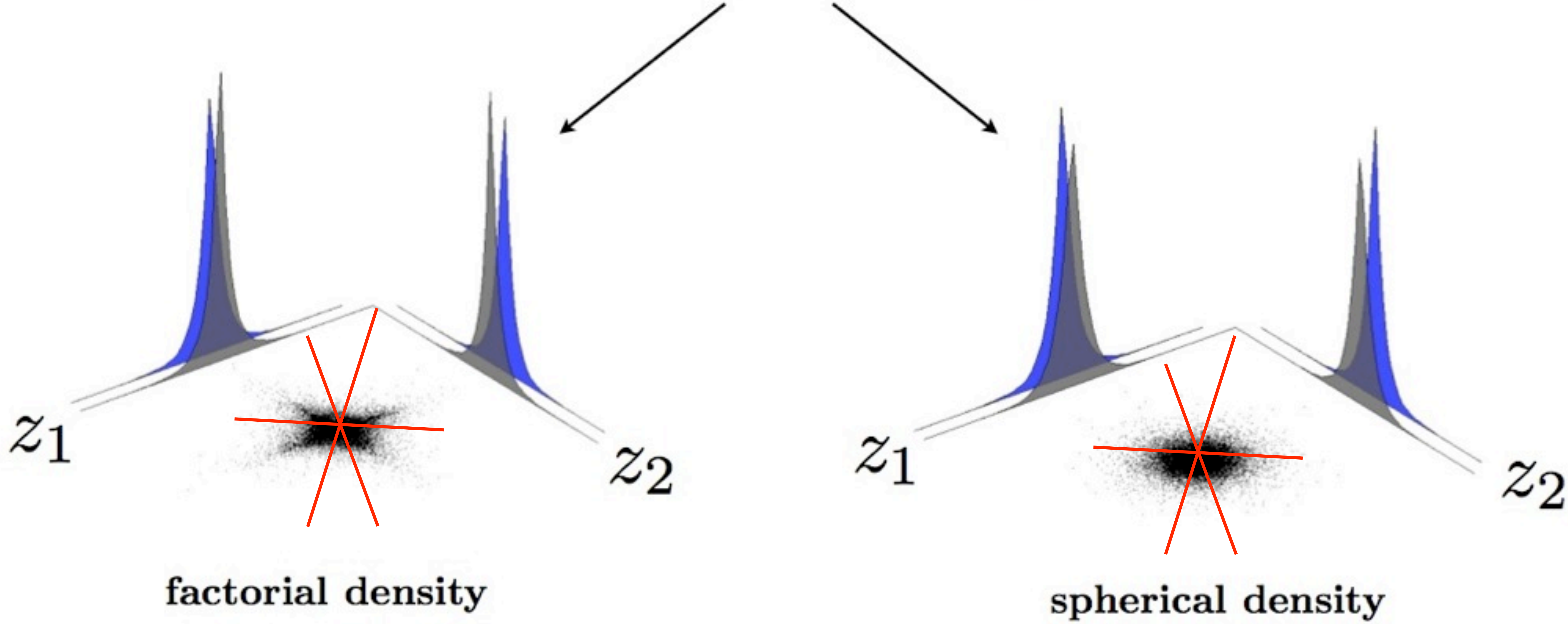
identical non-Gaussian marginals

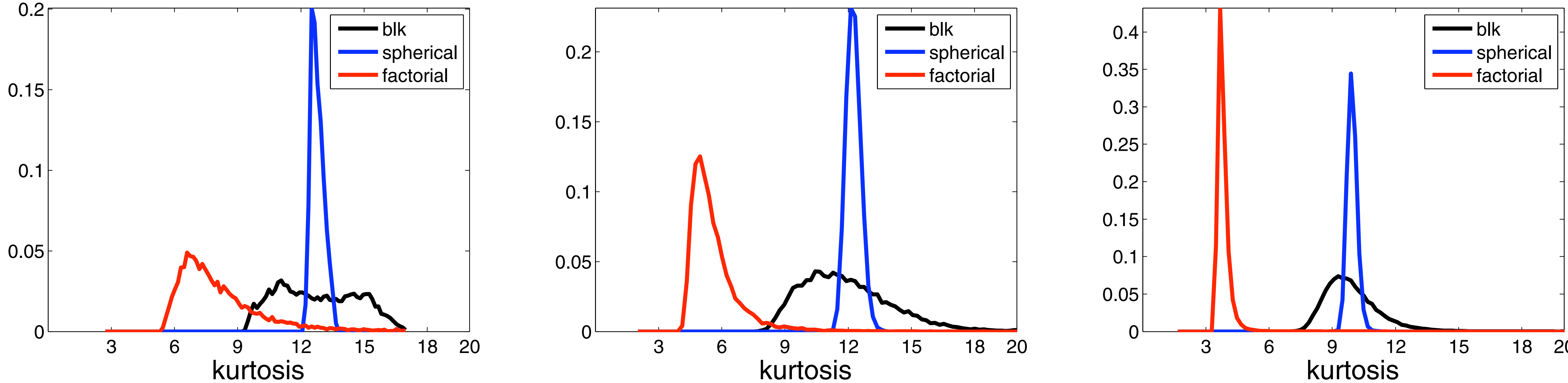


factorial density

spherical density

identical non-Gaussian marginals

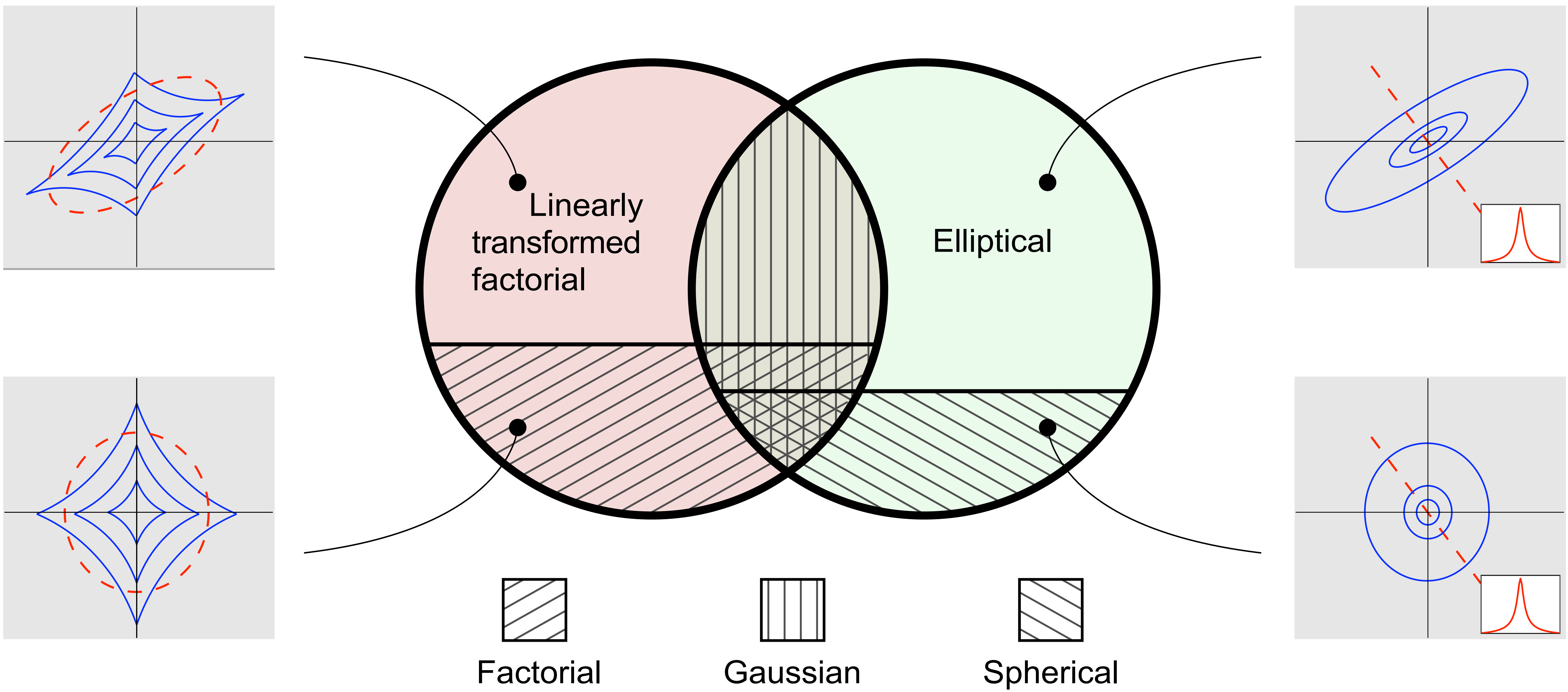


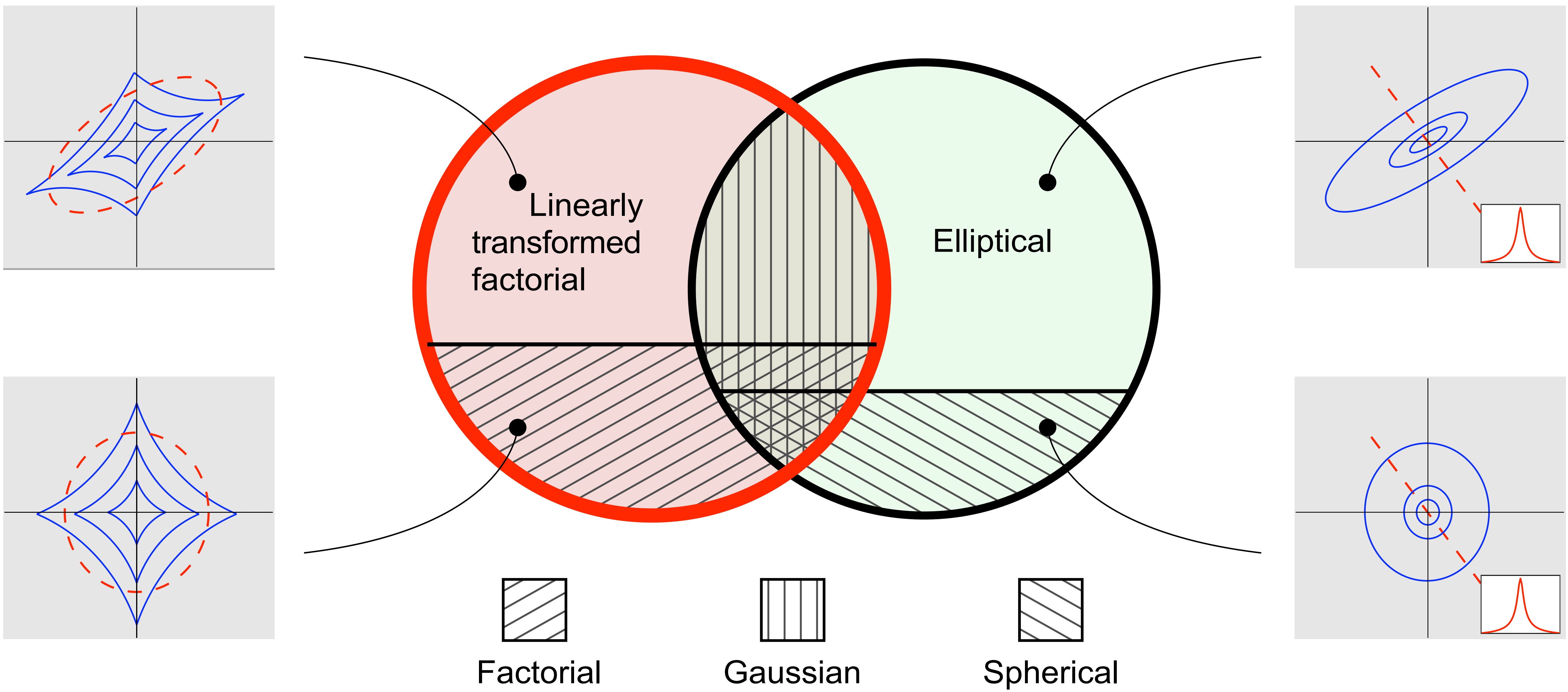


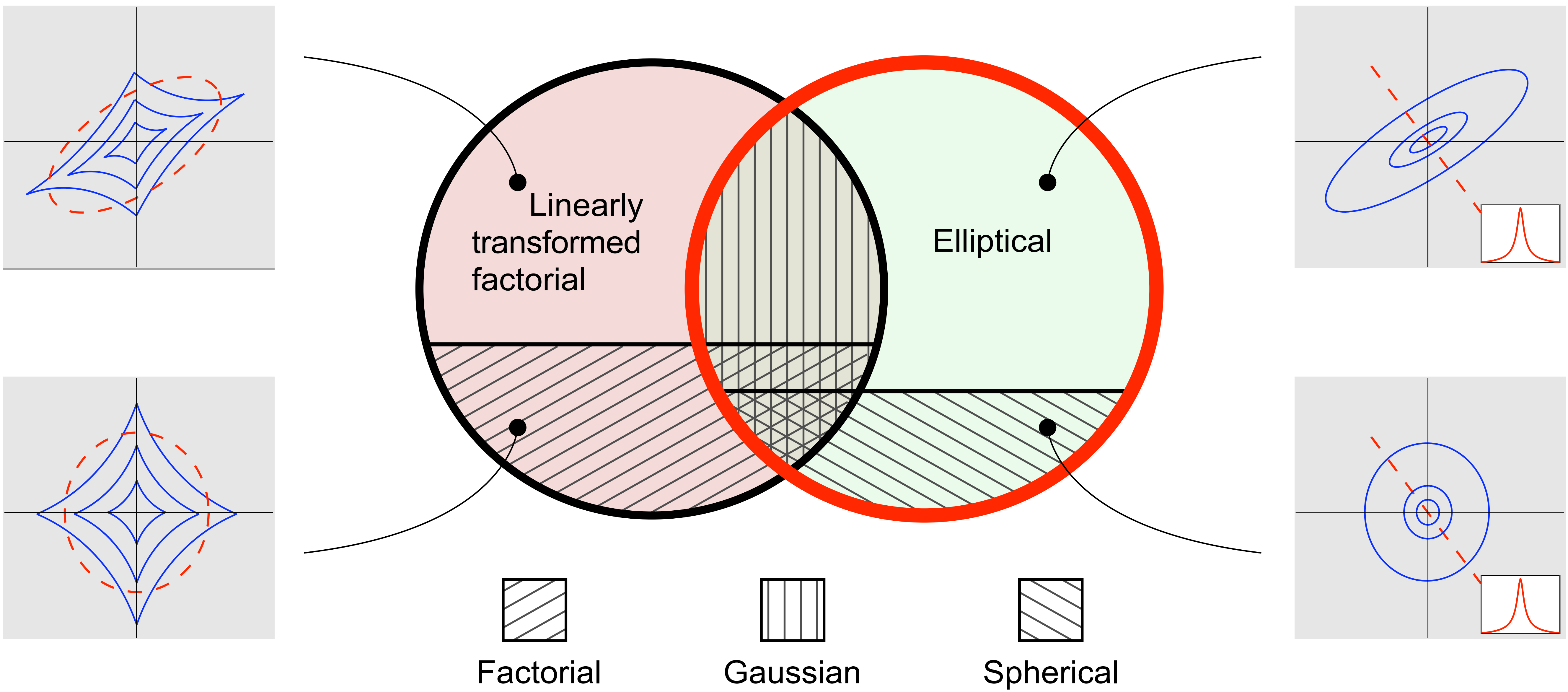
3x3 7x7 15x15

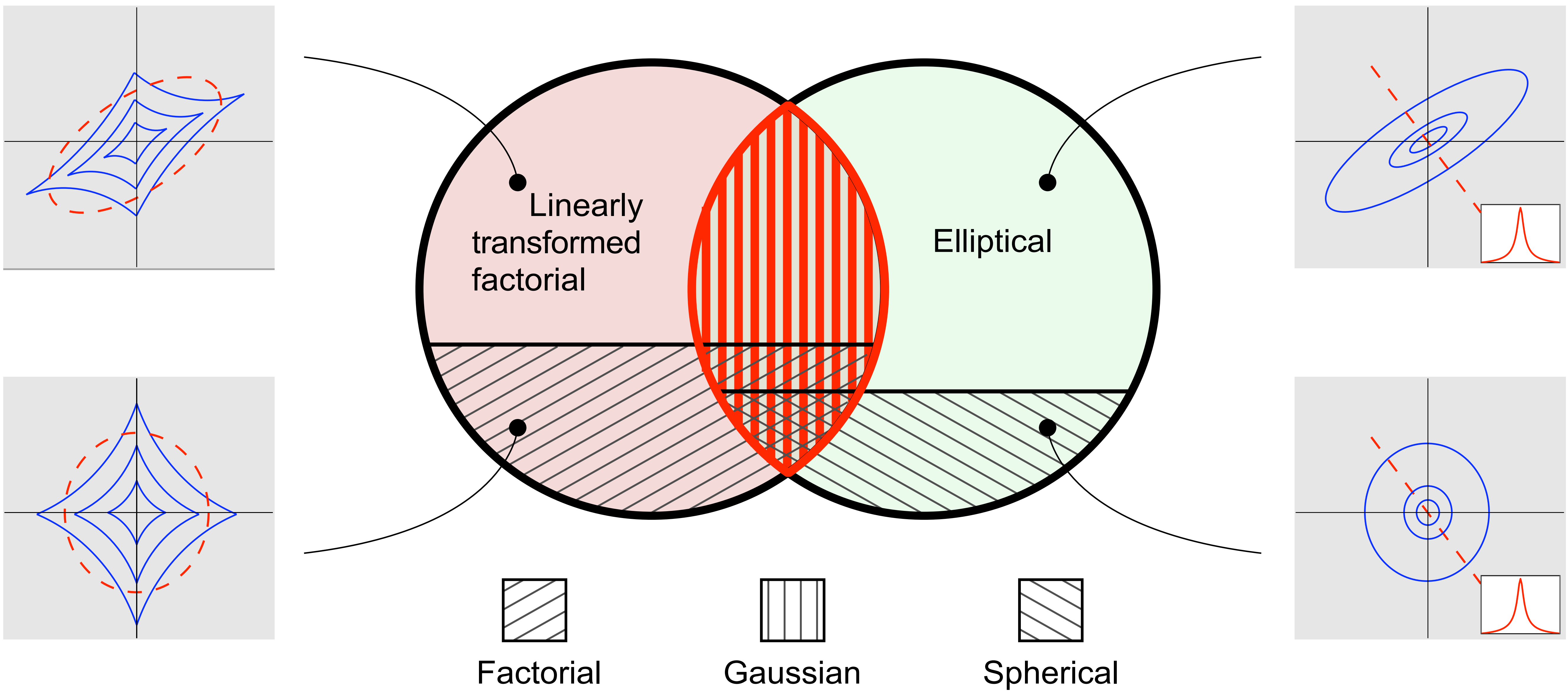
data (ICA'd): sphericalized: factorialized:

- Histograms, kurtosis of projections of image blocks onto random unit-norm basis functions.
- These imply data are closer to spherical than factorial


Linearly
transformed
factorial
Elliptical
Factorial
Gaussian
Spherical

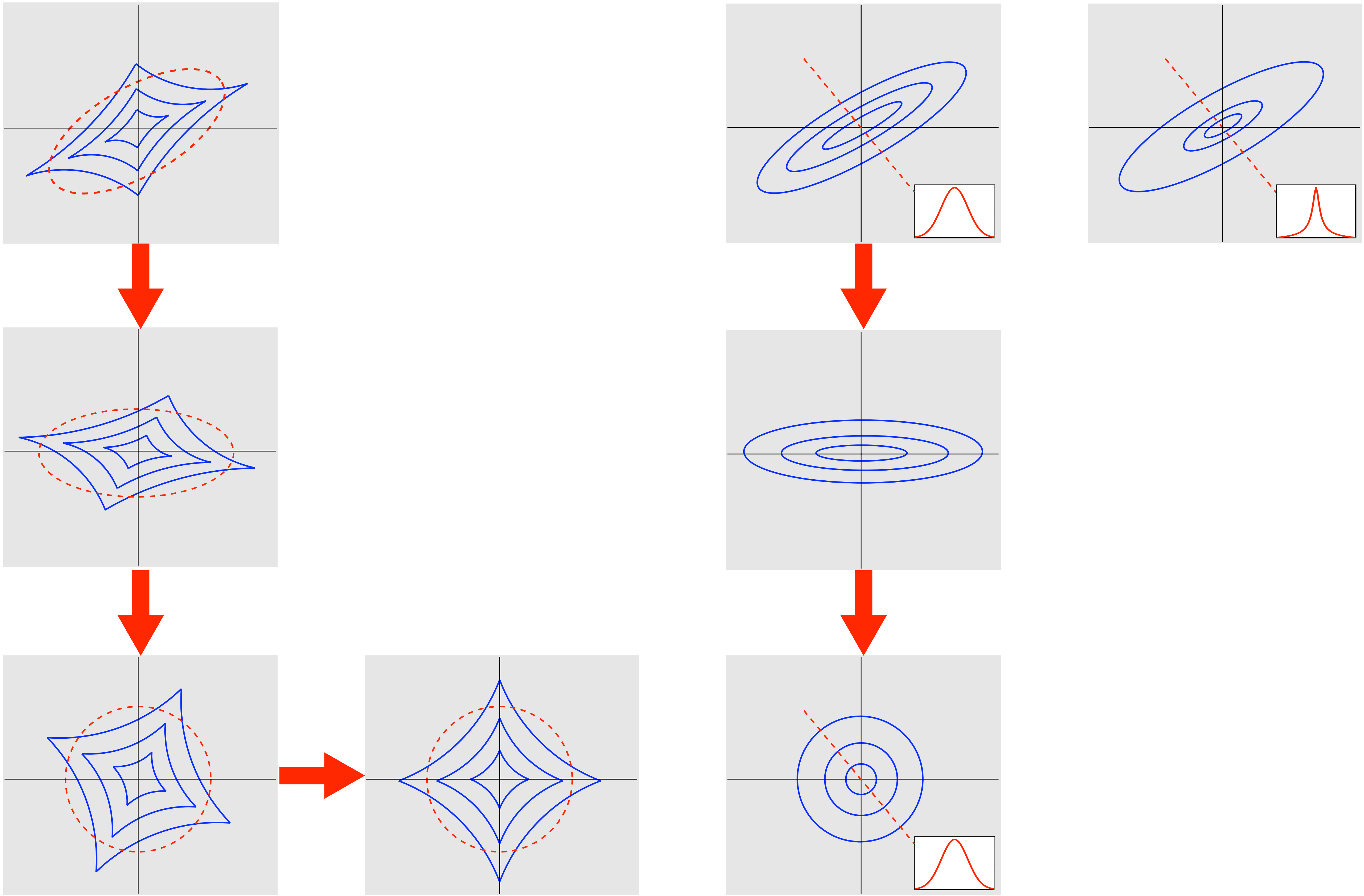

Linearly transformed factorial
Elliptical
Factorial
Gaussian
Spherical

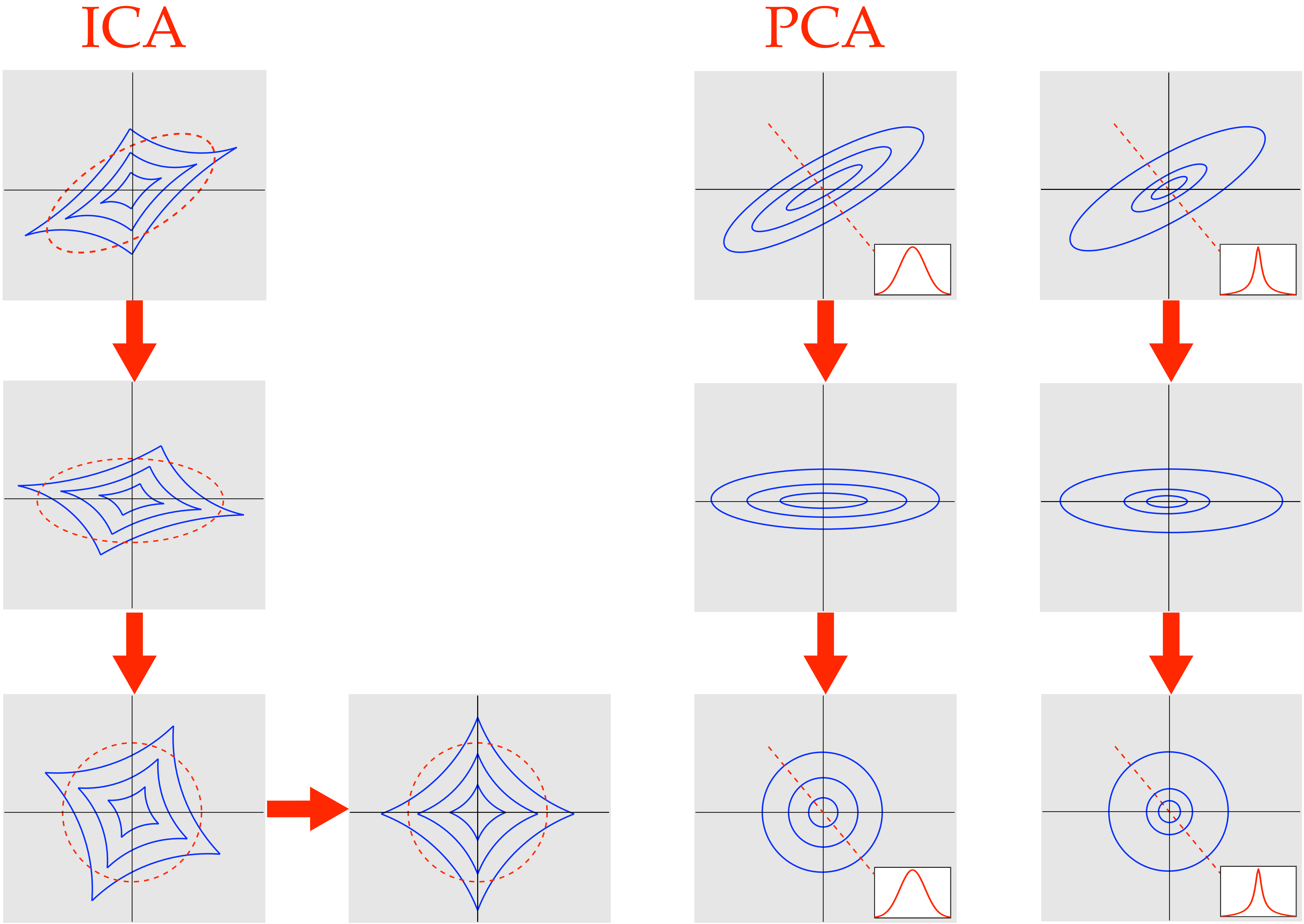

Linearly transformed factorial
Elliptical
Factorial
Gaussian
Spherical

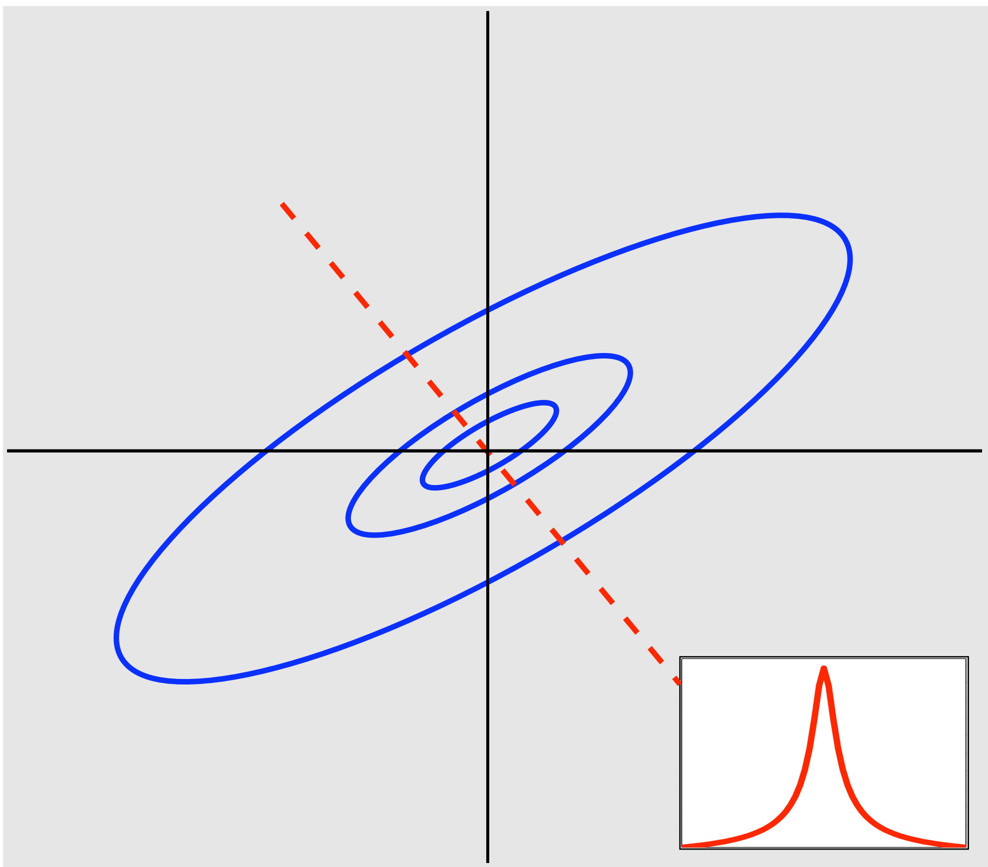

Linearly transformed factorial
Elliptical
Factorial
Gaussian
Spherical

ICA

PCA




ICA
PCA

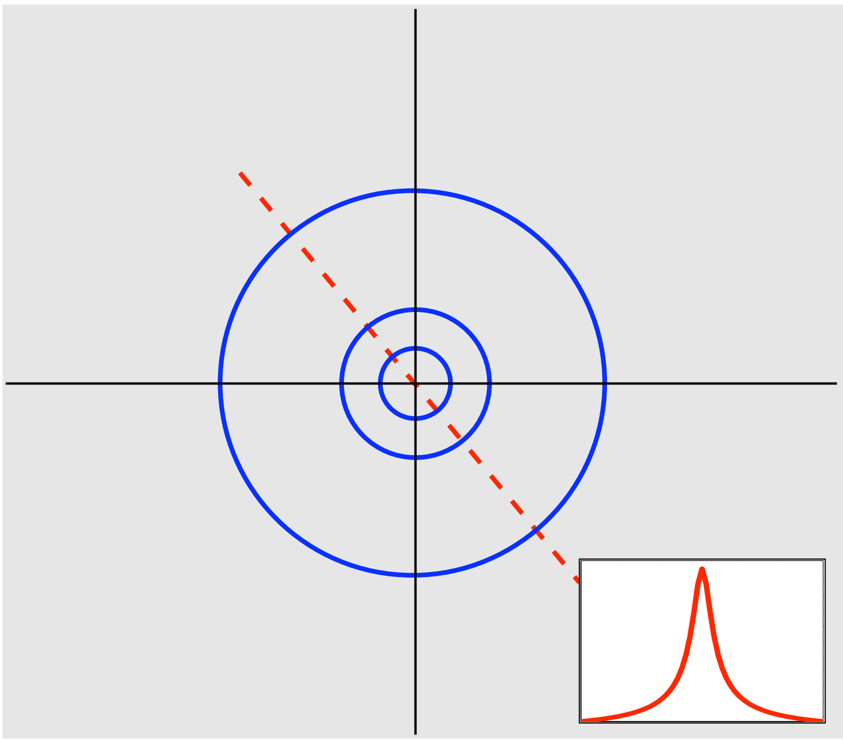


elliptical models of natural images

- Simoncelli, 1997;
- Zetzsche and Krieger, 1999;
- Huang and Mumford, 1999;
- Wainwright and Simoncelli, 2000;
- Hyvärinen et al., 2000;
- Parra et al., 2001;
- Srivastava et al., 2002;
- Sendur and Selesnick, 2002;
- Teh et al., 2003;
- Gehler and Welling, 2006
- etc.

[Fang et.al. 1990]

joint GSM model



$$\vec{x} = \vec{u}\sqrt{z}$$

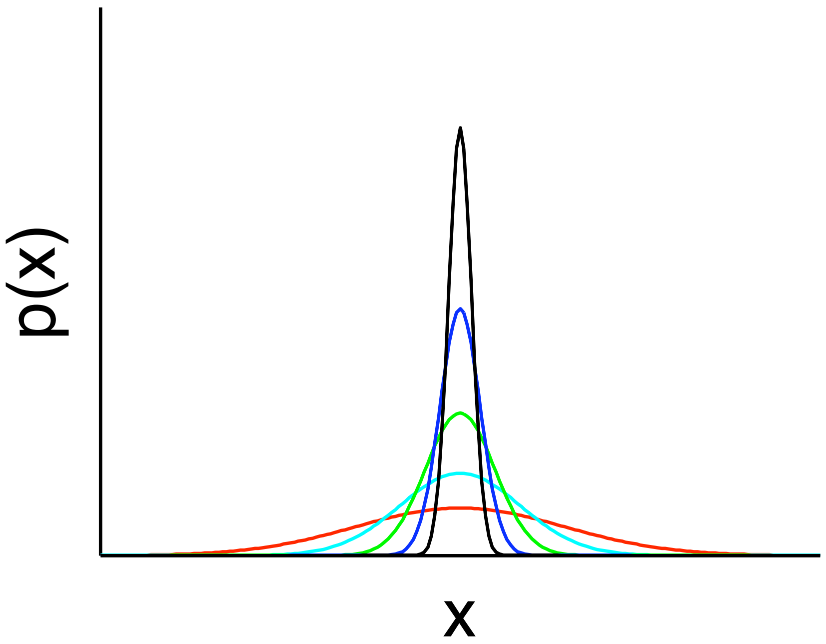
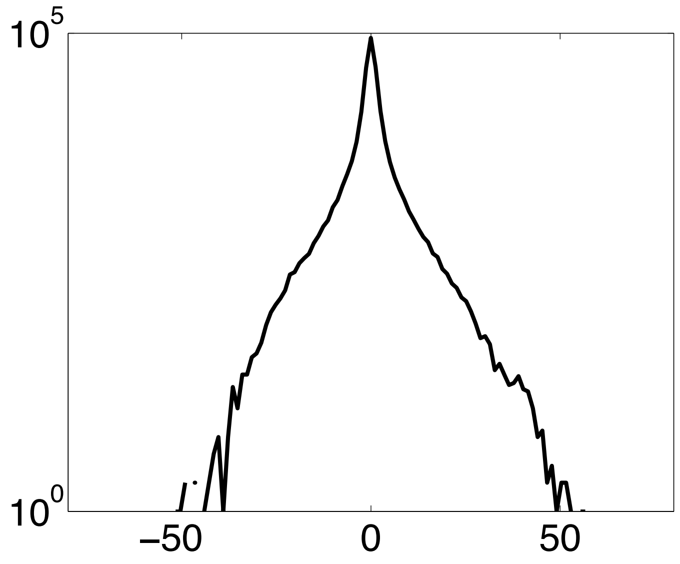
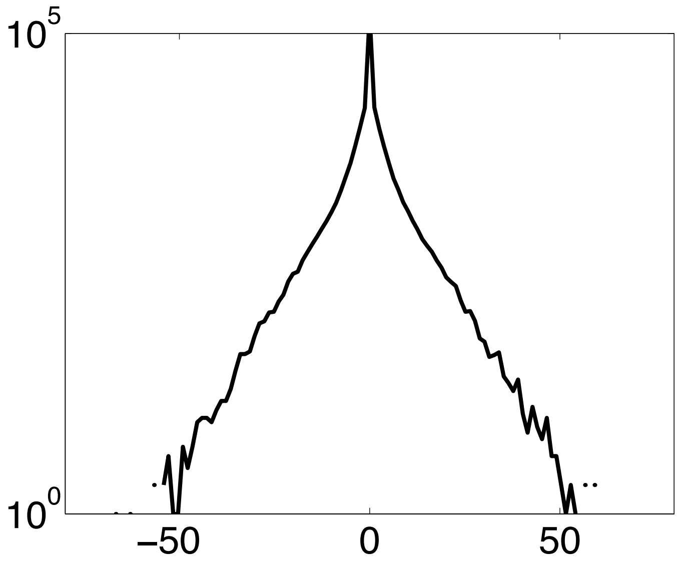


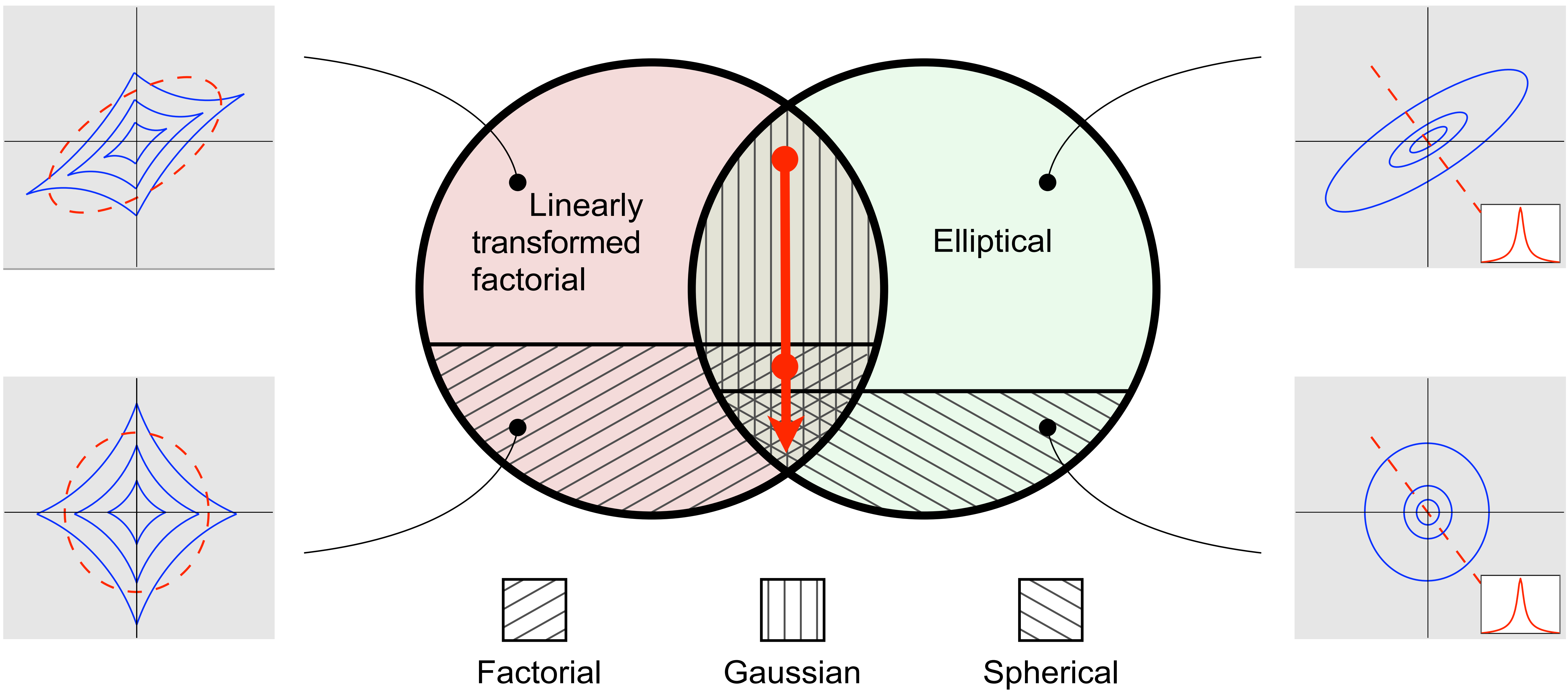
Image data

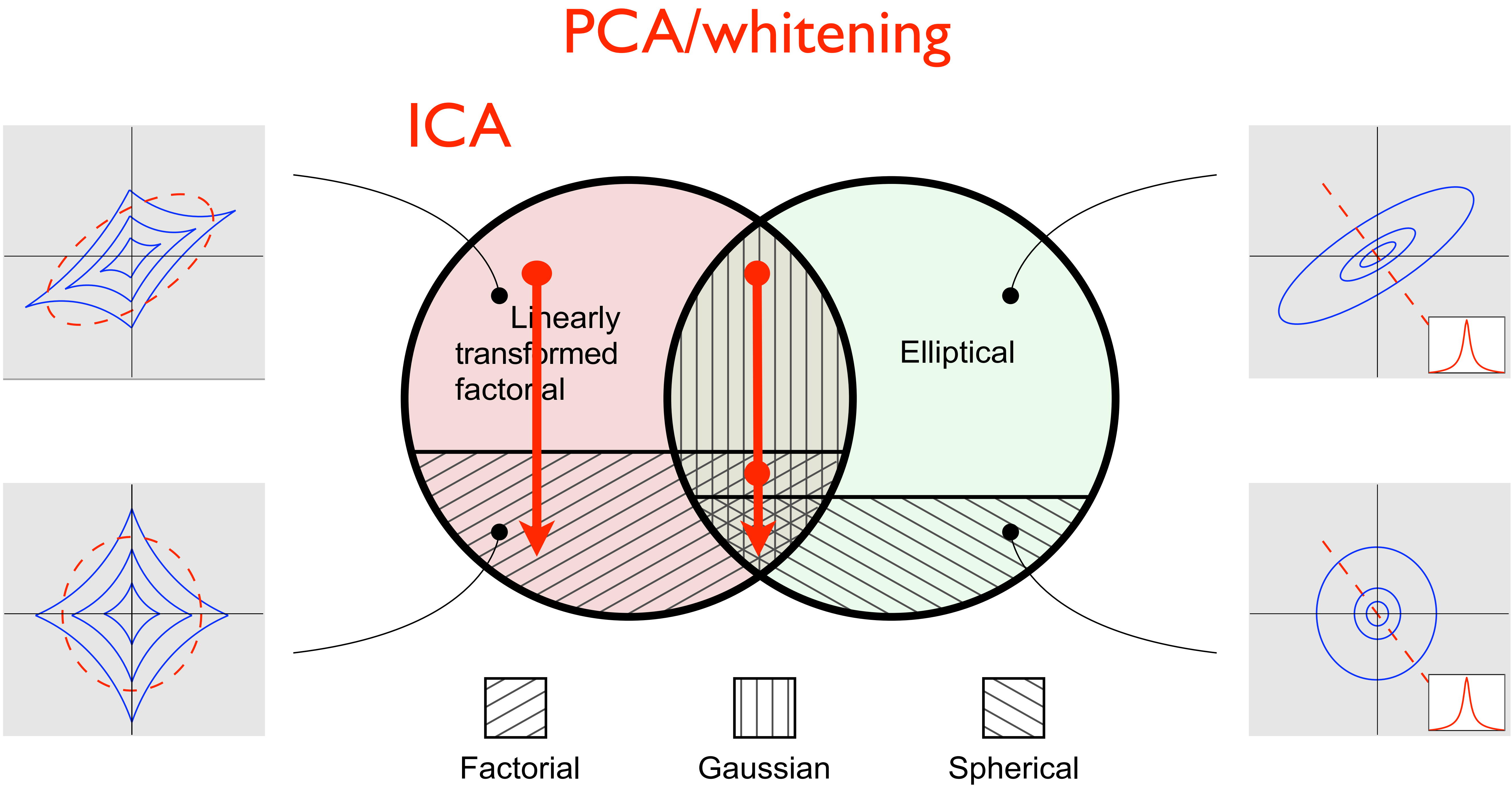


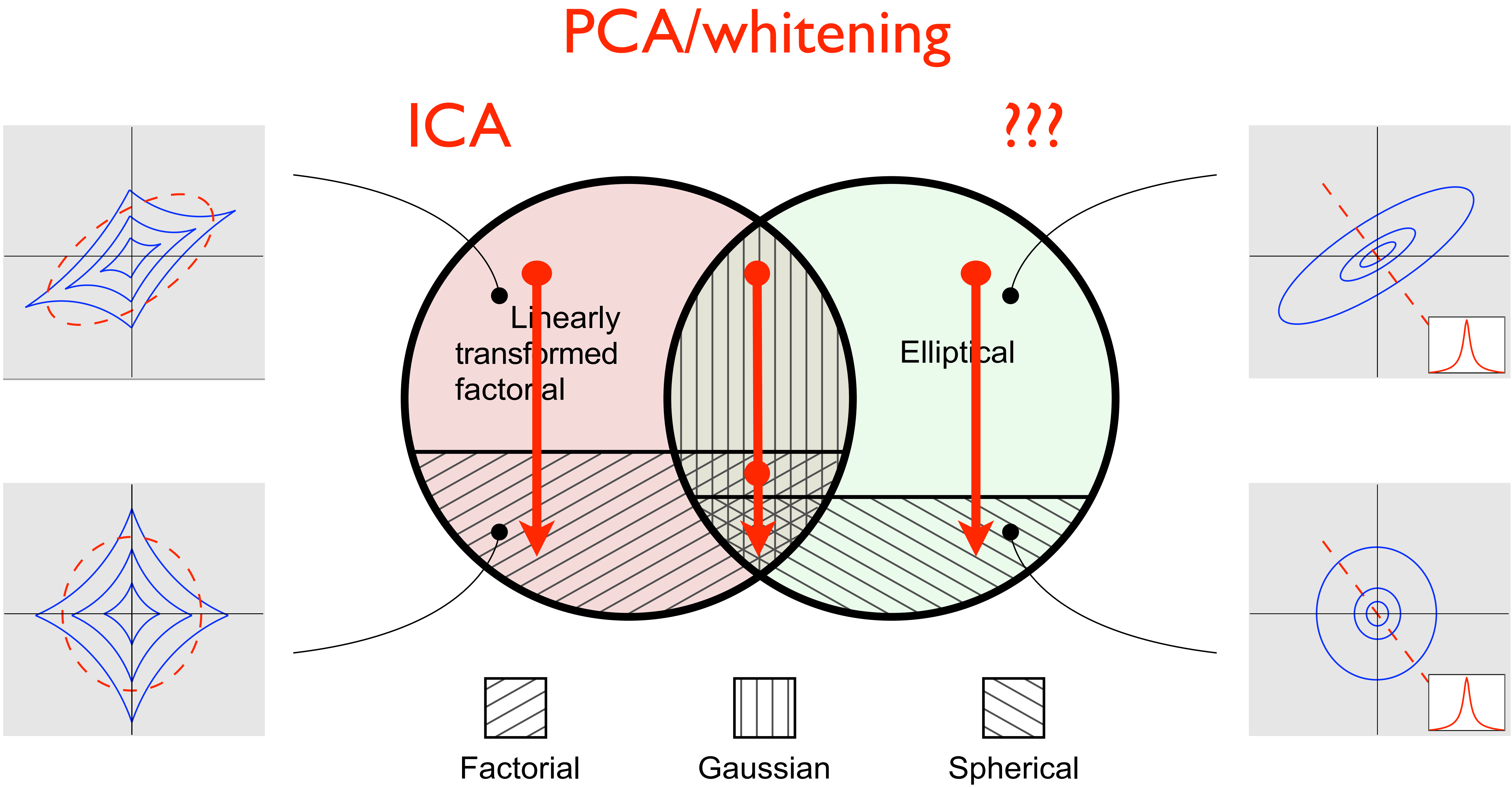
GSM simulation



PCA/whitening




PCA/whitening
ICA
Linearly transformed factorial
Elliptical
Factorial
Gaussian
Spherical

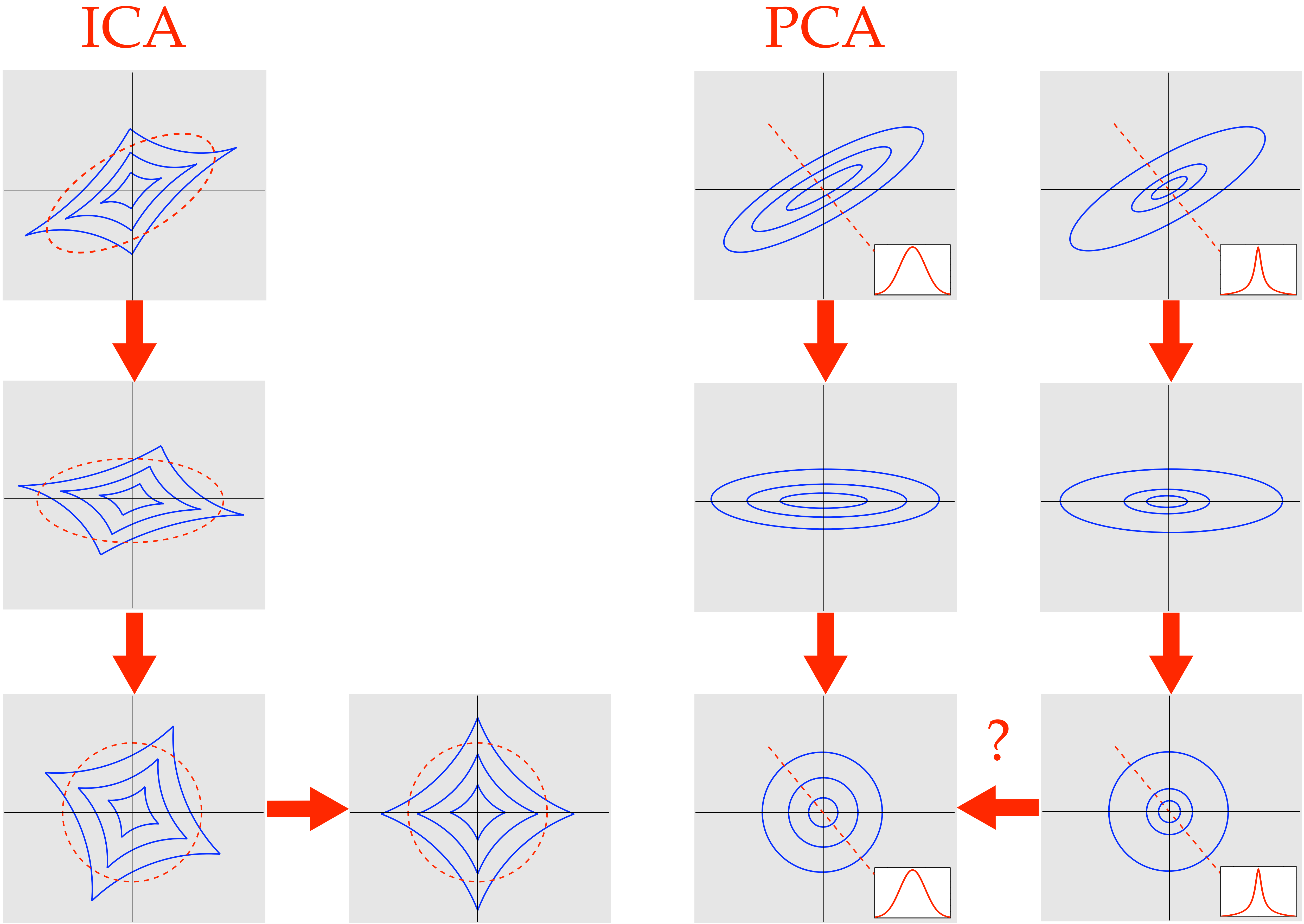

PCA/whitening
ICA
???
Linearly transformed factorial
Elliptical
Factorial
Gaussian
Spherical

nonlinear representations

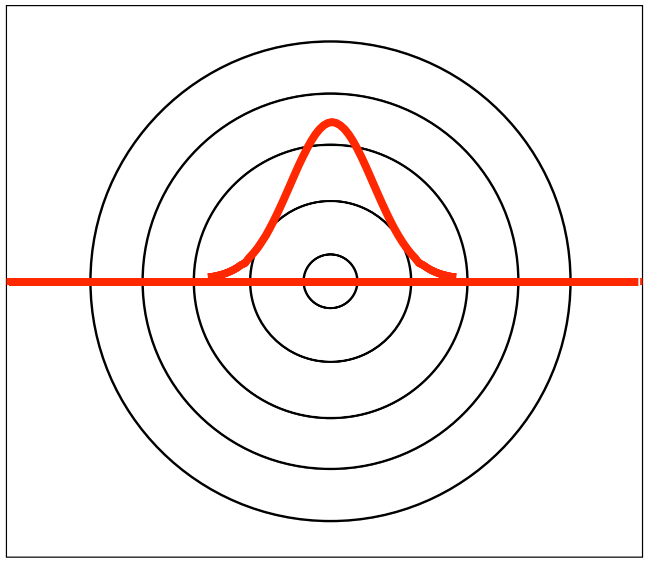
- complex wavelet phase-based [Ates & Orchid, 2003]
- orientation-based [Hammand & Simoncelli 2006]
- nonlinear whitening [Gluckman 2005]
- local divisive normalization [Malo et.al. 2004]
- global divisive normalization [Lyu & Simoncelli 2007,2008]

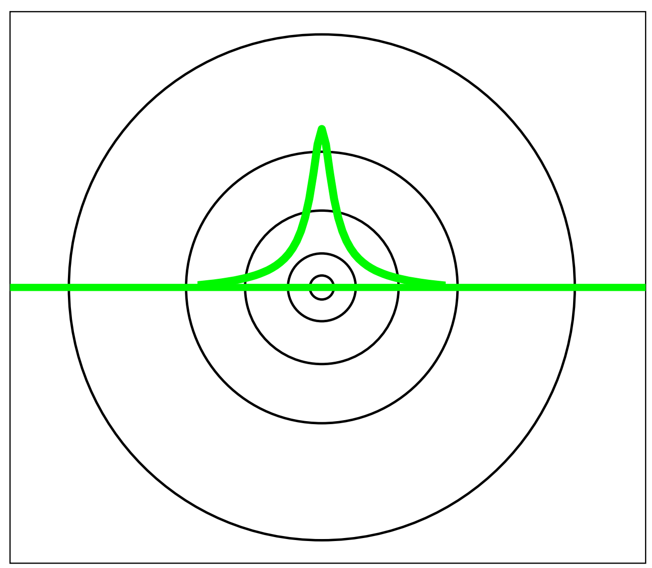
$$
\begin{aligned}
p(\vec{x}) &= \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{\vec{x}^T\vec{x}}{2}\right) \\
&= \frac{1}{\sqrt{(2\pi)^d}} \exp\left(-\frac{1}{2}\sum_{i=1}^{d} x_i^2\right) \\
&= \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_i^2\right) \\
&= \prod_{i=1}^{d} p(x_i)
\end{aligned}
$$

Gaussian is the **only** density that can be both factorial and spherically symmetric [Nash and Klamkin 1976]


ICA
PCA
?

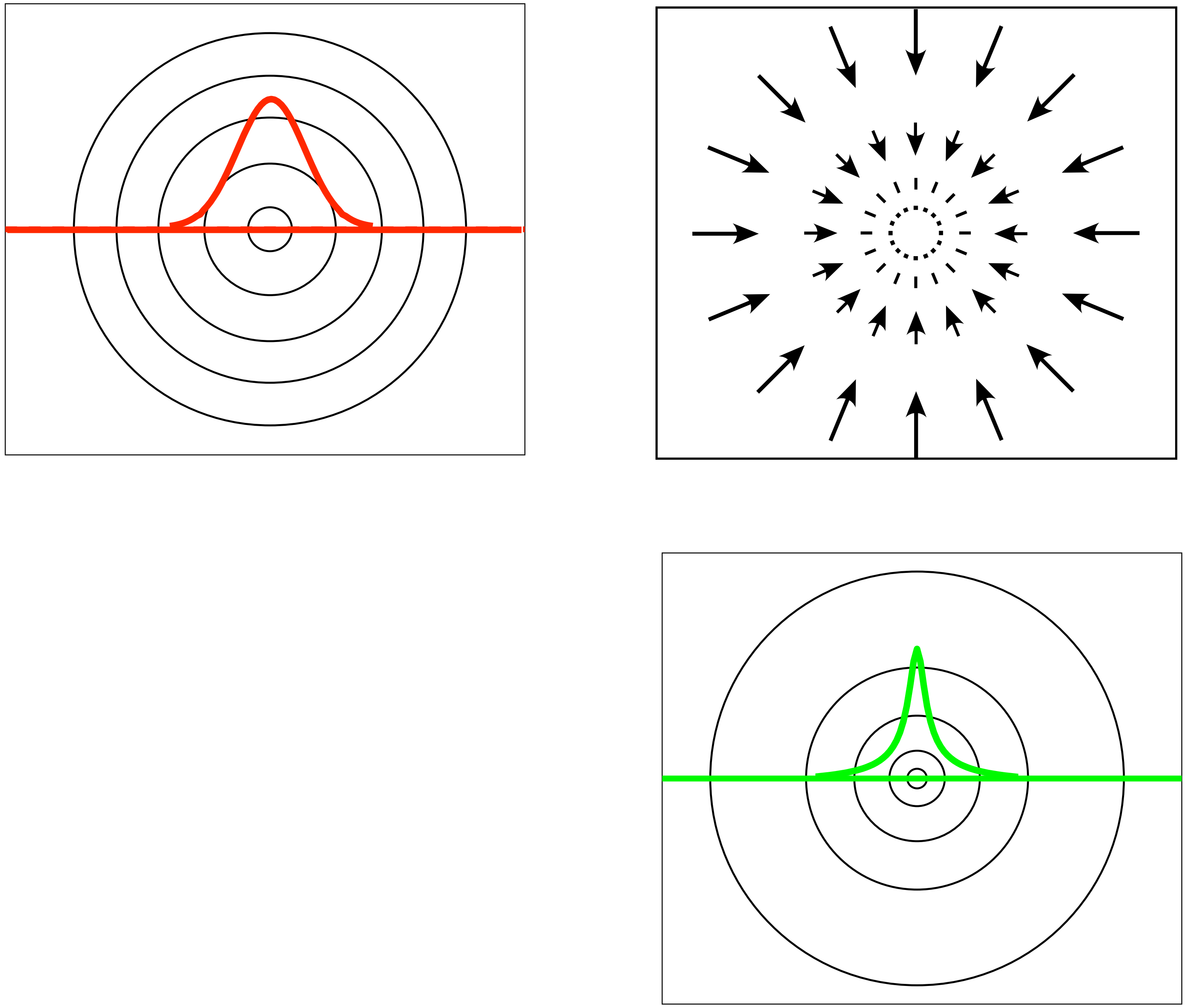
radial Gaussianization (RG)





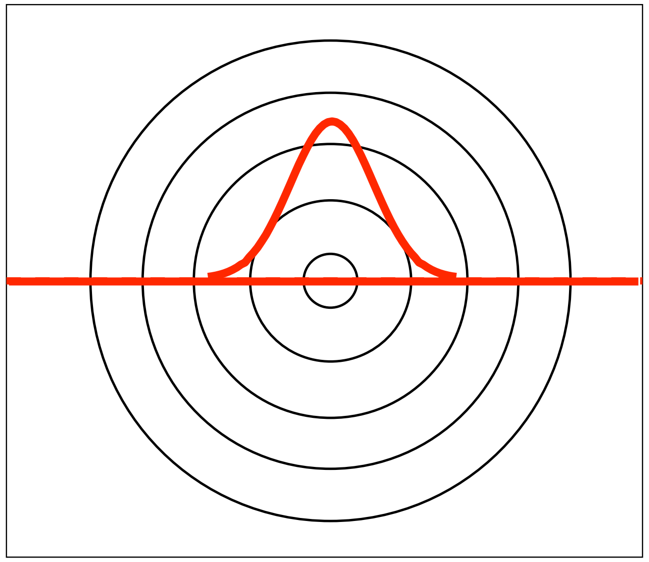
[Lyu & Simoncelli, 2008,2009]

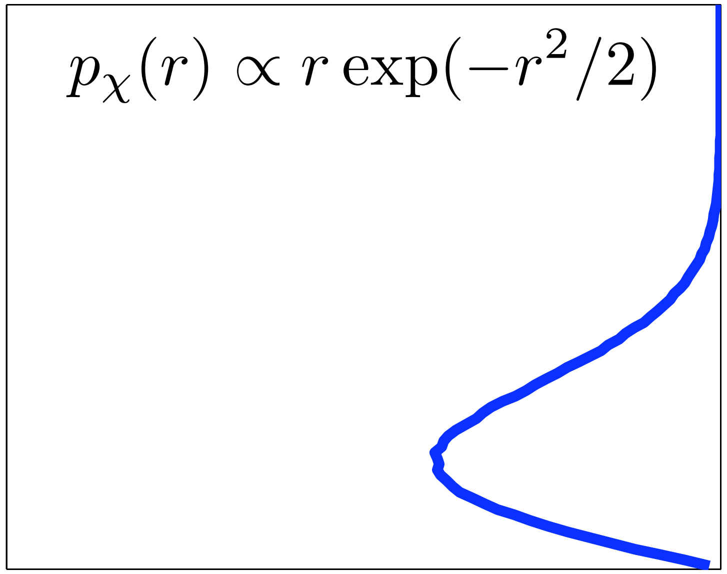
radial Gaussianization (RG)

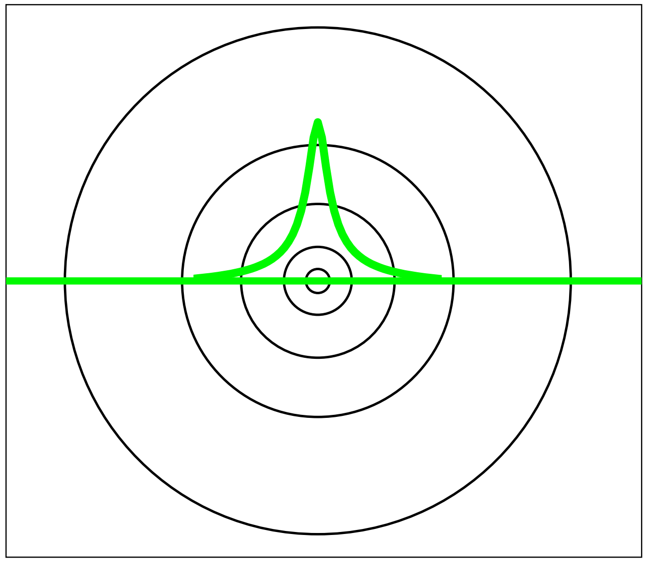


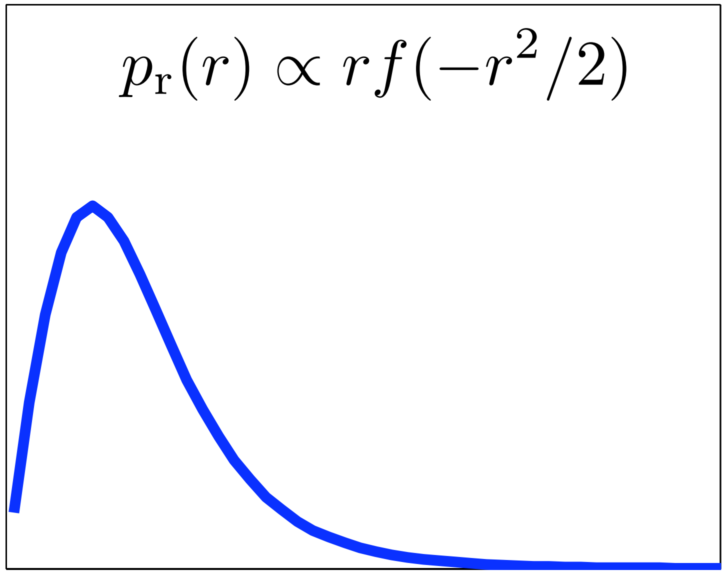
[Lyu & Simoncelli, 2008,2009]

radial Gaussianization (RG)



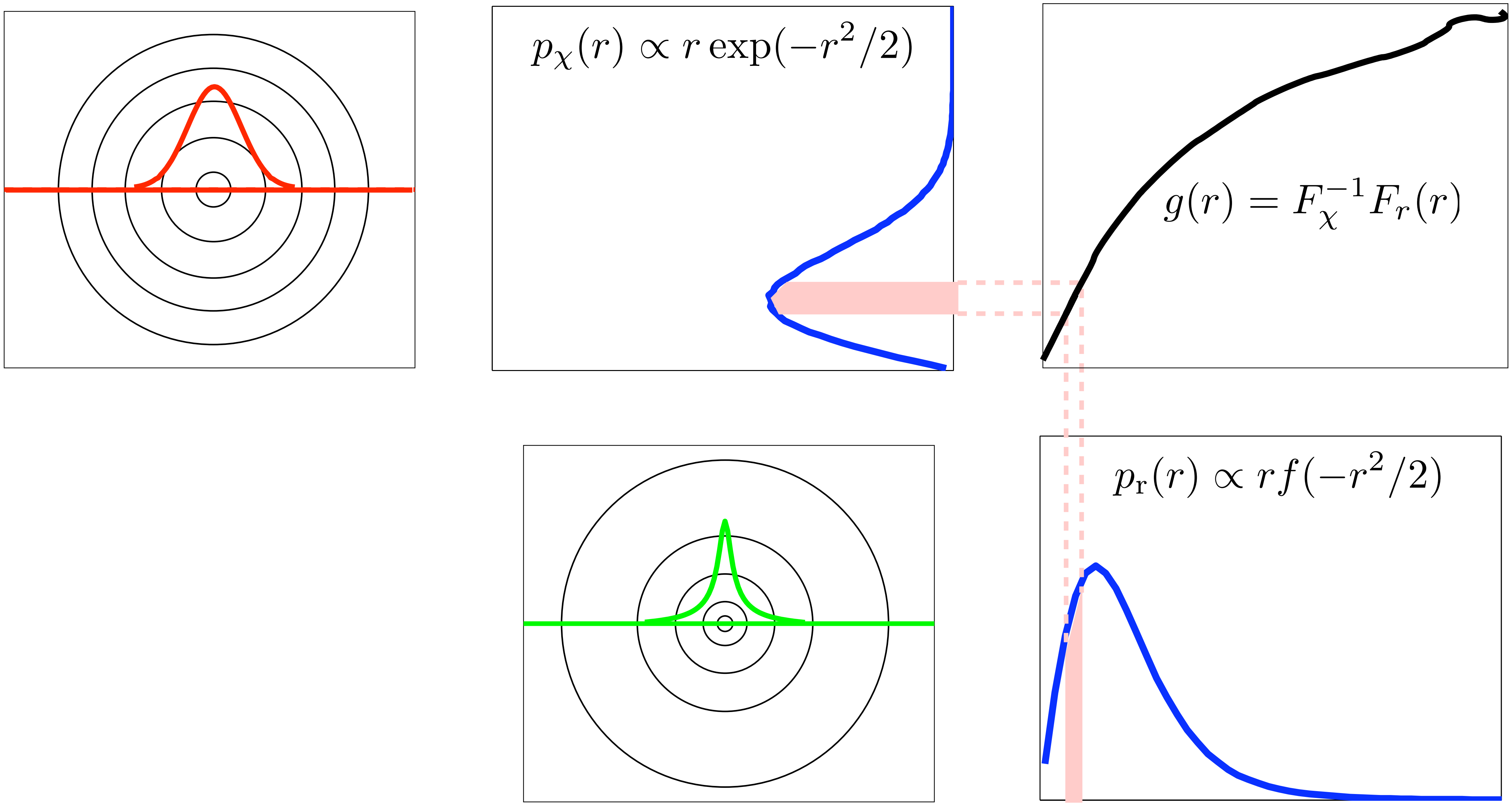






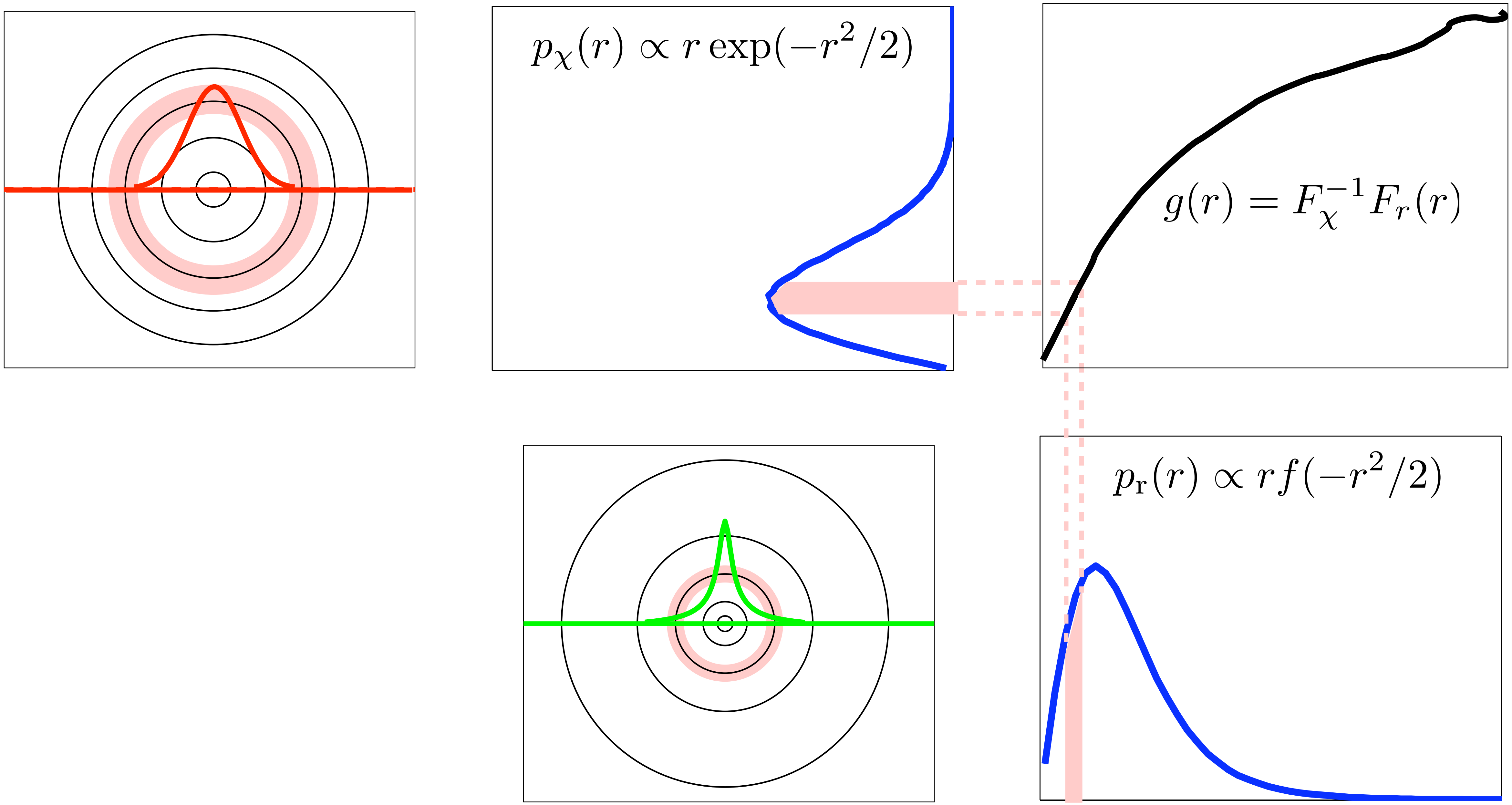
[Lyu & Simoncelli, 2008,2009]

radial Gaussianization (RG)



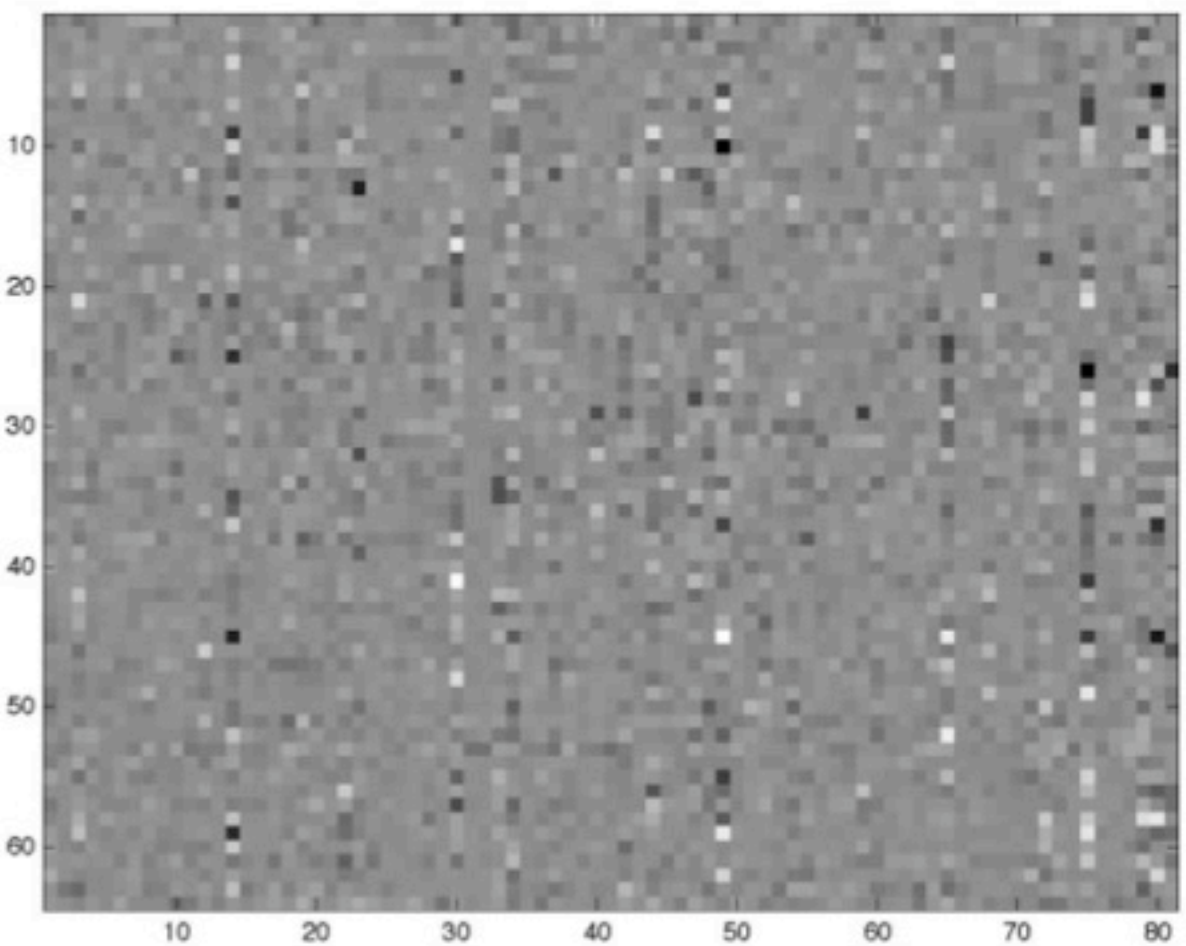
[Lyu & Simoncelli, 2008,2009]

radial Gaussianization (RG)

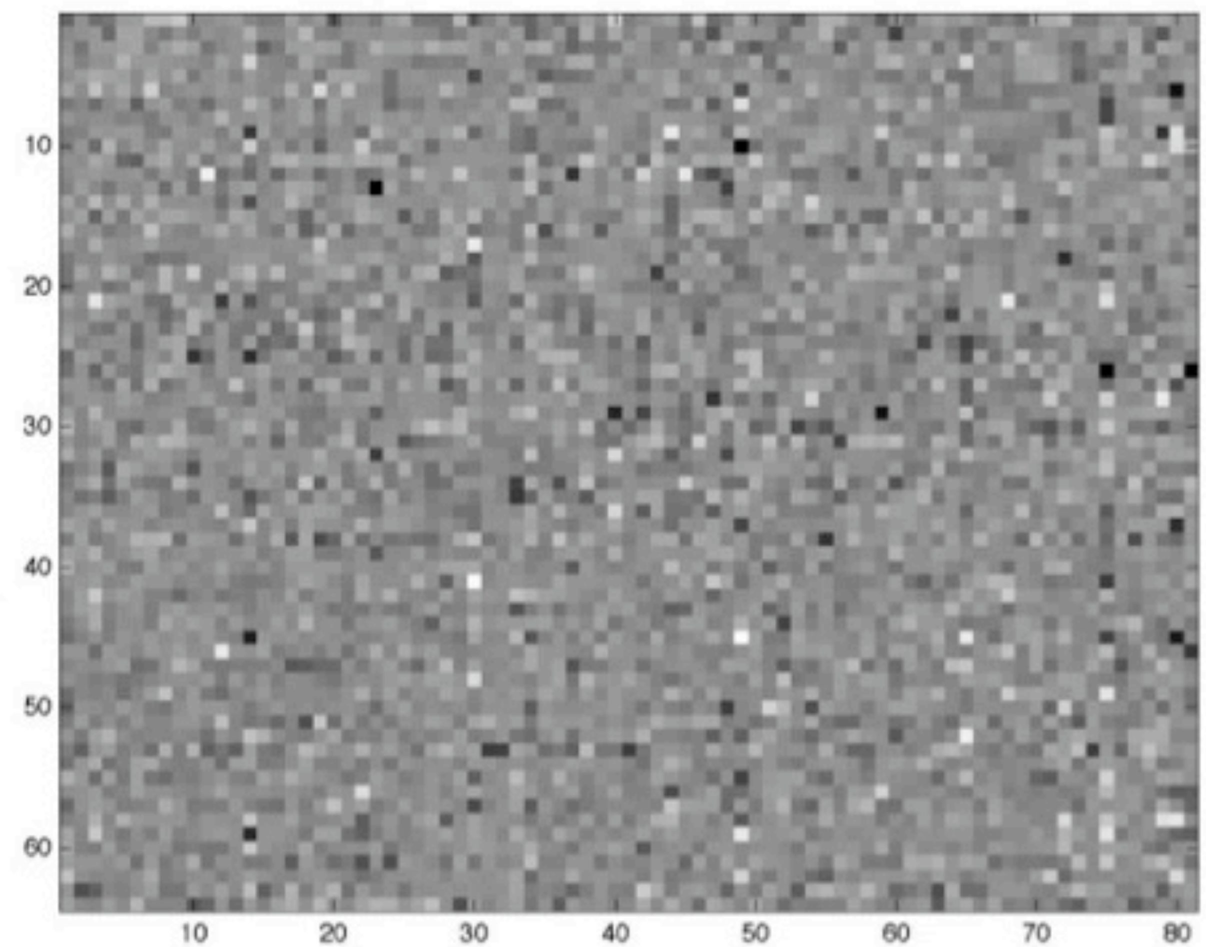


$$\vec{x}_{\mathrm{rg}} = \frac{g(\|\vec{x}_{\mathrm{wht}}\|)}{\|\vec{x}_{\mathrm{wht}}\|}\vec{x}_{\mathrm{wht}}$$

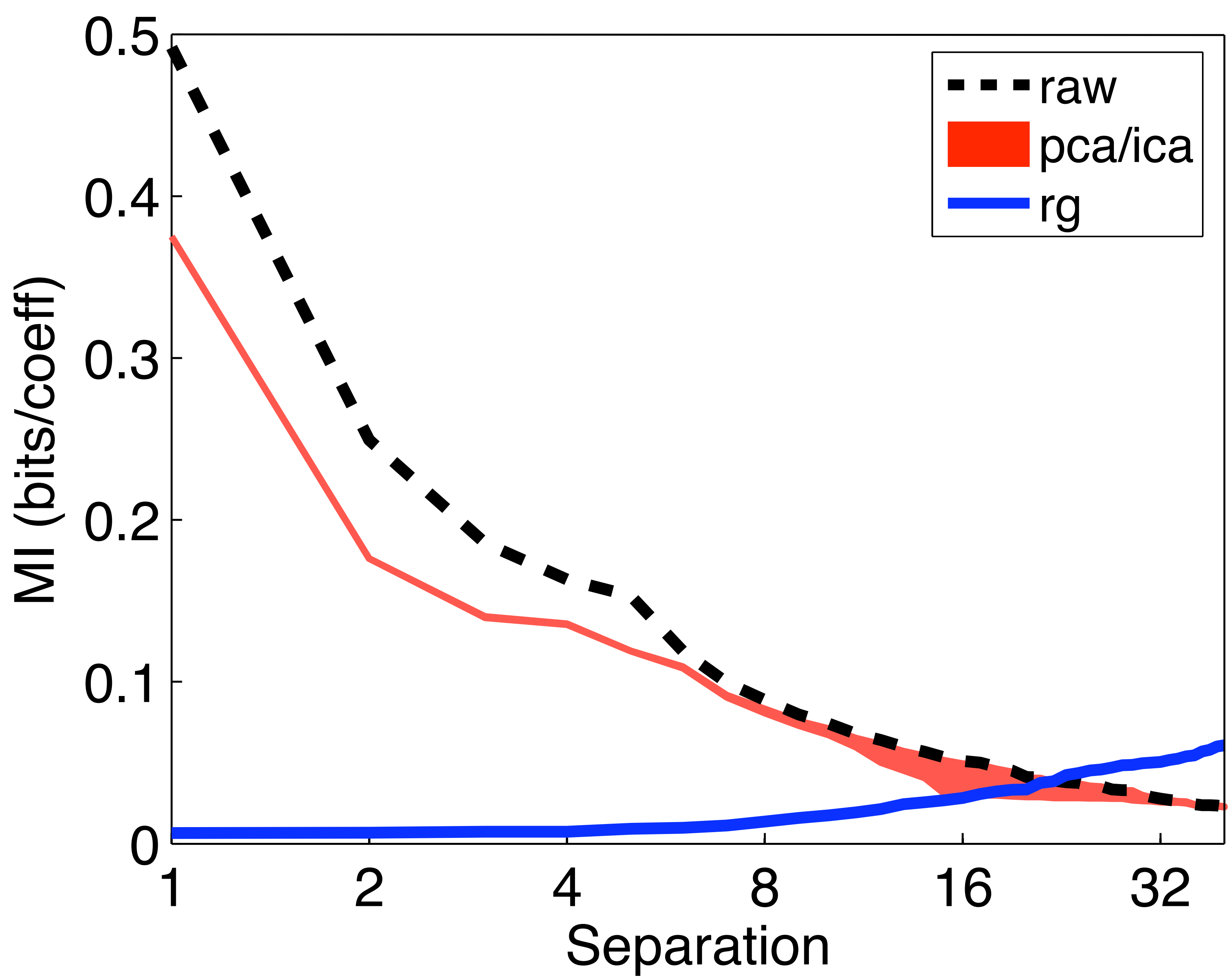
[Lyu & Simoncelli, 2008,2009]



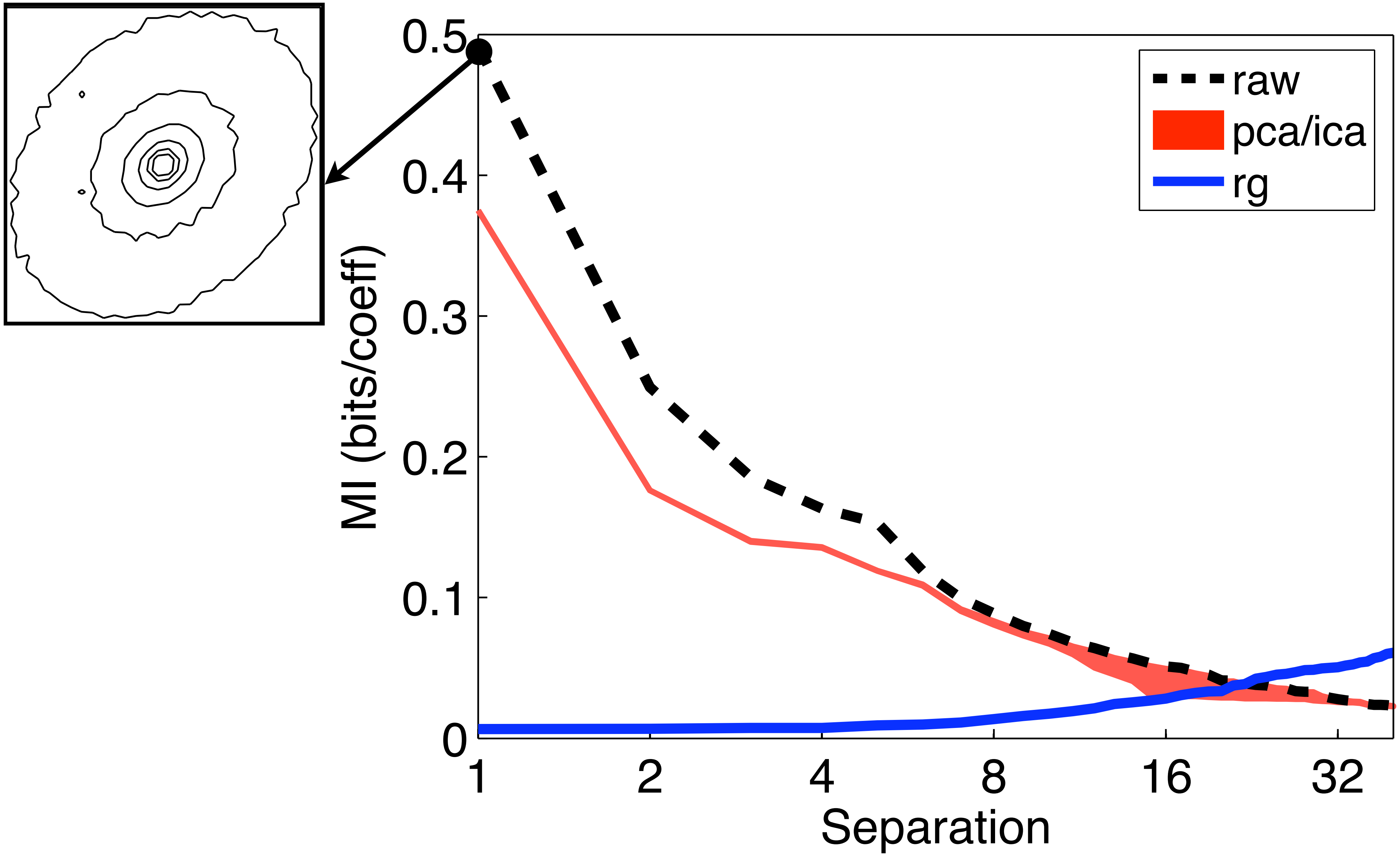
ICA coefficients



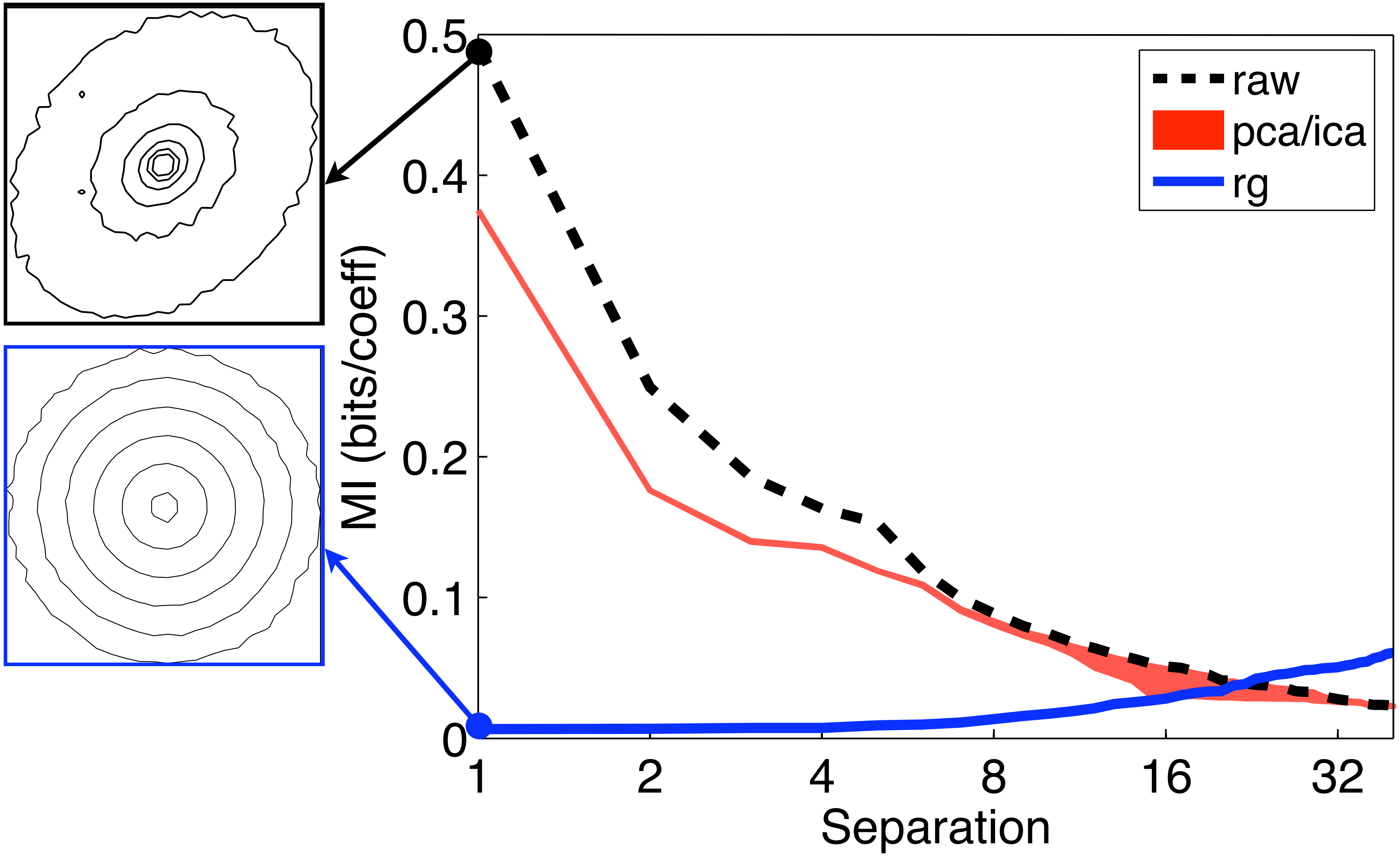
Radially factorized coefficients

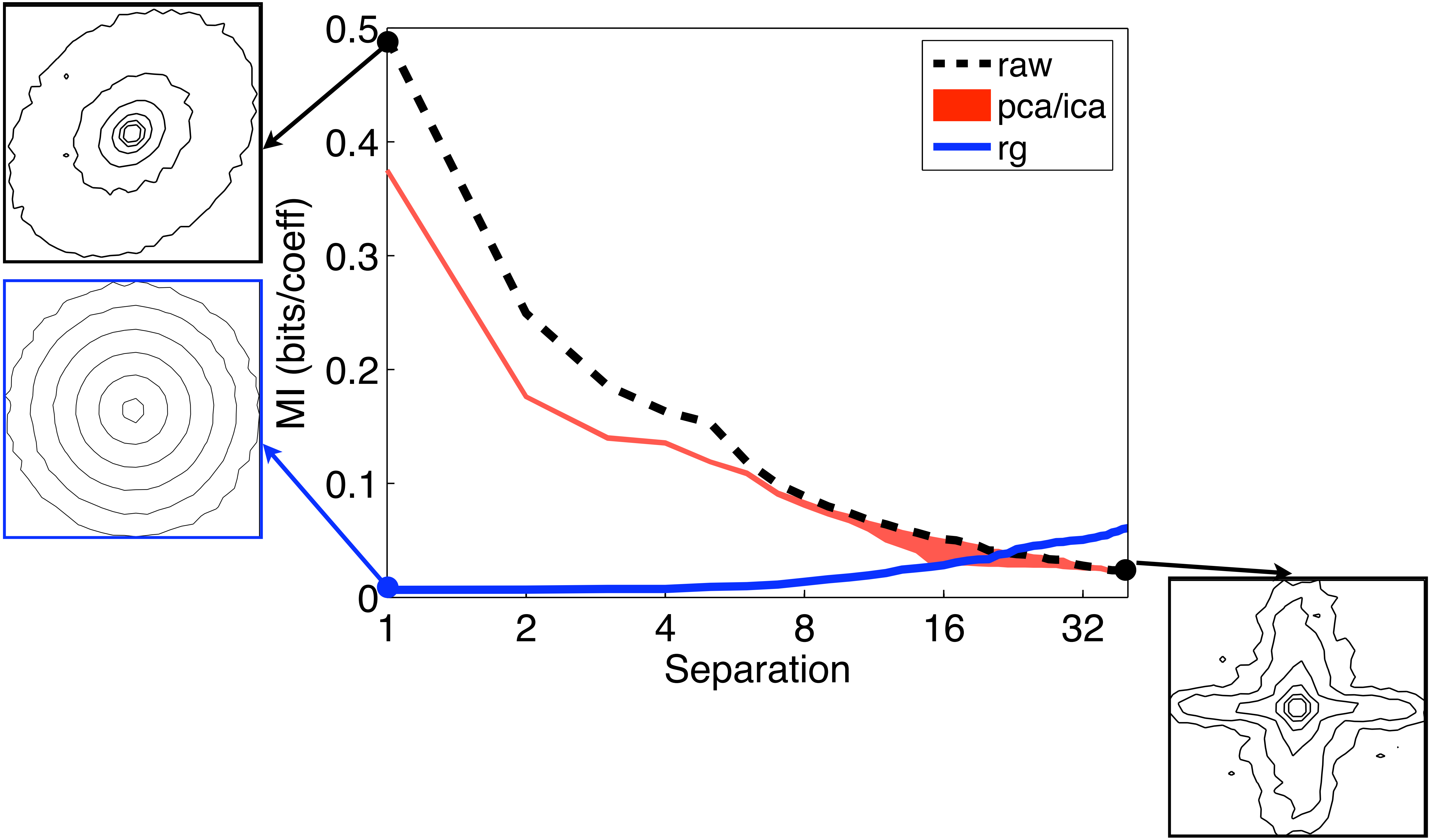


raw
pca/ica
rg
MI (bits/coeff)
Separation
0.5
0.4
0.3
0.2
0.1
0
1
2
4
8
16
32

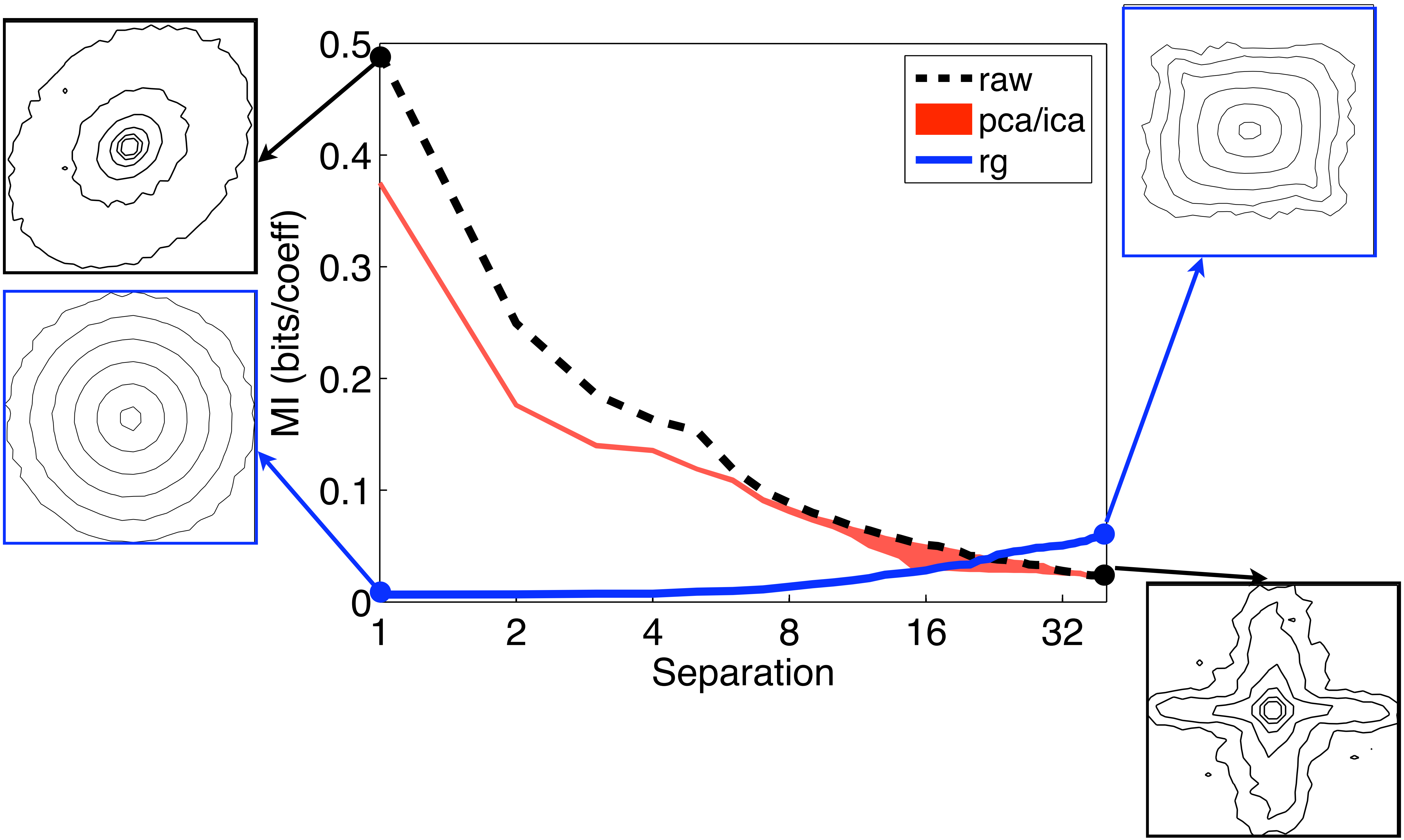


0.5
0.4
0.3
0.2
0.1
0
MI (bits/coeff)
1
2
4
8
16
32
Separation
raw
pca/ica
rg

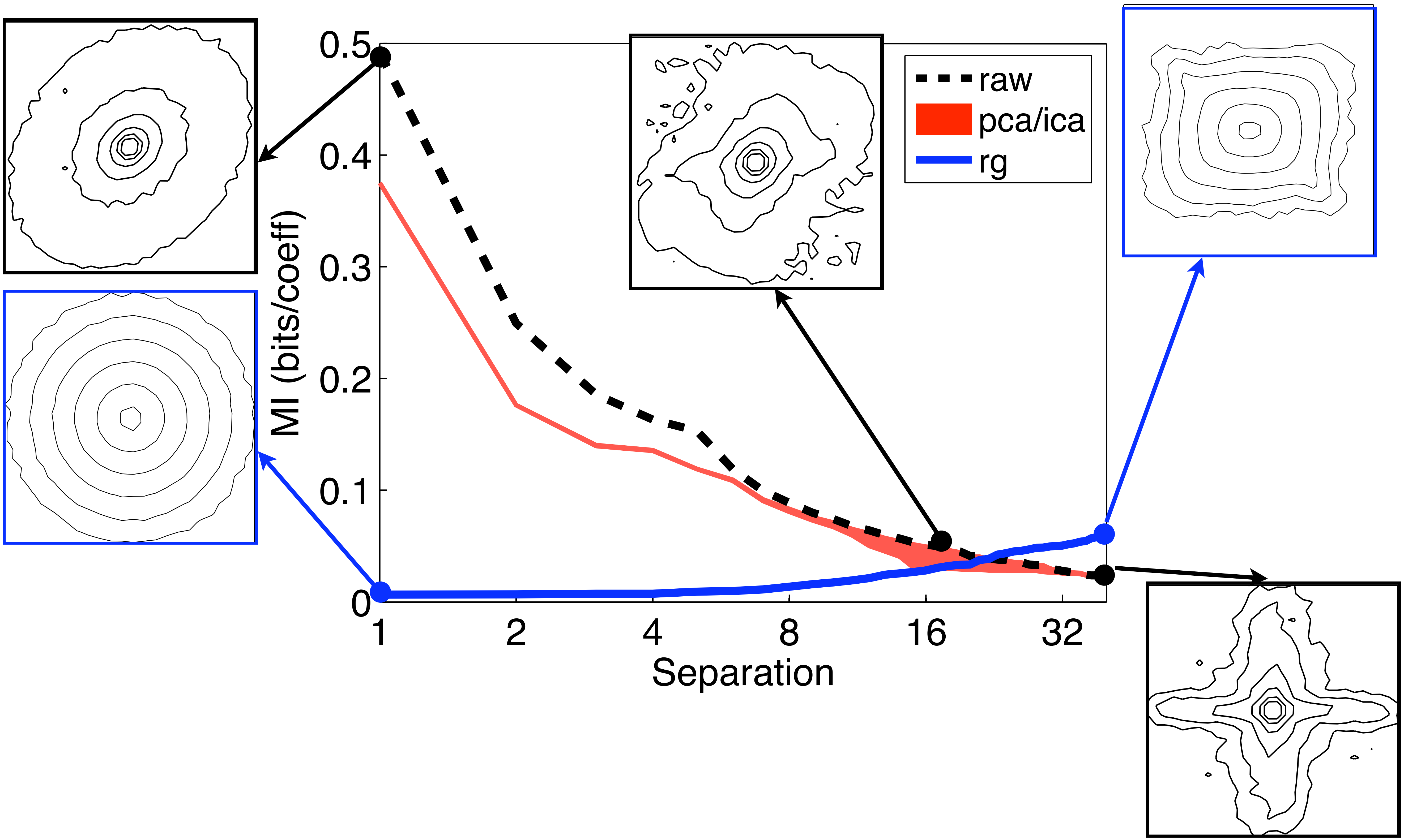

0.5
0.4
0.3
0.2
0.1
0
MI (bits/coeff)
1
2
4
8
16
32
Separation
raw
pca/ica
rg



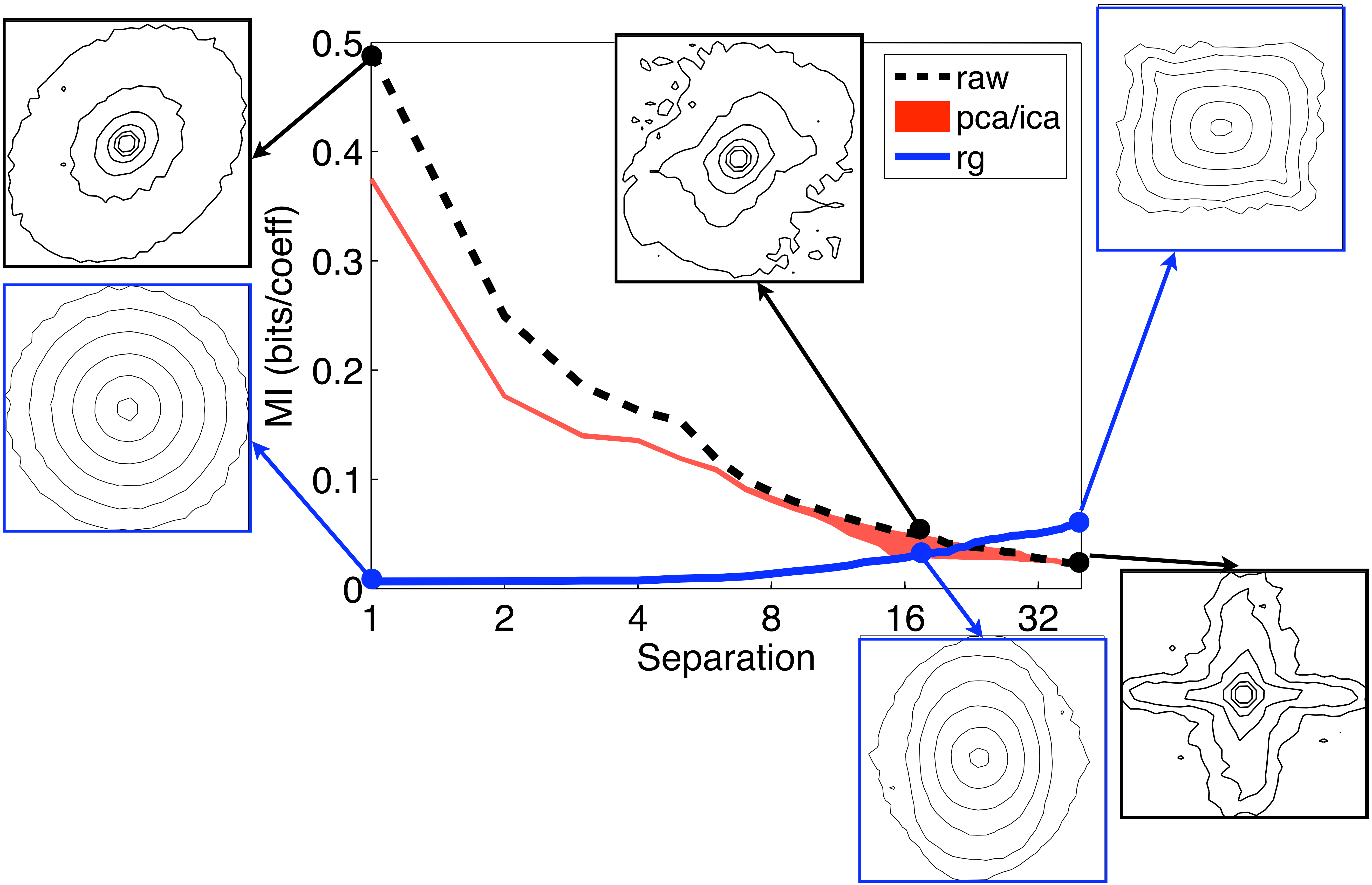
0.5
0.4
0.3
0.2
0.1
0
MI (bits/coeff)
1
2
4
8
16
32
Separation
raw
pca/ica
rg



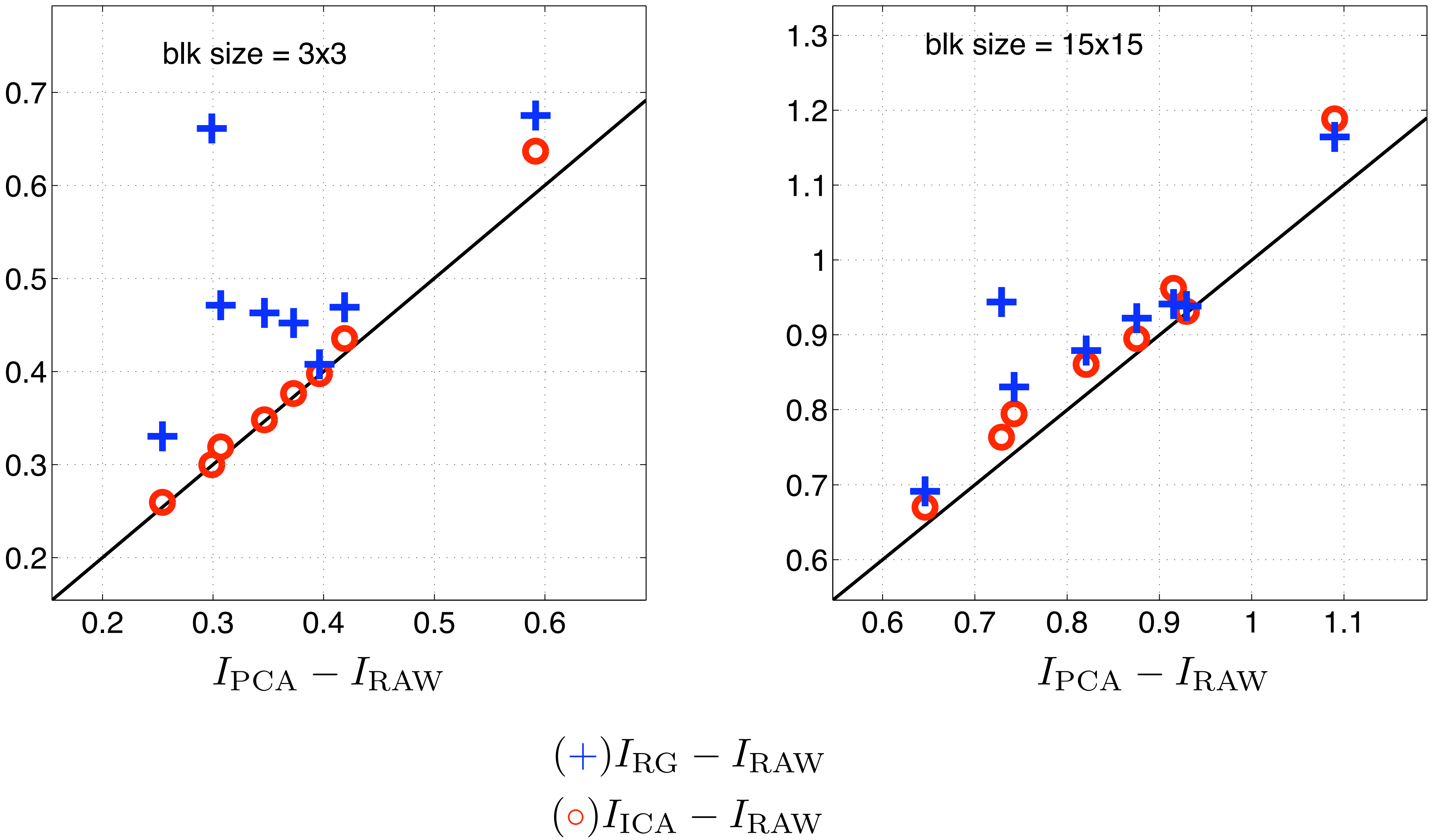
0.5
0.4
0.3
0.2
0.1
0
MI (bits/coeff)
1
2
4
8
16
32
Separation
raw
pca/ica
rg



0.5
0.4
0.3
0.2
0.1
0
MI (bits/coeff)
1
2
4
8
16
32
Separation
raw
pca/ica
rg

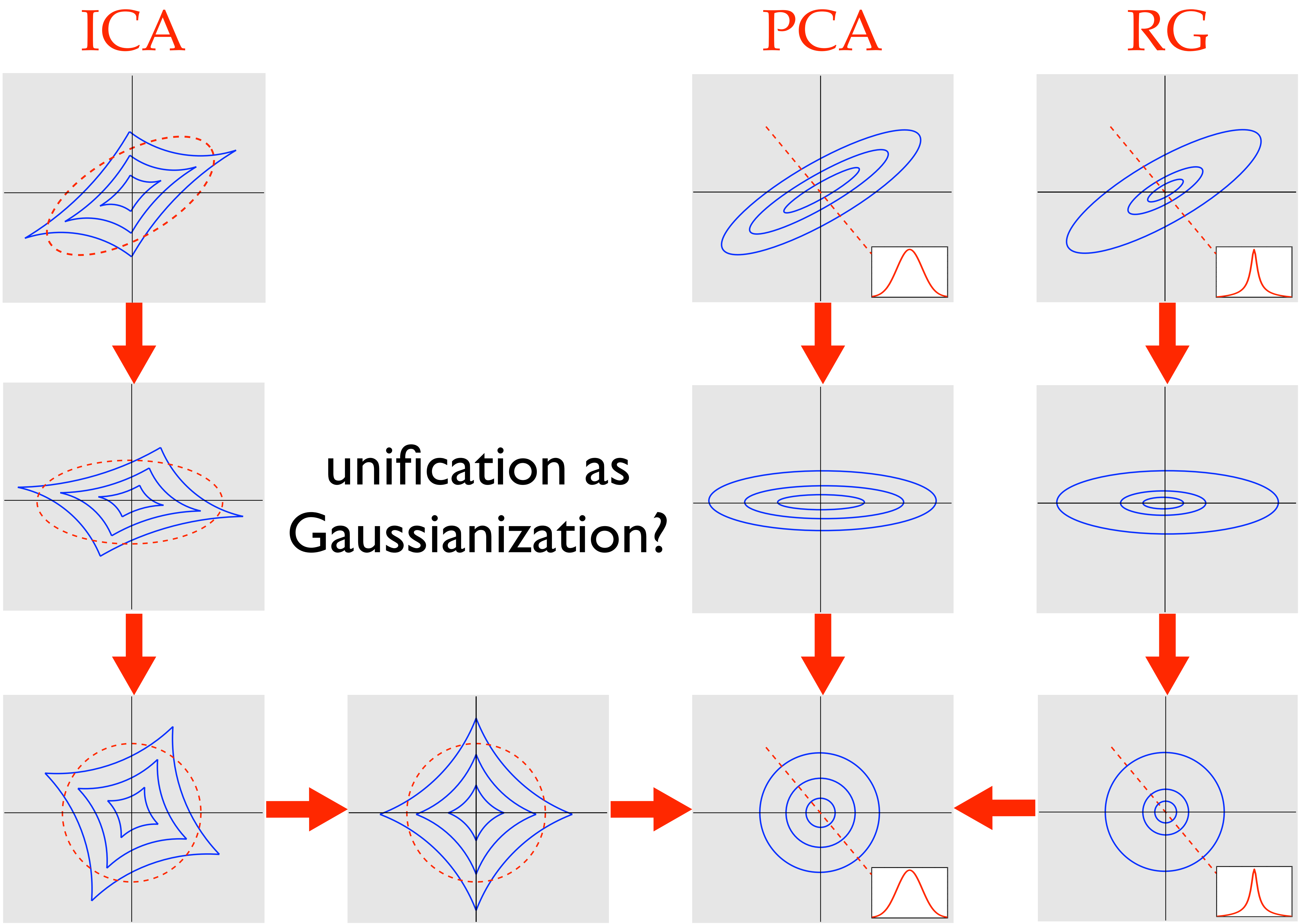


0.5
0.4
0.3
0.2
0.1
0
MI (bits/coeff)
1
2
4
8
16
32
Separation
raw
pca/ica
rg

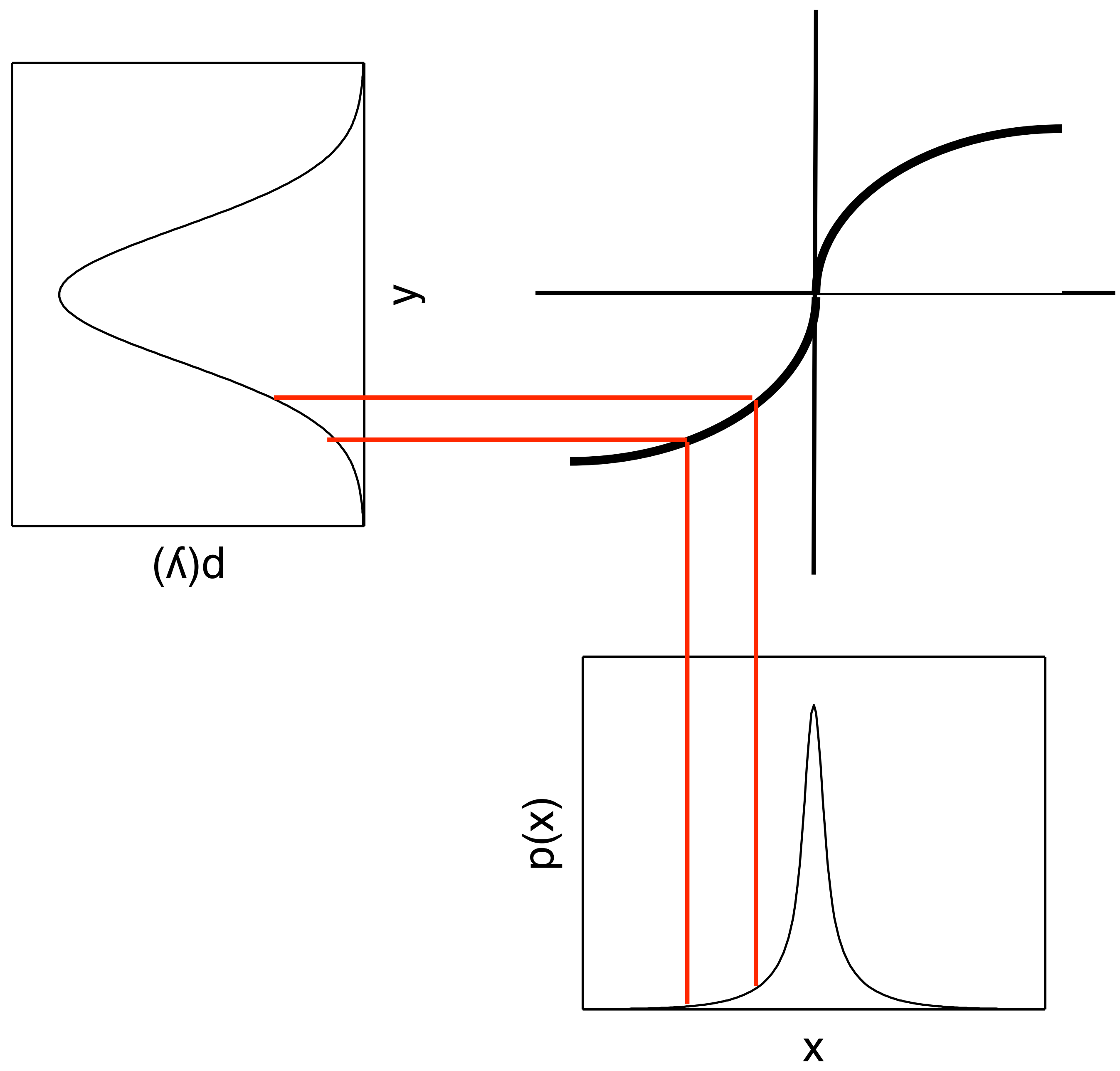


blocks of local mean removed pixel blocks of natural images

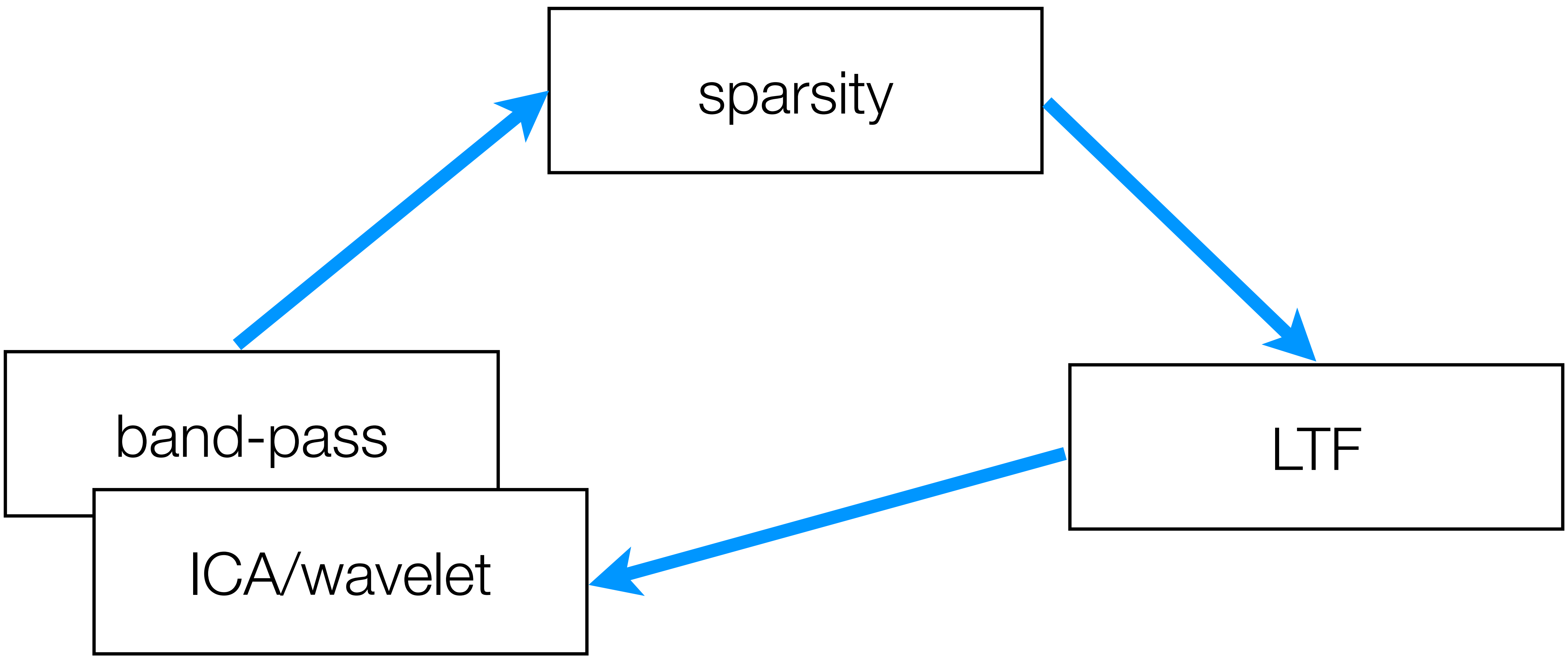
(Lyu & Simoncelli, Neural Computation, to appear)


ICA
PCA
RG
unification as
Gaussianization?

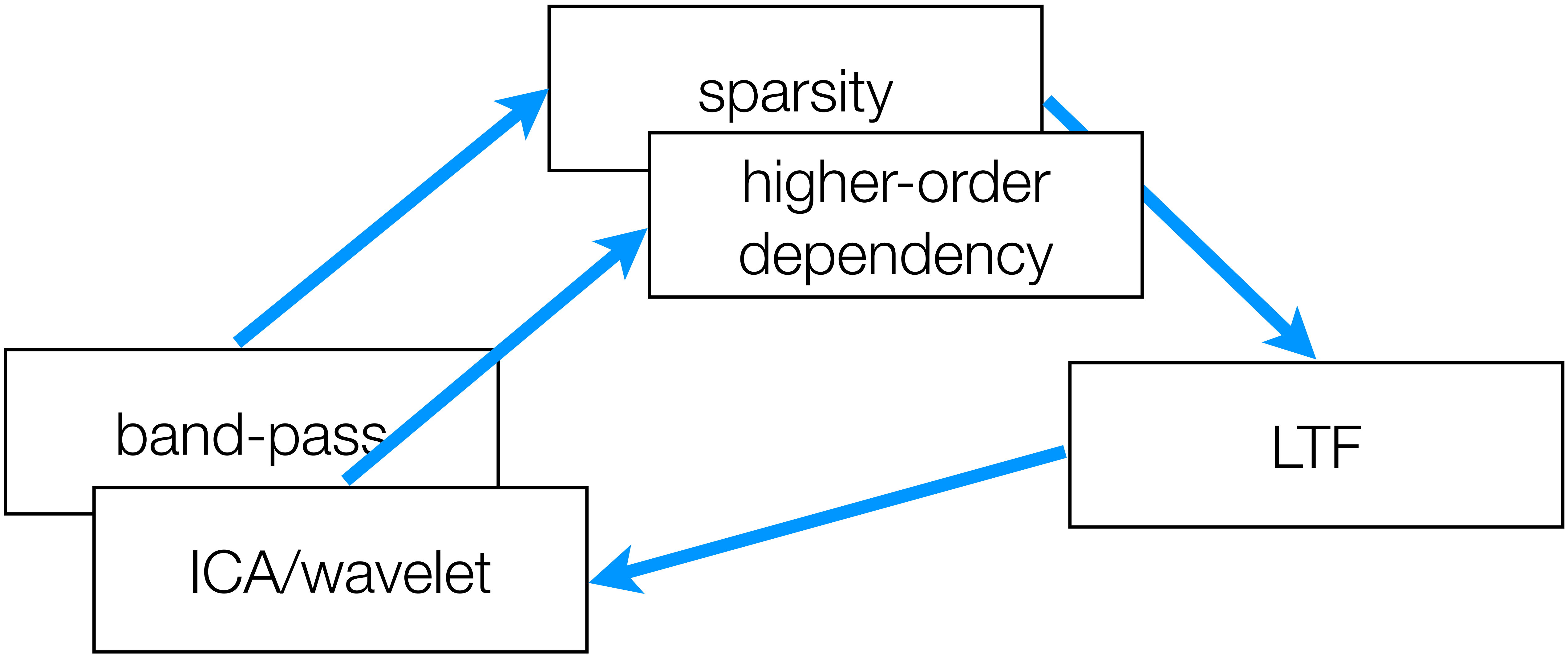
marginal Gaussianization



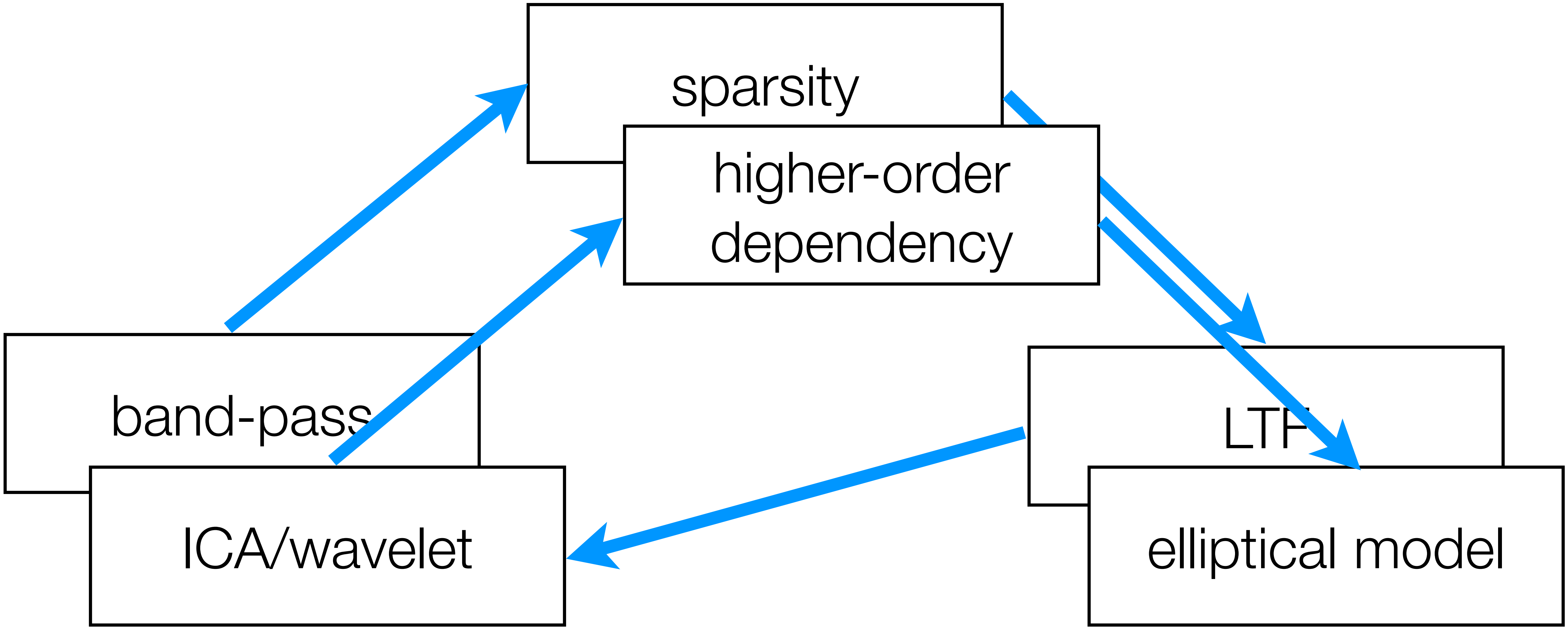
summary



summary

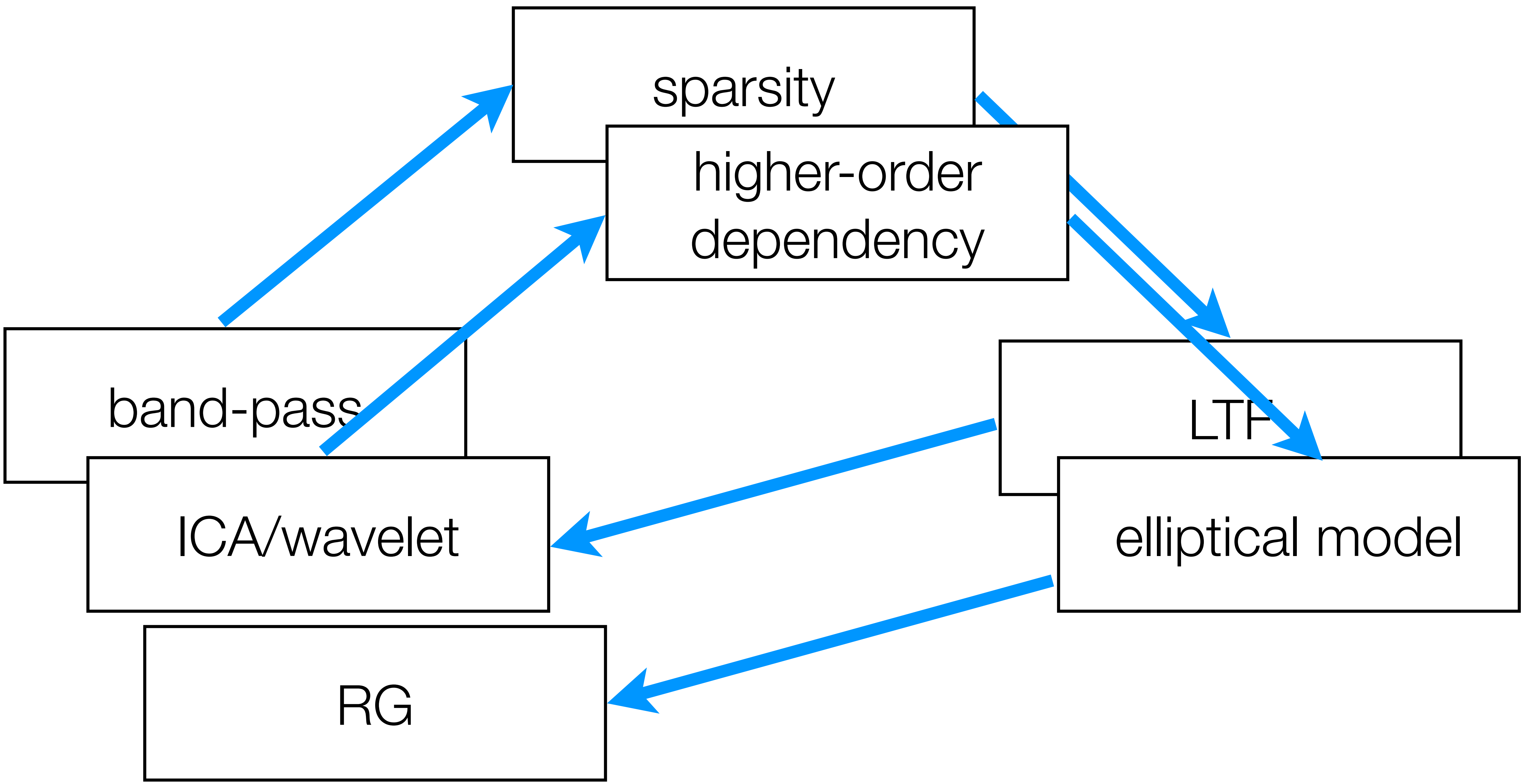

sparsity
higher-order dependency
band-pass
ICA/wavelet
LTF

summary

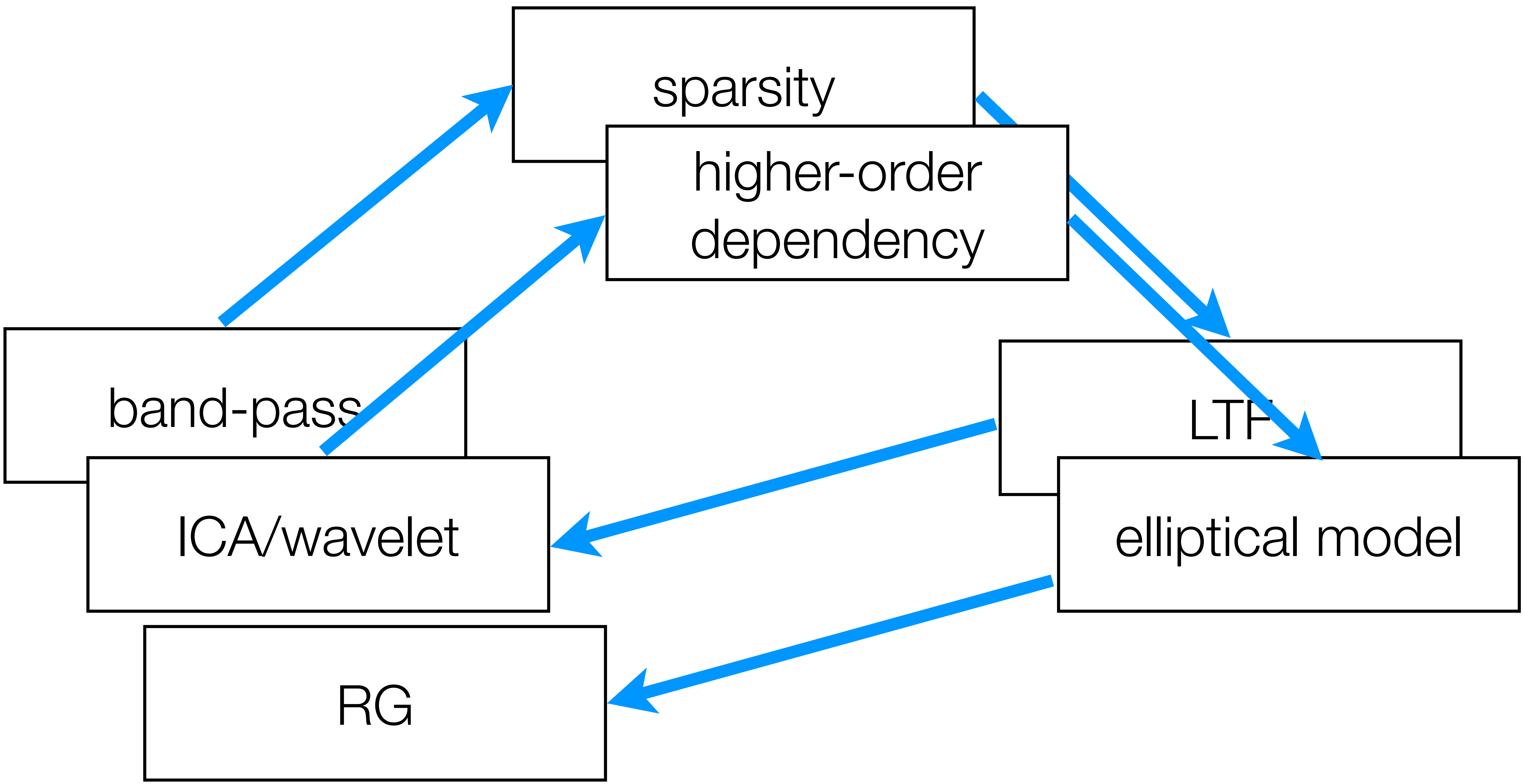


sparsity
higher-order
dependency
band-pass
ICA/wavelet
LTF
elliptical model

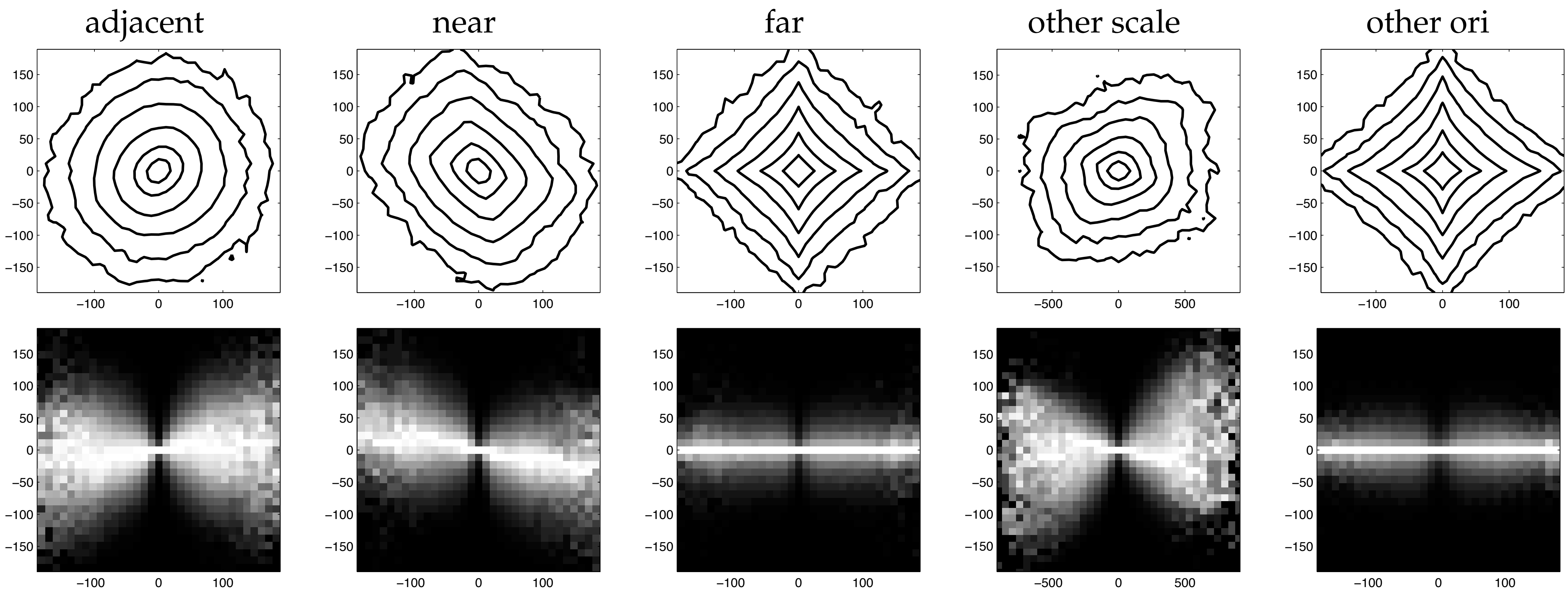
summary


sparsity
higher-order dependency
band-pass
ICA/wavelet
LTF
elliptical model
RG

summary



Not enough!

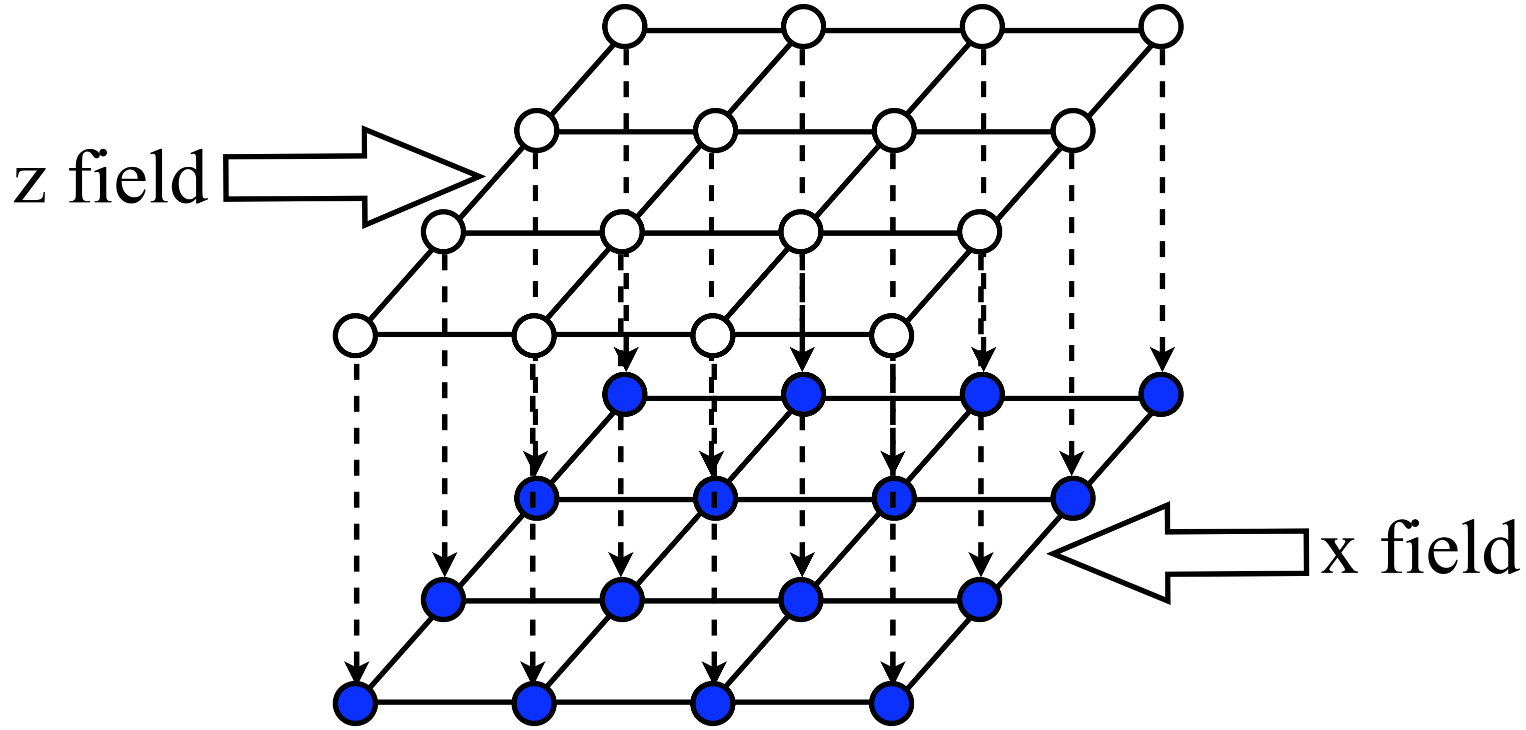


- Nearby: densities are approximately circular/elliptical
- Distant: densities are approximately factorial

[Simoncelli, ‘97; Wainwright&Simoncelli, ‘99]

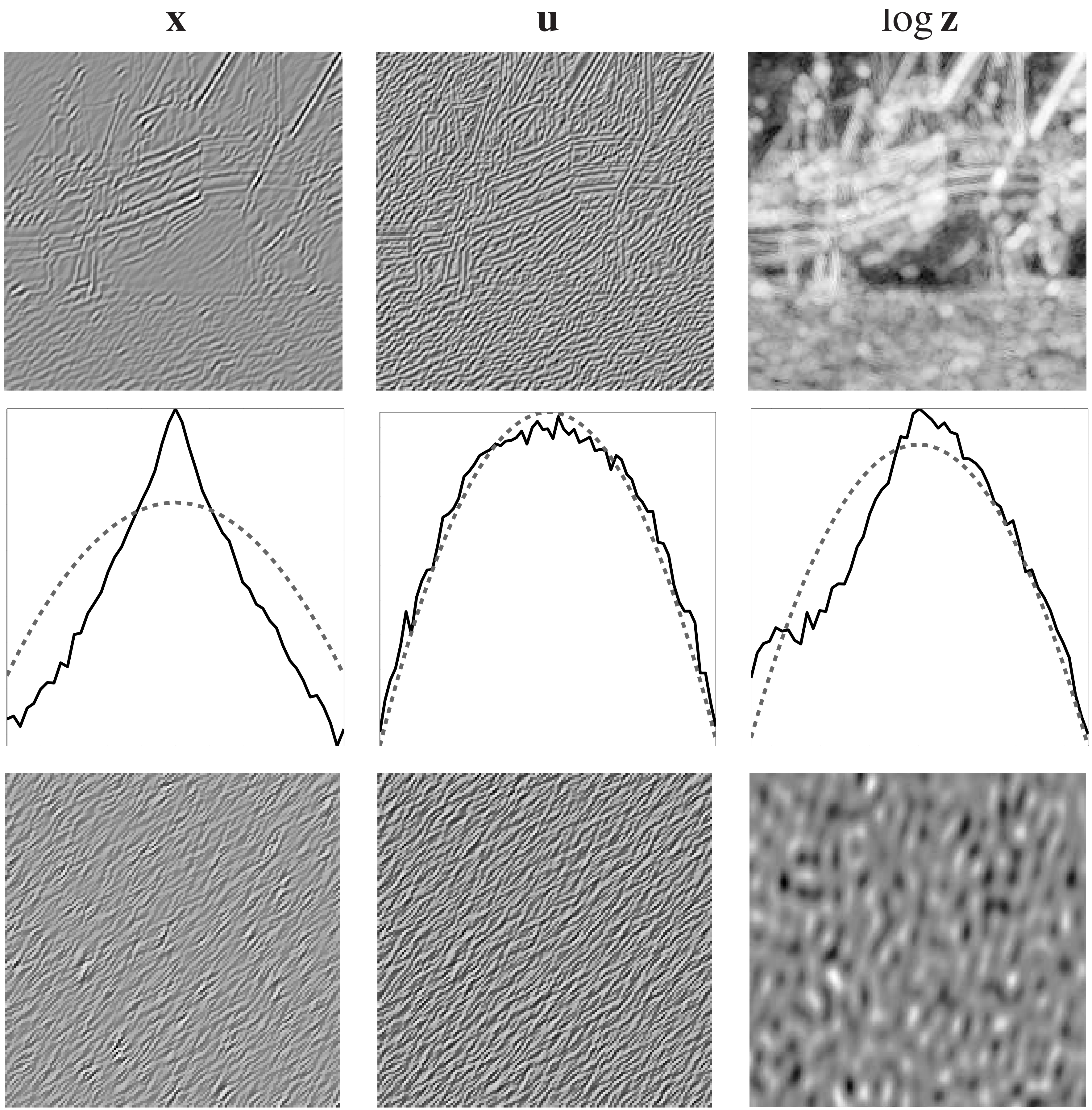
extended models

- independent subspace and topographical ICA [Hoyer & Hyvarinen, 2001,2003; Karklin & Lewicki 2005]
- adaptive covariance structures [Hammond & Simoncelli, 2006; Guerrero-Colon et.al. 2008; Karklin & Lewicki 2009]
- product of *t* experts [Osindero et.al. 2003]
- fields of experts [Roth & Black, 2005]
- tree and fields of GSMs [Wainwright & Simoncelli, 2003; Lyu & Simoncelli, 2008]
- implicit MRF model [Lyu 2009]



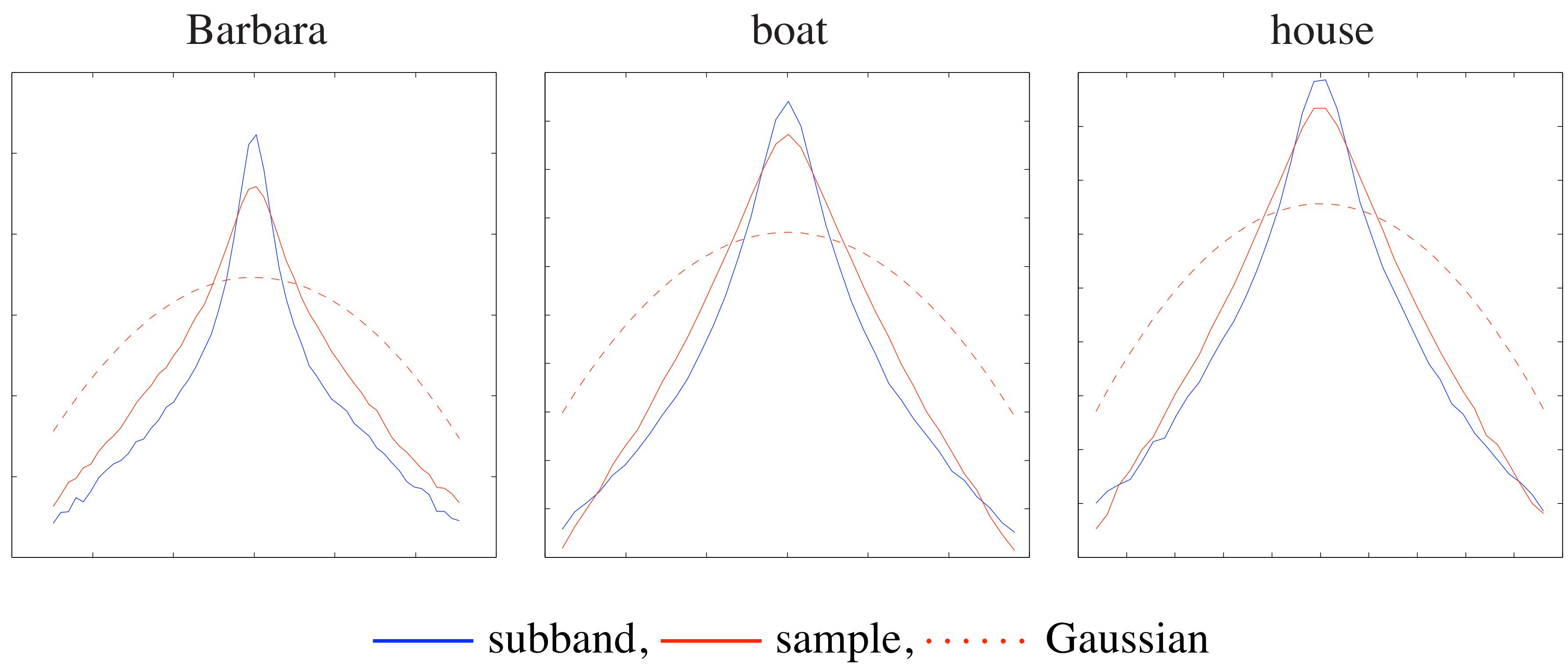
FoGSM: $\vec{x} \stackrel{d}{=} \vec{u} \otimes \sqrt{\vec{z}}$

- $\vec{u}$: zero mean homogeneous Gauss MRF
- $\vec{z}$: exponentiated homogeneous Gauss MRF
- $\vec{x}|\vec{z}$: *inhomogeneous* Gauss MRF
- $\vec{x} \oslash \sqrt{\vec{z}}$: homogeneous Gauss MRF
- marginal distribution is GSM
- generative model: efficient sampling

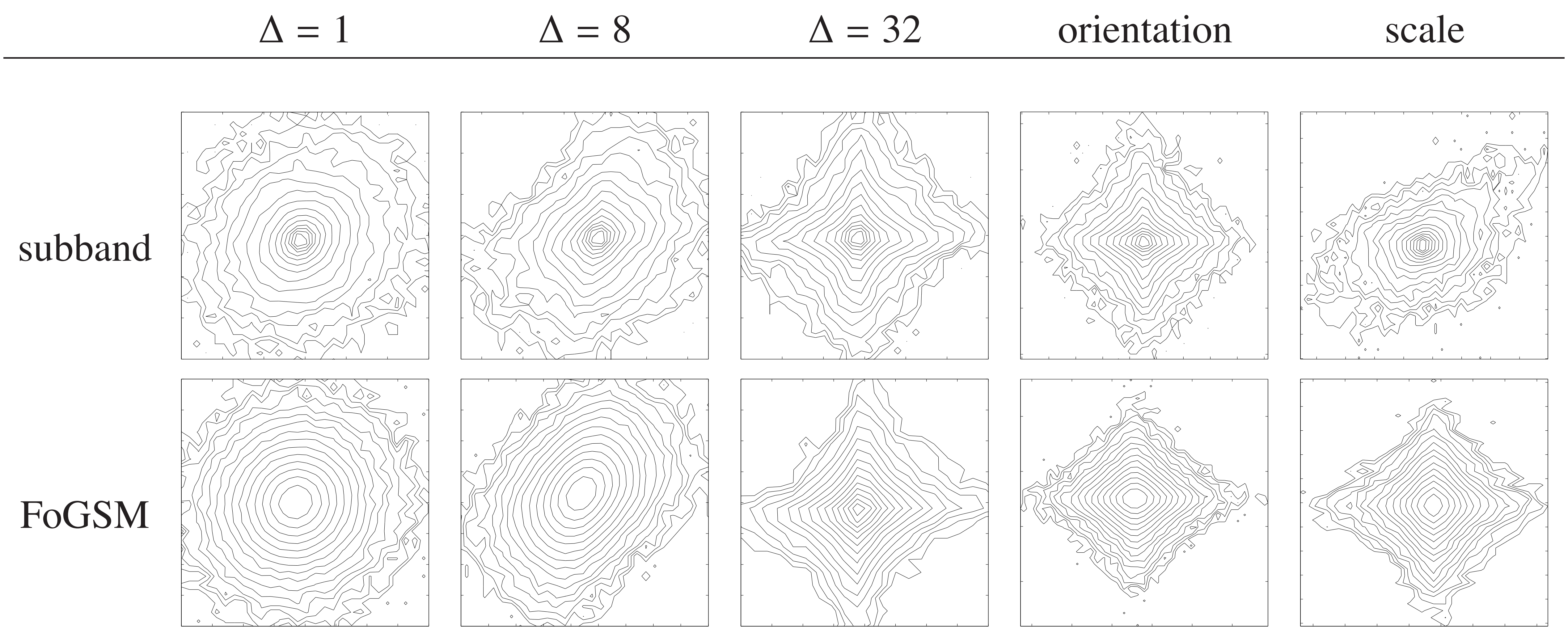


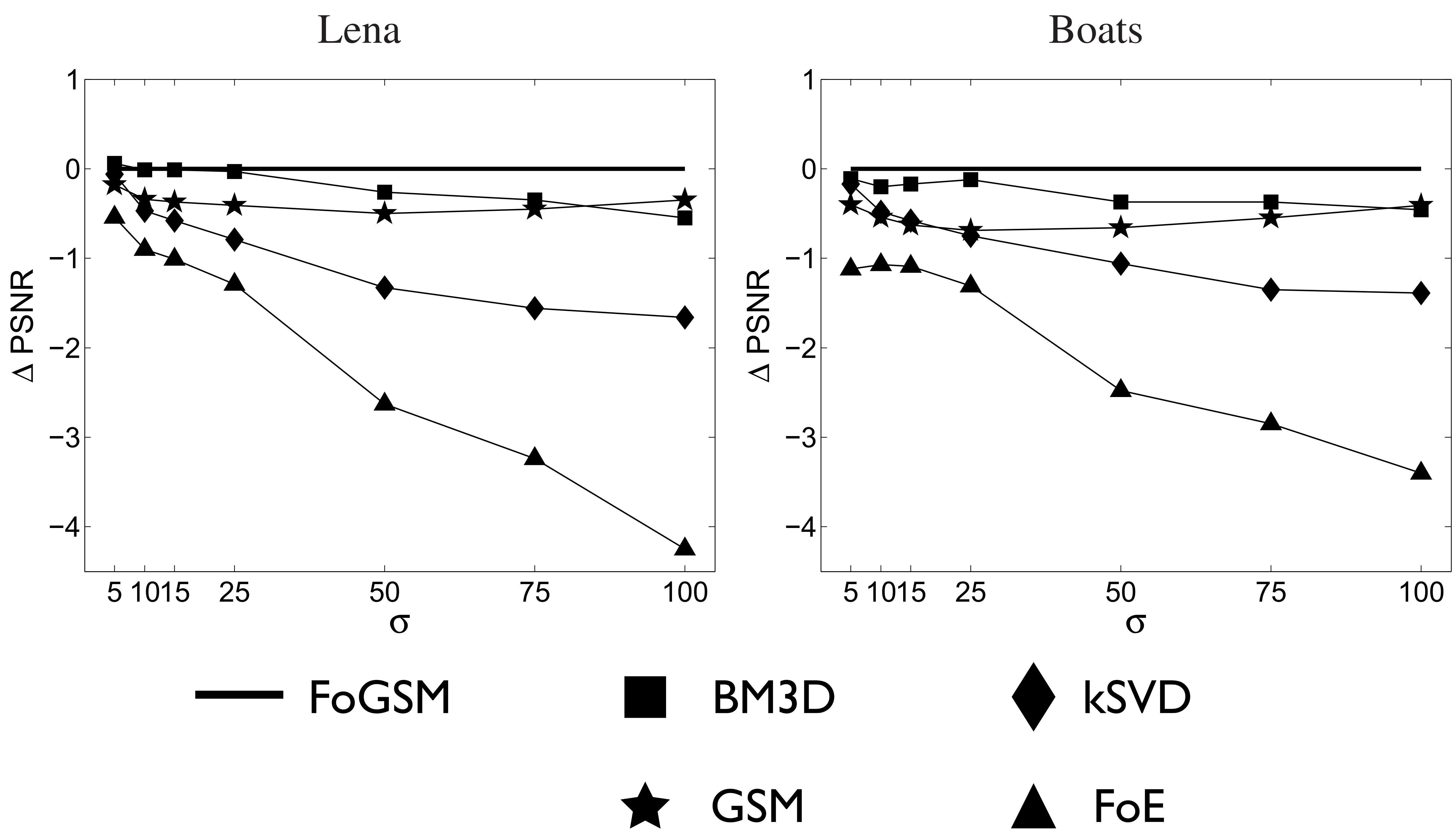
$\mathbf{x}$
$\mathbf{u}$
$\log \mathbf{z}$

marginal



joint

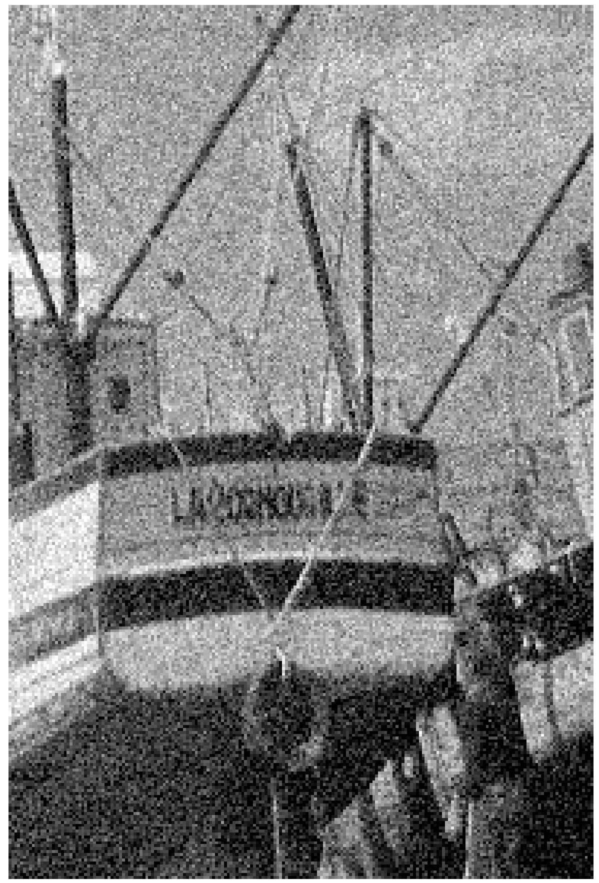




peak-signal-to-noise-ratio (PSNR)

$$20 * \log_{10} \frac{255}{\sqrt{\sum_{i,j} \left(I_{\text{original}}(i,j) - I_{\text{denoised}}(i,j)\right)^2}}$$

original image



noisy image (σ = 25)
(14.15dB)

matlab `wiener2`
(27.19dB)

FoGSM
(30.02dB)

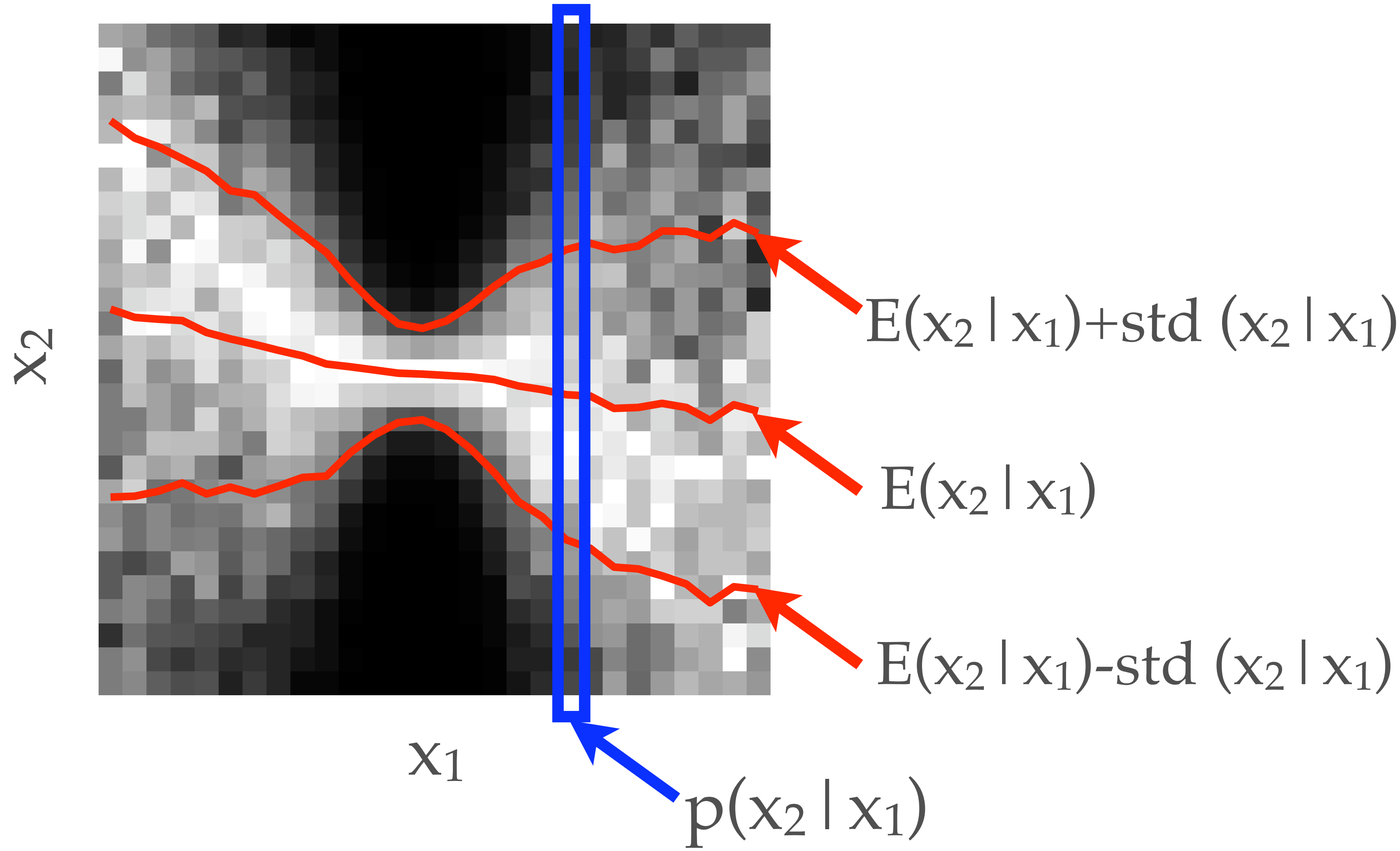
original image

noisy image (σ = 100) (8.13dB)

matlab `wiener2`(29.32dB) (18.38dB)

FoGSM (23.01dB)

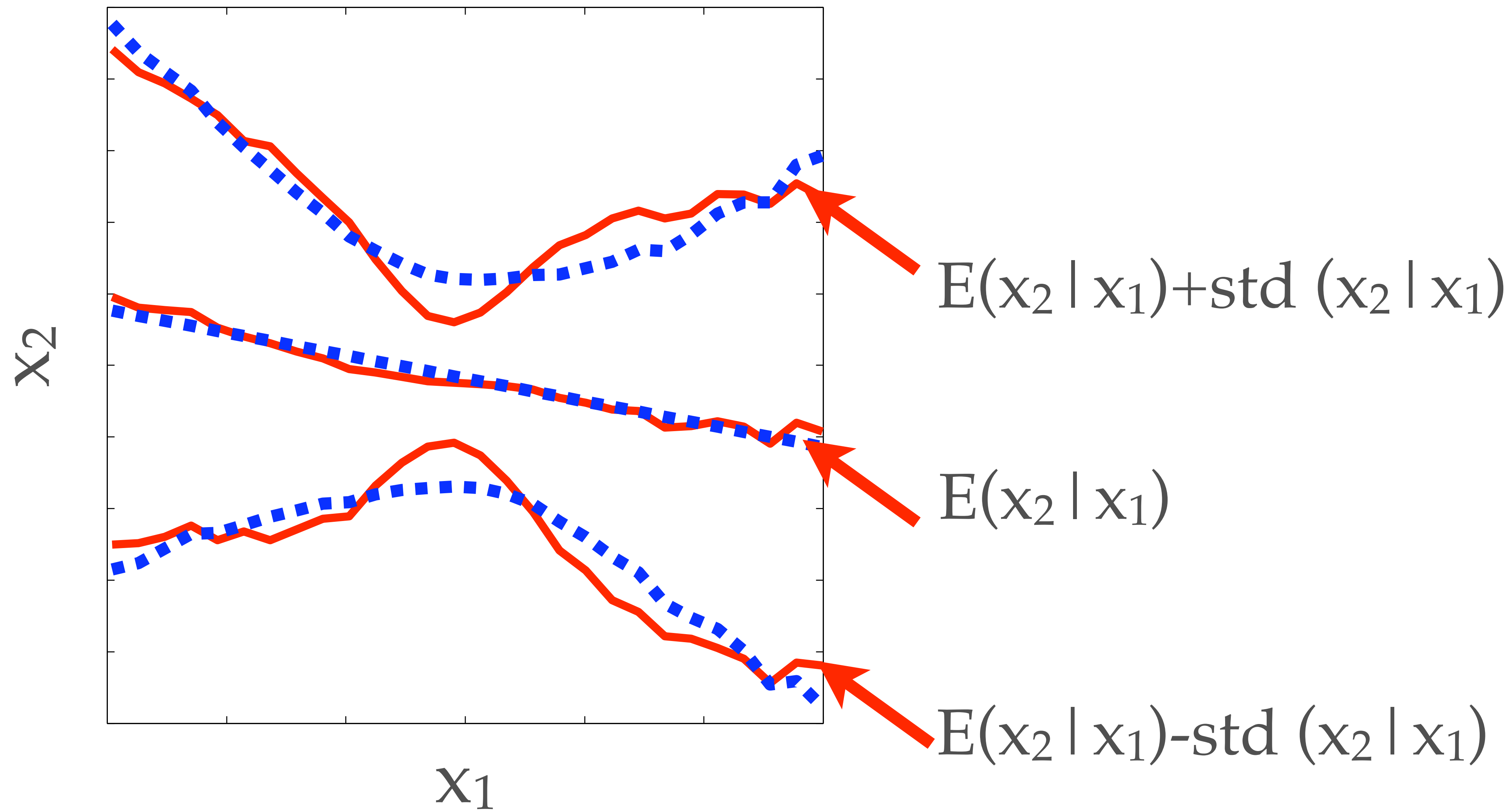
pairwise conditional density



"bow-tie"

[Buccigrossi & Simoncelli, 97]

pairwise conditional density



$$E(x_2|x_1) \approx ax_1$$

$$\mathrm{var}(x_2|x_1) \approx b + cx_1^2$$

conditional density

$$\mu_i = E(x_i | x_{j, j \in \mathcal{N}(i)}) = \sum_{j \in N(i)} a_j x_j$$

$$\sigma_i^2 = \mathrm{var}(x_i | x_{j, j \in \mathcal{N}(i)}) = b + \sum_{j \in N(i)} c_j x_j^2$$

- maxEnt conditional density

$$p(x_i | x_{j, j \in \mathcal{N}(i)}) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right)$$

  - singleton conditionals
  - joint MRF density can be determined by all singletons (Brook's lemma)

implicit MRF

- defined by all singletons
- joint density (and clique potential) is implicit
- learning: maximum pseudo-likelihood

ICM-MAP denoising

$$\operatorname*{argmax}_{\vec{x}} p(\vec{x}|\vec{y}) = \operatorname*{argmax}_{\vec{x}} p(\vec{y}|\vec{x})p(\vec{x}) = \operatorname*{argmax}_{\vec{x}} \log p(\vec{y}|\vec{x}) + \log p(\vec{x})$$

- set initial value for $\vec{x}^{(0)}$, and $t = 1$
- repeat until convergence
  - repeat for all i
    - compute the current estimation for x_i, as

$$\begin{aligned} x_i^{(t)} &= \operatorname*{argmax}_{x_i} \log p(x_1^{(t)}, \cdots, x_{i-1}^{(t)}, \\ &\quad x_i, x_{i+1}^{(t-1)}, \cdots, x_d^{(t-1)} | \vec{y}). \end{aligned}$$

  - $t \leftarrow t + 1$

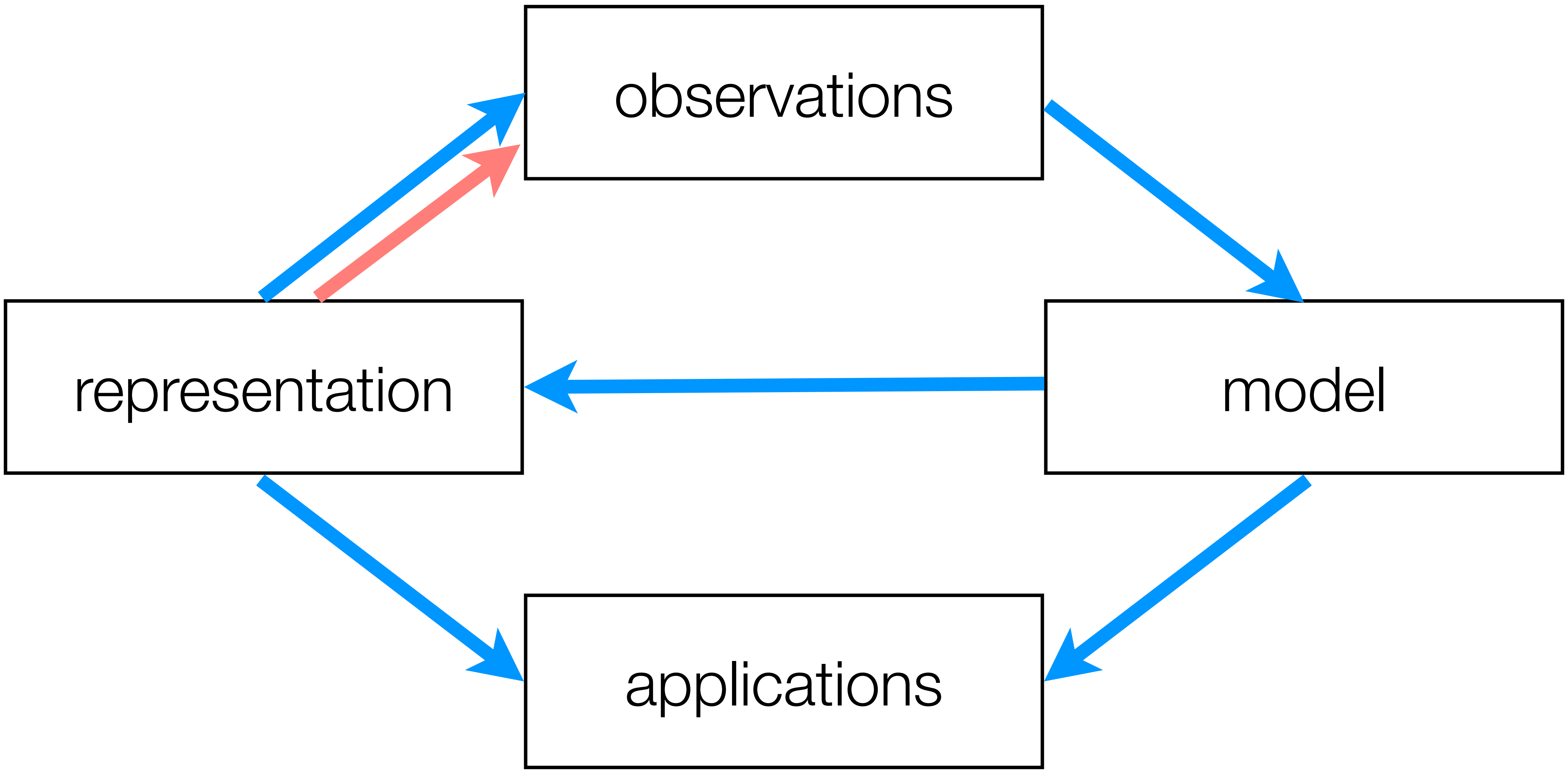
ICM-MAP denoising

$$
\begin{aligned}
& \underset{x_i}{\operatorname{argmax}} \log p(\vec{y}|\vec{x}) + \log p(\vec{x}) \\
= \quad & \underset{x_i}{\operatorname{argmax}} \log p(\vec{y}|\vec{x}) + \log p(x_1, \cdots, x_{i-1}, x_i, x_{i+1}, \cdots, x_n) \\
= \quad & \underset{x_i}{\operatorname{argmax}} \quad \underbrace{\log p(\vec{y}|\vec{x})}_{\text{can be further simplified}} + \underbrace{\log p(x_i|x_{j,j\in N(i)})}_{\text{singleton conditional}} \\
+ \quad & \underbrace{\cancel{\log p(x_{j,j\in N(i)})}}_{\text{constant w.r.t } x_i}.
\end{aligned}
$$

local adaptive and iterative Wiener filtering

$$
x_i = \frac{\sigma_w^2 \sigma_i^2}{\sigma_w^2 + \sigma_i^2} \left(\frac{y_i}{\sigma_w^2} + \frac{\mu_i}{\sigma_i^2} - \sum_{i \neq j} w_{ij}(x_j - y_j) \right).
$$

summary



what need to be done

- inhomogeneous structures
  - structural (edge, contour, etc.)
  - textual (grass, leaves, etc.)
  - smooth (fog, sky, etc.)
- local orientations and relative phases

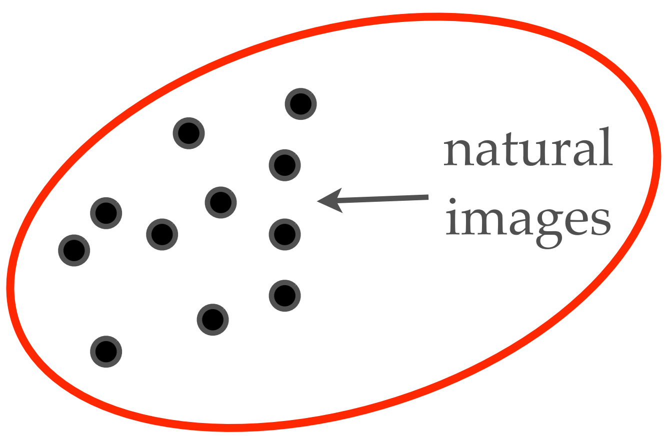
holy grail: comprehensive model & representations to capture all these variations

big question marks

- what are natural images, anyway?

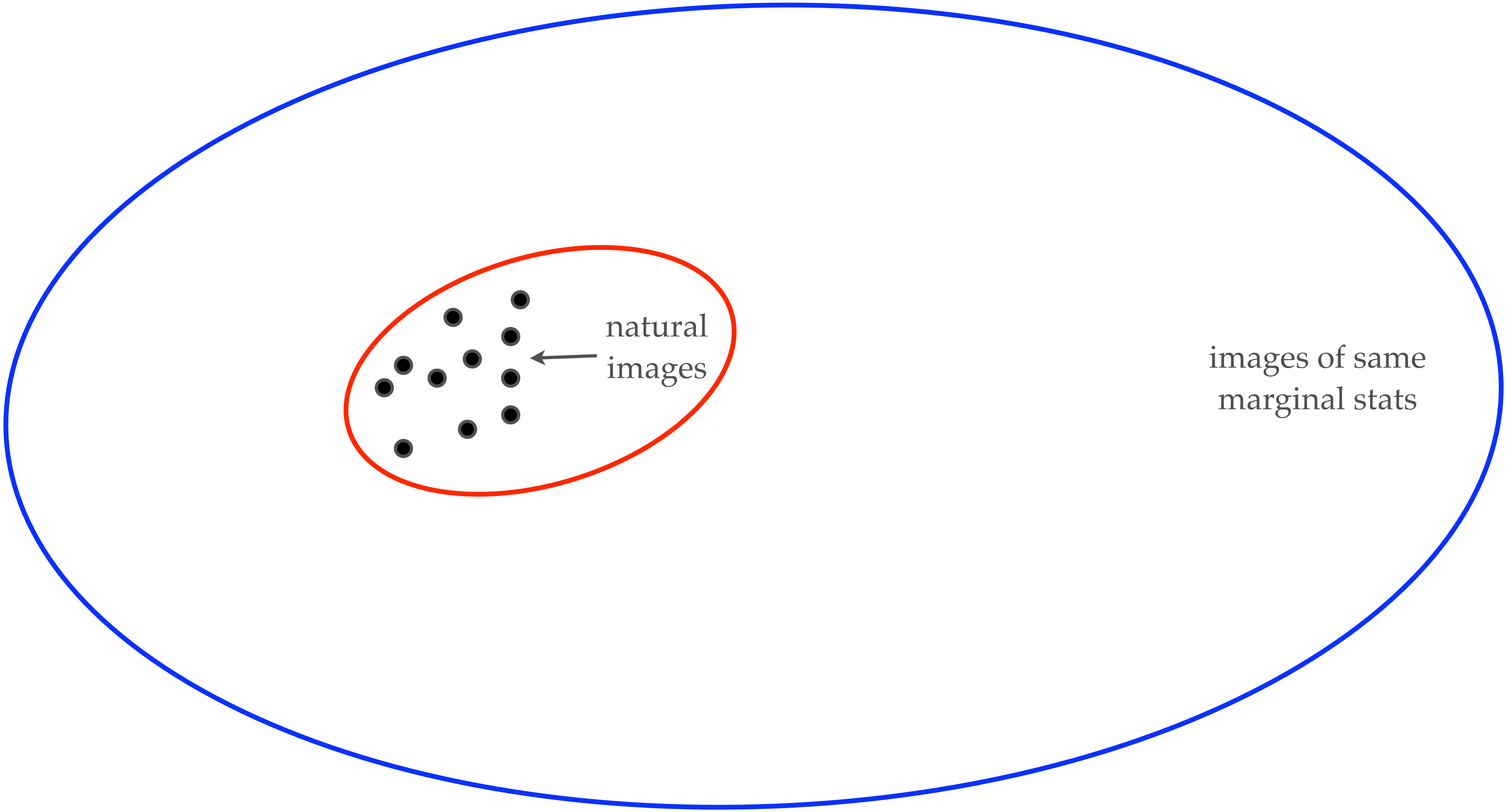
- ironically, white noises are "natural" as they are the result of cosmic radiations
- naturalness is subjective

peeling the onion



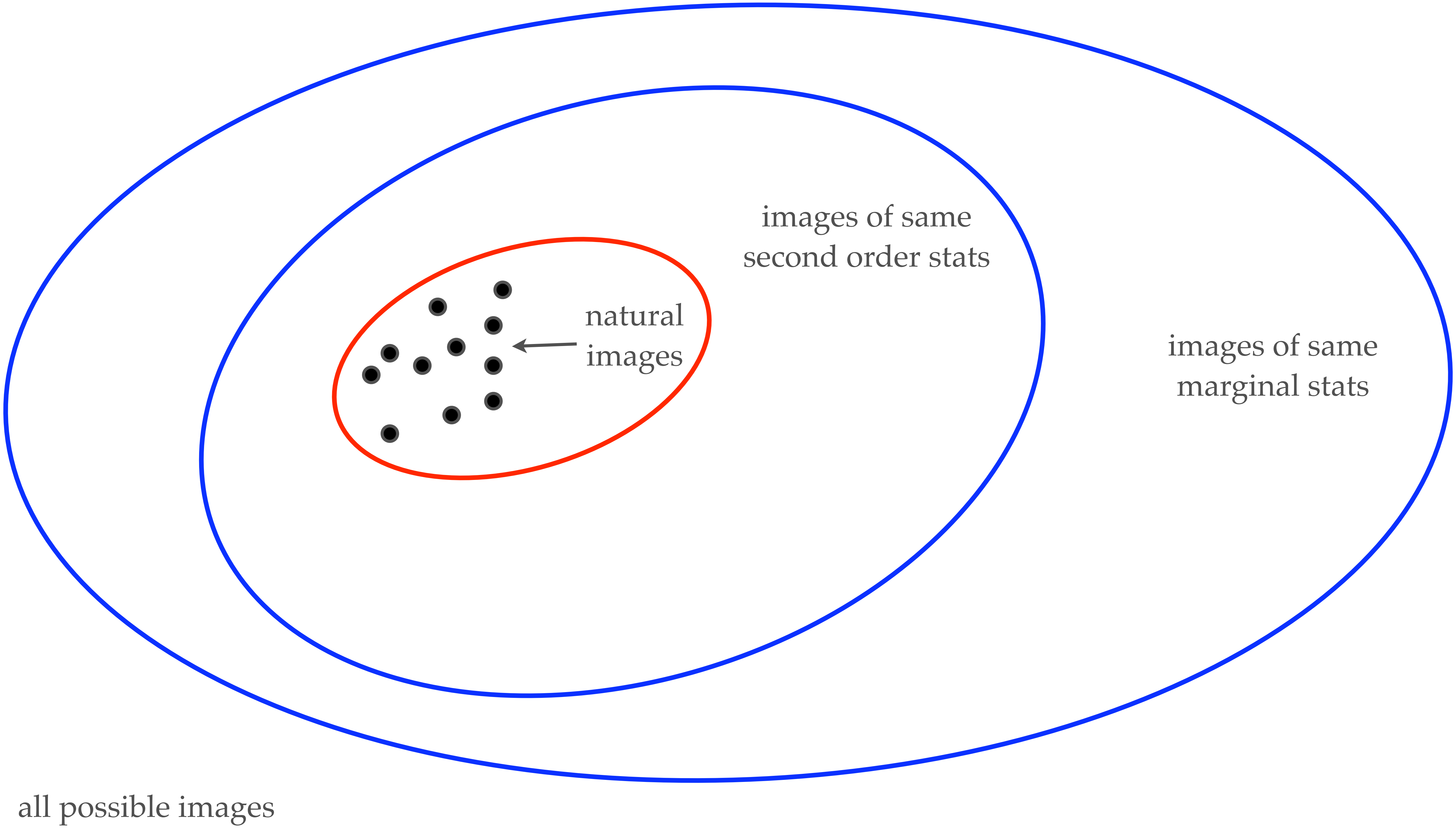
all possible images

peeling the onion

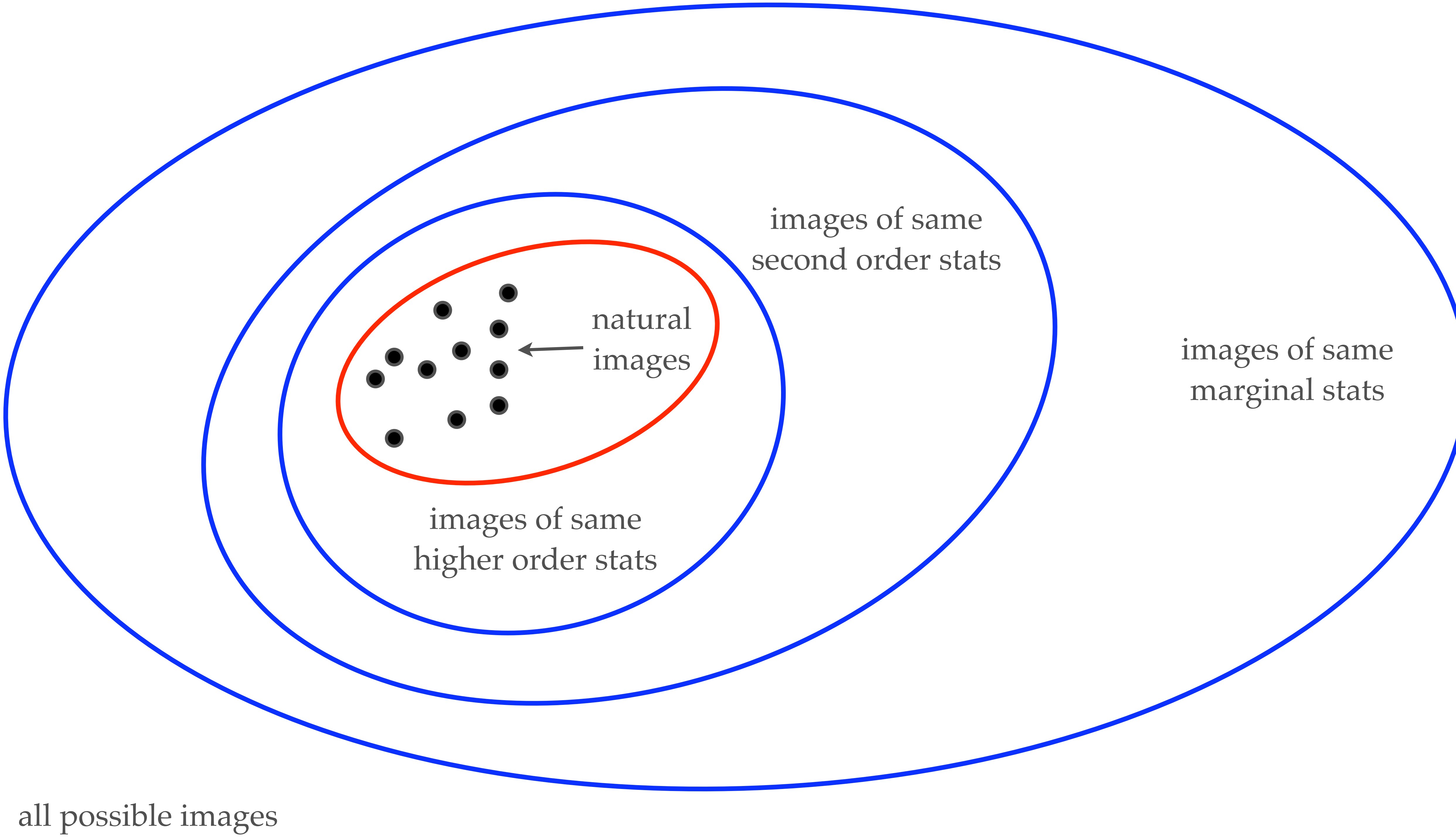


all possible images

peeling the onion



peeling the onion



resources

- D. L. Ruderman. *The statistics of natural images*. Network: Computation in Neural Systems, 5:517–548, 1996.
- E. P. Simoncelli and B. Olshausen. *Natural image statistics and neural representation*. Annual Review of Neuroscience, 24:1193–1216, 2001.
- S.-C. Zhu. *Statistical modeling and conceptualization of visual patterns*. IEEE Trans PAMI, 25(6), 2003
- A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu. *On advances in statistical modeling of natural images*. J. Math. Imaging and Vision, 18(1):17–33, 2003.
- E. P. Simoncelli. *Statistical modeling of photographic images*. In Handbook of Image and Video Processing, 431–441. Academic Press, 2005.
- A. Hyvärinen, J. Hurri, and P. O. Hoyer. *Natural Image Statistics: A probabilistic approach to early computational vision*. Springer, 2009.

thank you