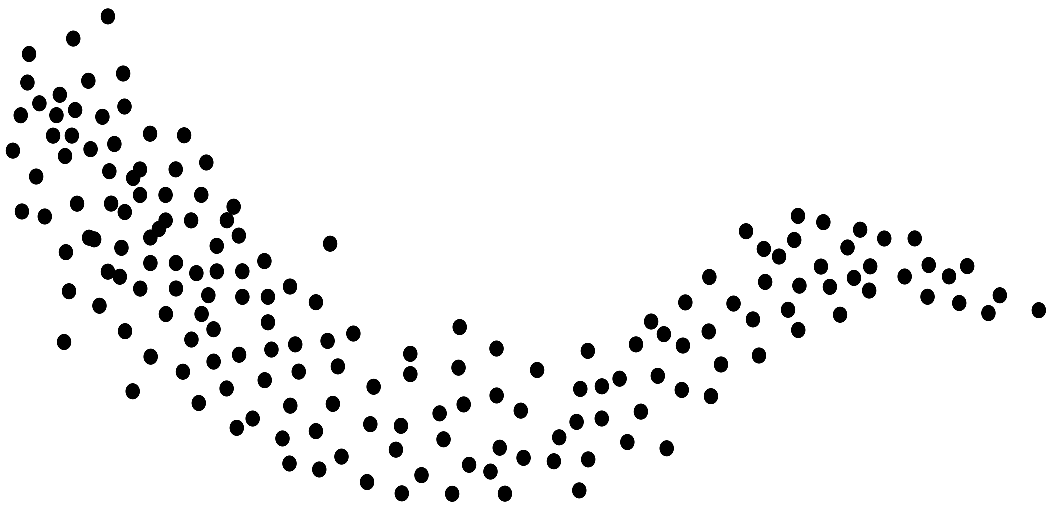
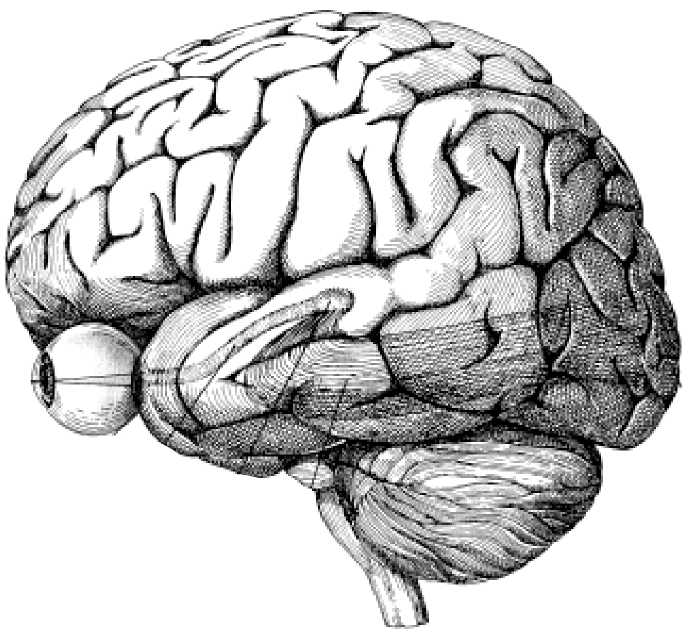
Density Modeling of Images with Generalized Divisive Normalization

Johannes Ballé
Valero Laparra
Eero P. Simoncelli

New York University
Howard Hughes Medical Institute

Why unsupervised learning?





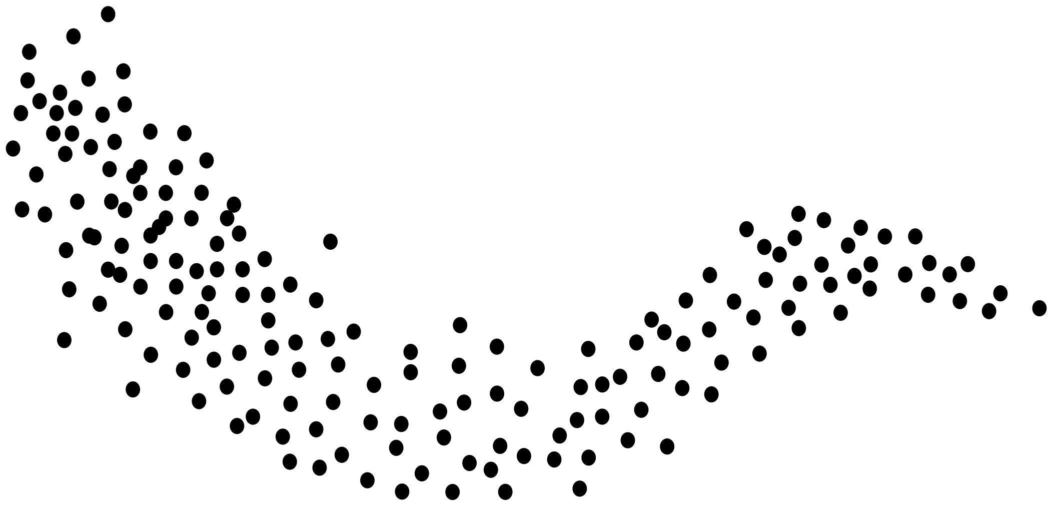
find structure
in unlabeled data

understand sensory
representation

figure: Hubel, 1995

Density estimation (parametric density)

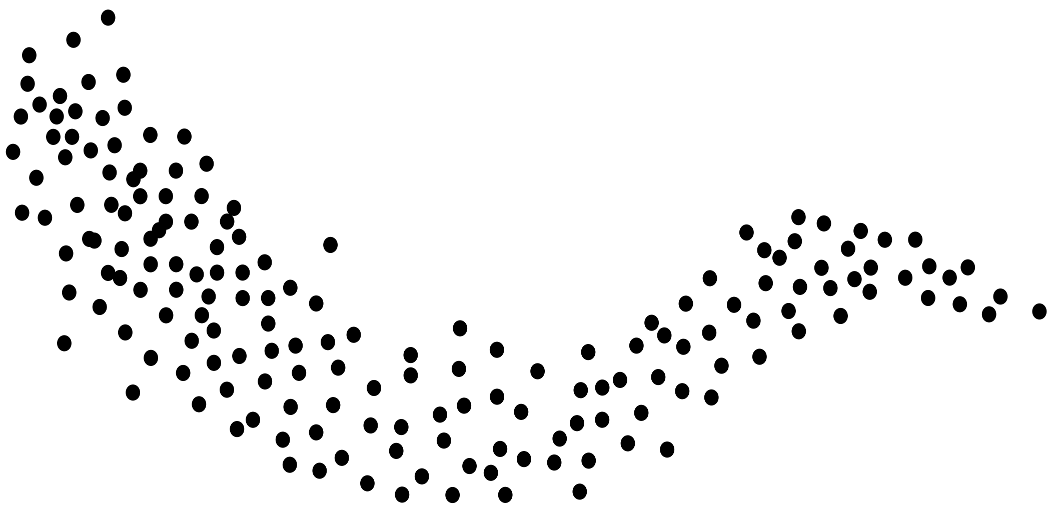
$$p_x(\boldsymbol{x}) = \frac{1}{Z(\boldsymbol{\theta})} \exp\big(-f(\boldsymbol{x}; \boldsymbol{\theta})\big)$$



Density estimation (parametric density)

$$p_x(\boldsymbol{x}) = \frac{1}{Z(\boldsymbol{\theta})} \exp\Big(-f(\boldsymbol{x}; \boldsymbol{\theta})\Big)$$

$$Z(\boldsymbol{\theta}) = \int \exp\Big(-f(\boldsymbol{x}; \boldsymbol{\theta})\Big) \, \mathrm{d}\boldsymbol{x}$$

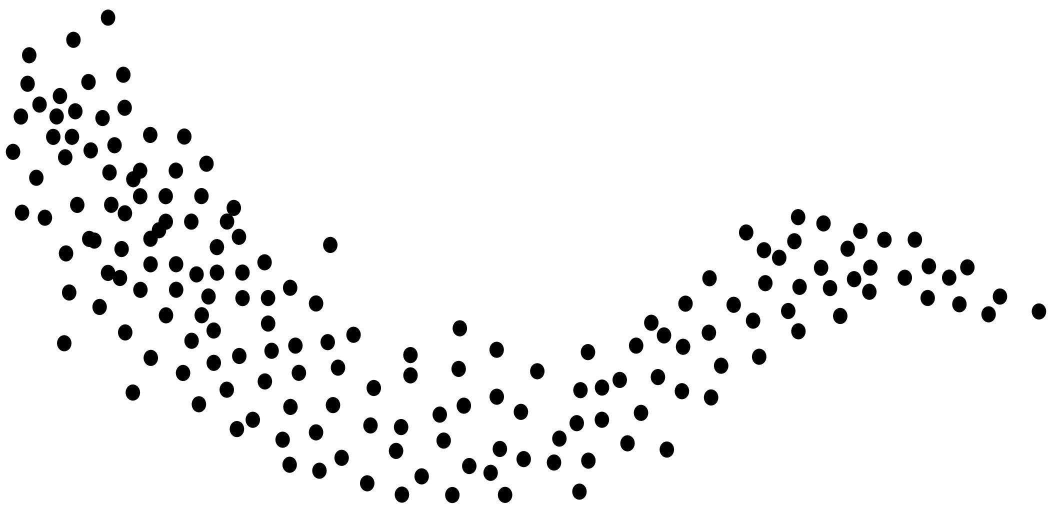


Density estimation (parametric density)

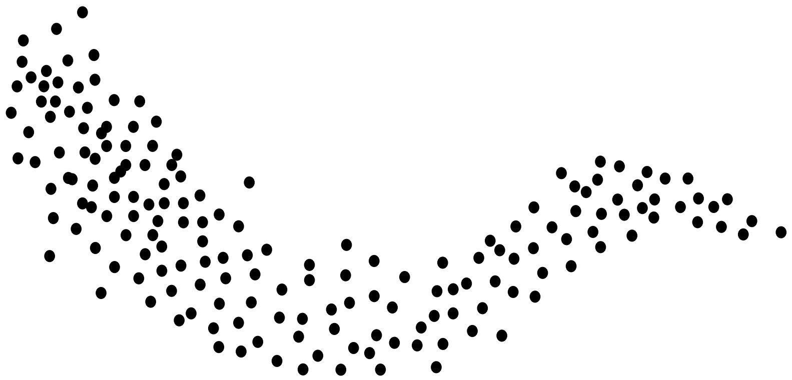
$$p_x(\boldsymbol{x}) = \frac{1}{Z(\boldsymbol{\theta})} \exp\big(-f(\boldsymbol{x}; \boldsymbol{\theta})\big)$$

$$Z(\boldsymbol{\theta}) = \int \exp\big(-f(\boldsymbol{x}; \boldsymbol{\theta})\big)\, \mathrm{d}\boldsymbol{x}$$

tractable?



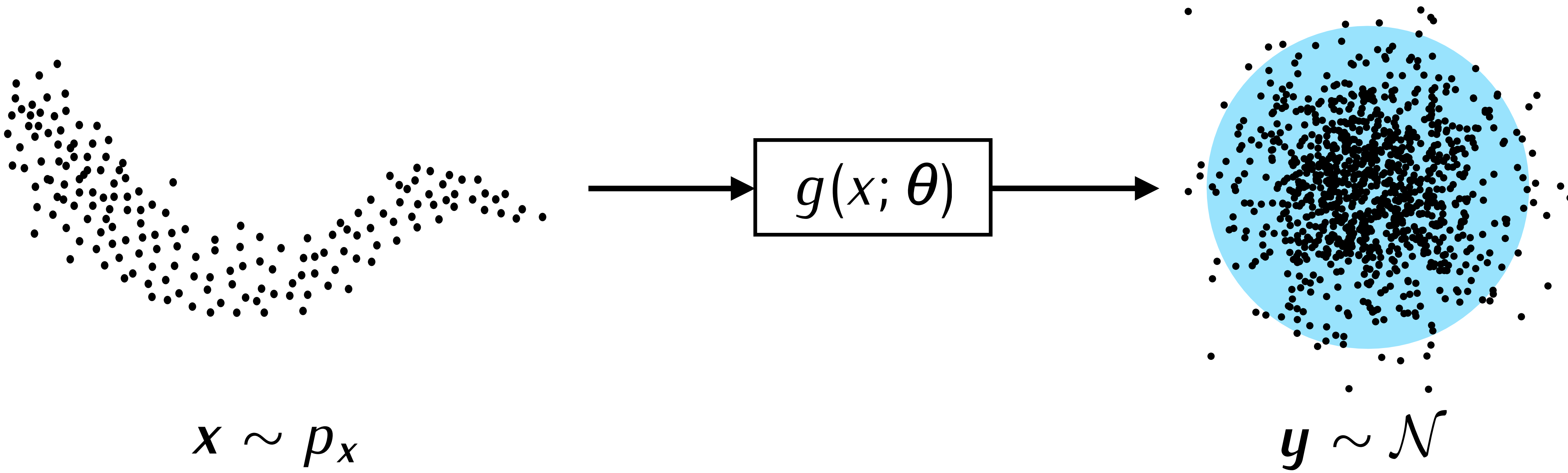
Density estimation (parametric transformation)



$\boldsymbol{x} \sim p_x$

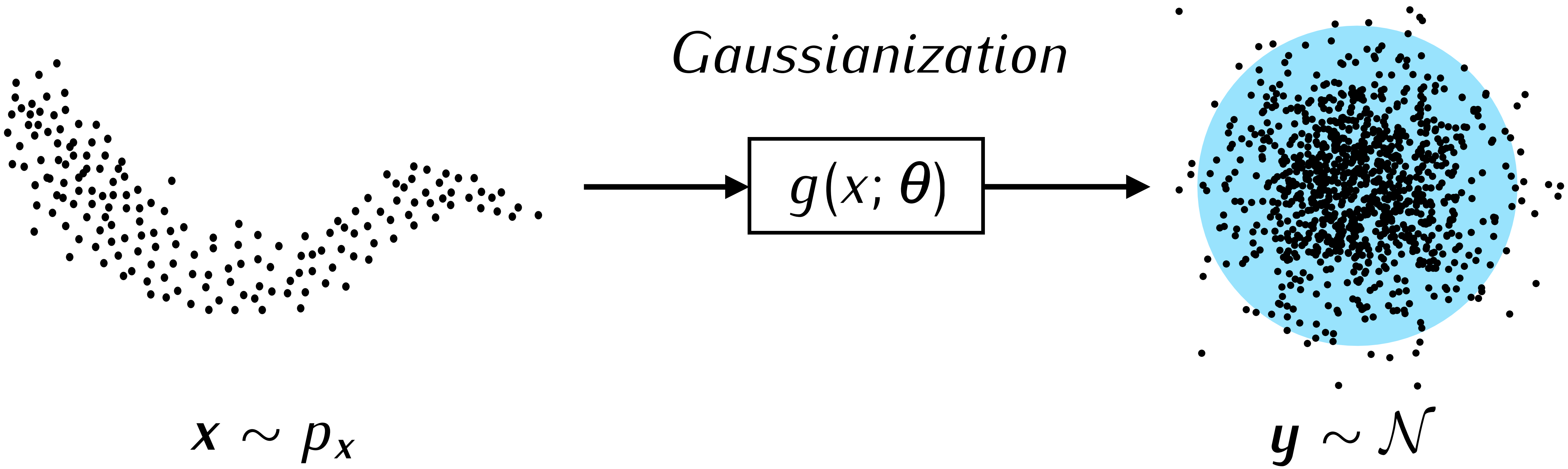
Friedman, 1984
Chen & Gopinath, 2001
Lyu & Simoncelli, 2009
Laparra et al., 2010
Dinh et al., 2015

Density estimation (parametric transformation)



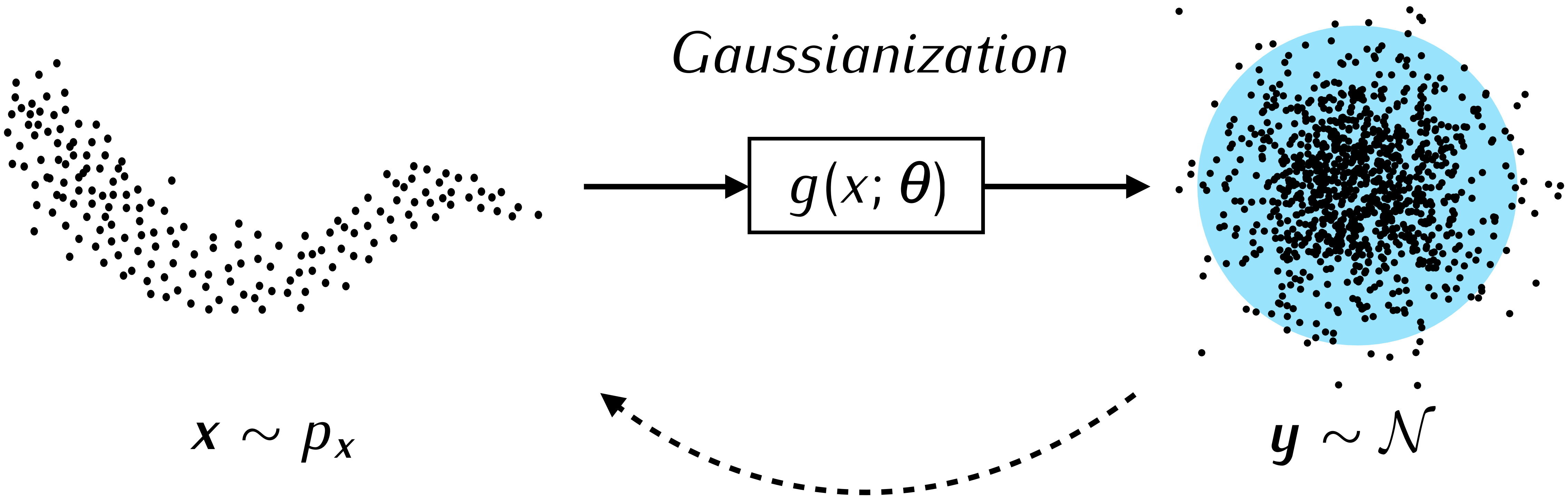
Friedman, 1984
Chen & Gopinath, 2001
Lyu & Simoncelli, 2009
Laparra et al., 2010
Dinh et al., 2015

Density estimation (parametric transformation)



Friedman, 1984
Chen & Gopinath, 2001
Lyu & Simoncelli, 2009
Laparra et al., 2010
Dinh et al., 2015

Density estimation (parametric transformation)

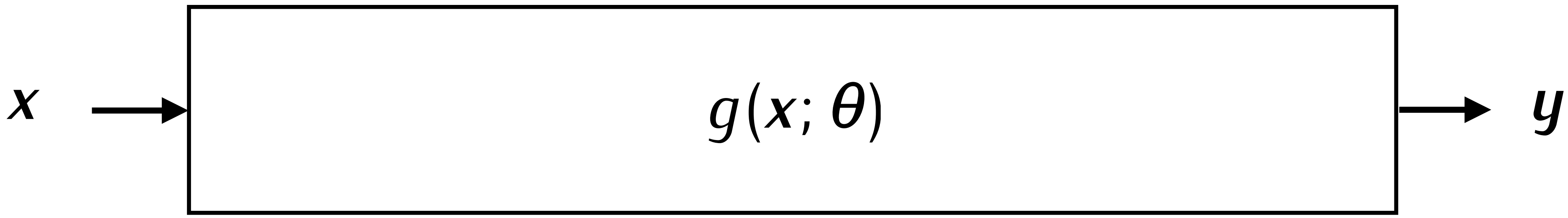


"inferred" density:

$$p_x(\boldsymbol{x}) = \left| \frac{\partial g(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{x}} \right| \mathcal{N}\big(g(\boldsymbol{x};\boldsymbol{\theta})\big)$$

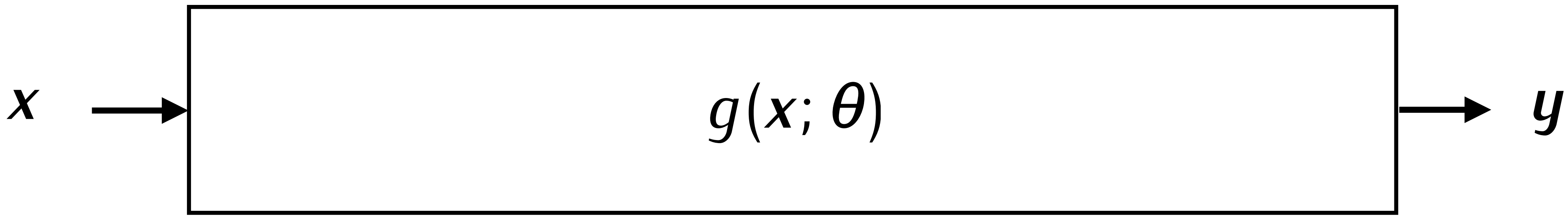
Friedman, 1984
Chen & Gopinath, 2001
Lyu & Simoncelli, 2009
Laparra et al., 2010
Dinh et al., 2015

Parameter estimation



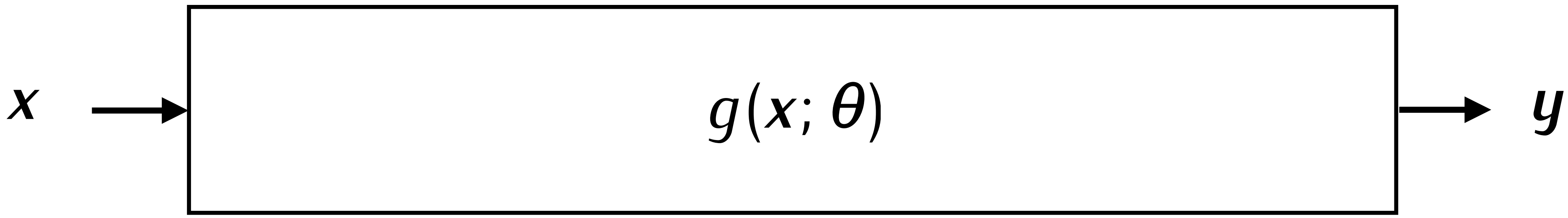
$$p_x(\boldsymbol{x}) = \left|\frac{\partial g(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{x}}\right| \mathcal{N}\big(g(\boldsymbol{x};\boldsymbol{\theta})\big)$$

Parameter estimation



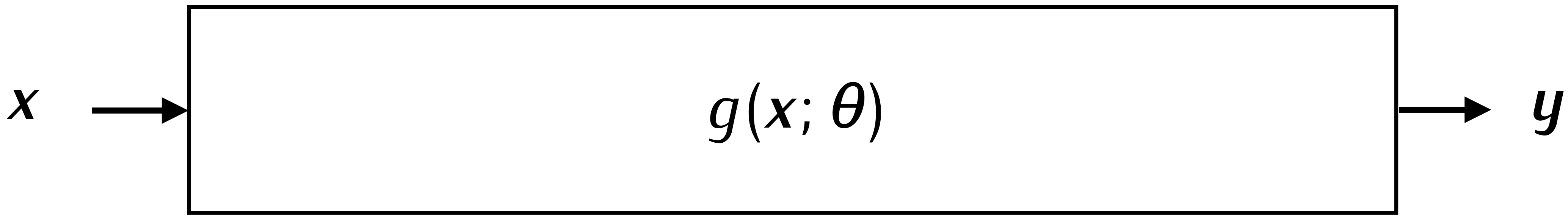
$$-\log p_{\boldsymbol{x}}(\boldsymbol{x}) = -\log\left|\frac{\partial g(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{x}}\right| - \log\mathcal{N}\big(g(\boldsymbol{x};\boldsymbol{\theta})\big)$$

Parameter estimation



$$-\log p_{x}(\boldsymbol{x}) = -\log\left|\frac{\partial g(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{x}}\right| - \frac{1}{2}\left\|g(\boldsymbol{x};\boldsymbol{\theta})\right\|_2^2 + C$$

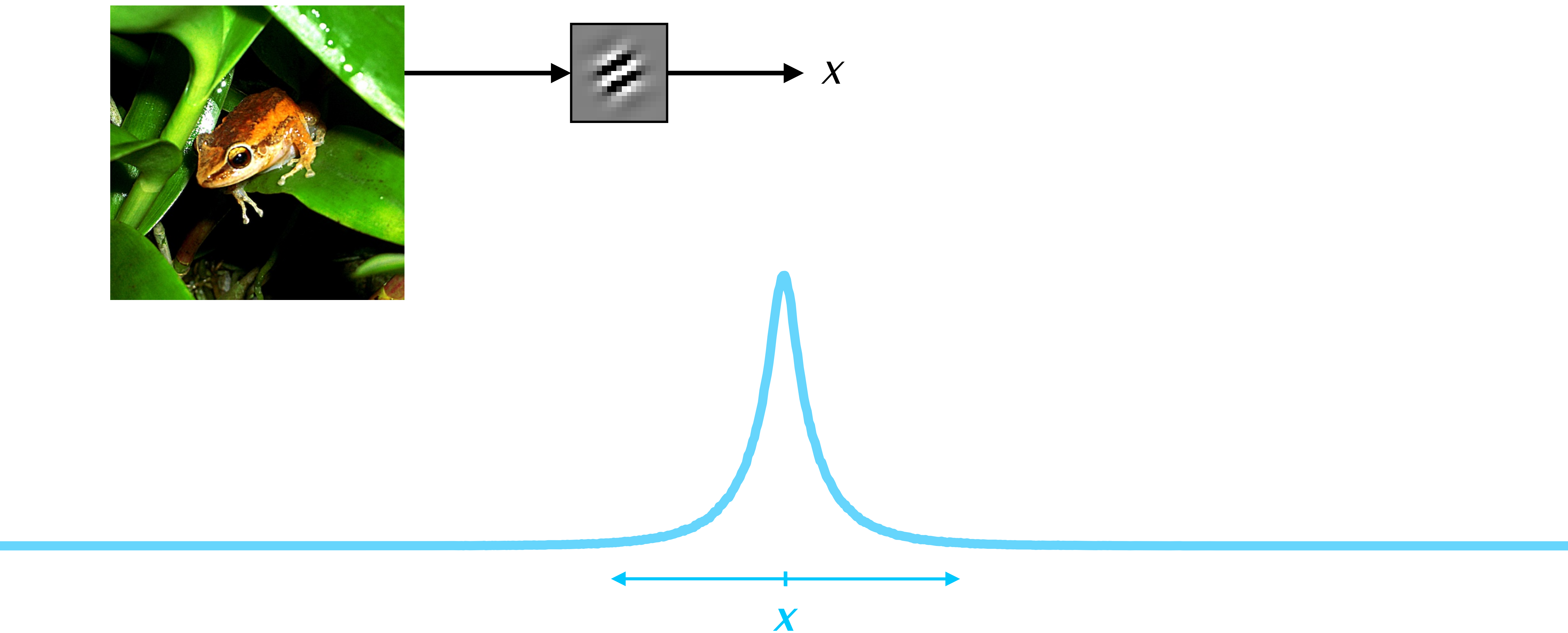
Parameter estimation



$$-\log p_{\boldsymbol{x}}(\boldsymbol{x}) = -\log \left| \frac{\partial g(\boldsymbol{x}; \boldsymbol{\theta})}{\partial \boldsymbol{x}} \right| - \frac{1}{2} \left\| g(\boldsymbol{x}; \boldsymbol{\theta}) \right\|_2^2 + C$$

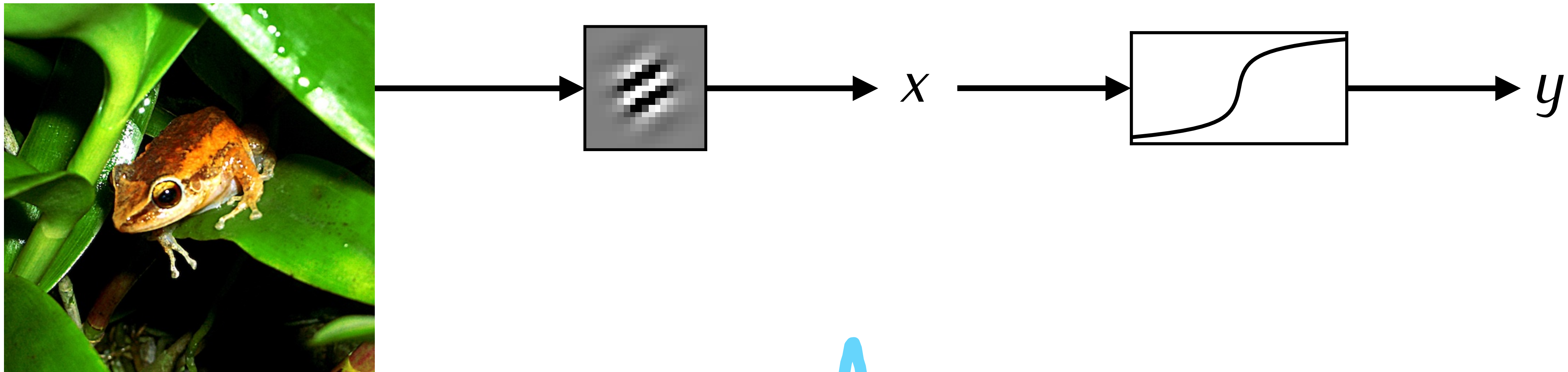
minimize wrt. $\boldsymbol{\theta}$ using stochastic gradient descent

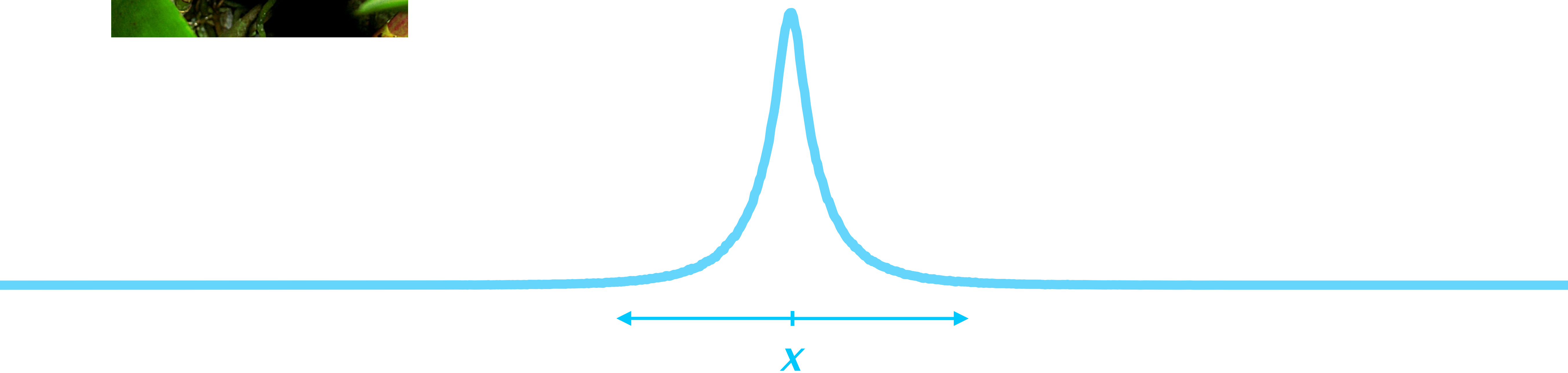
Marginal distribution of linear filter responses



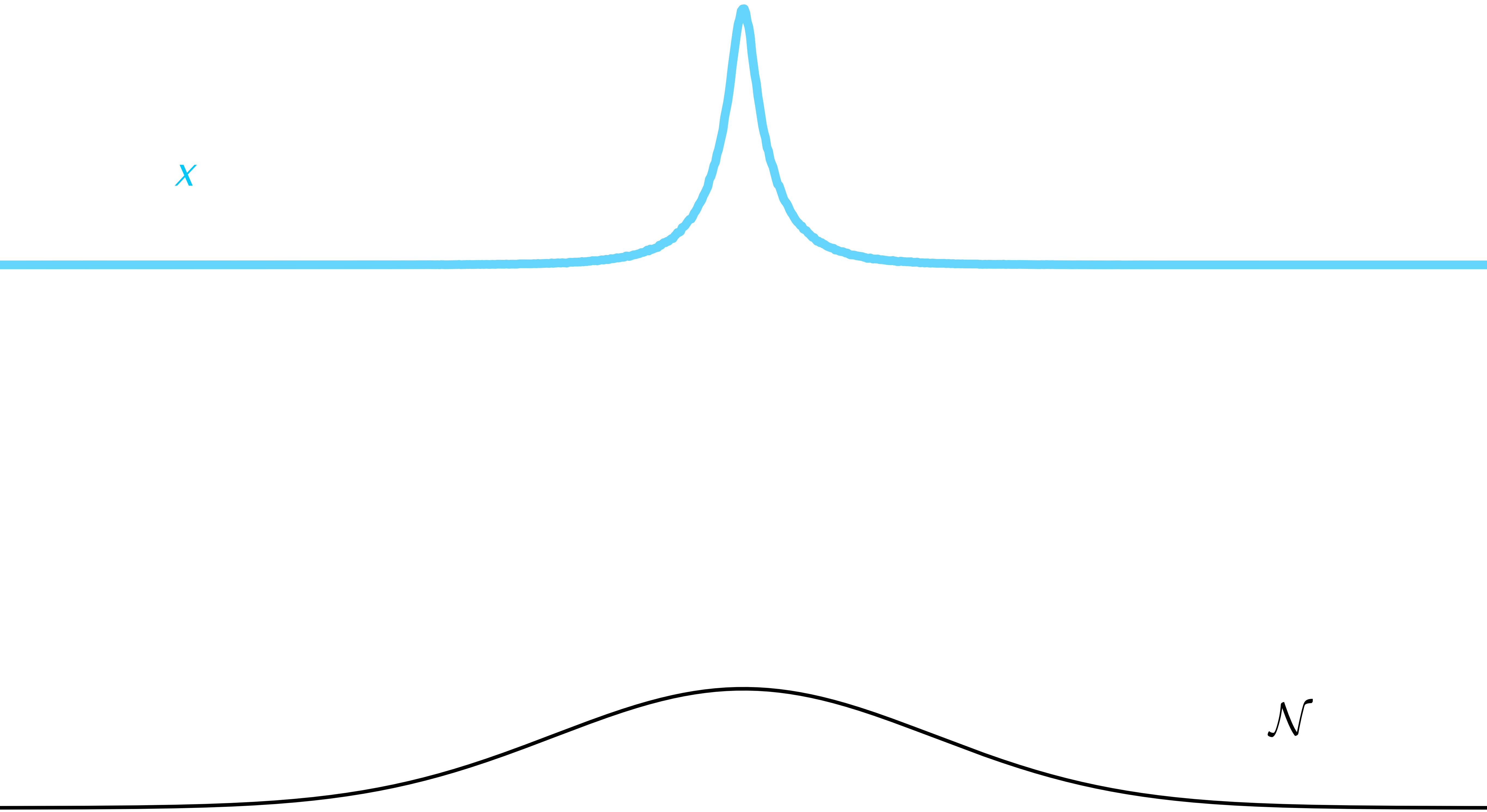
Burt & Adelson, 1981
Field, 1987
Mallat, 1989

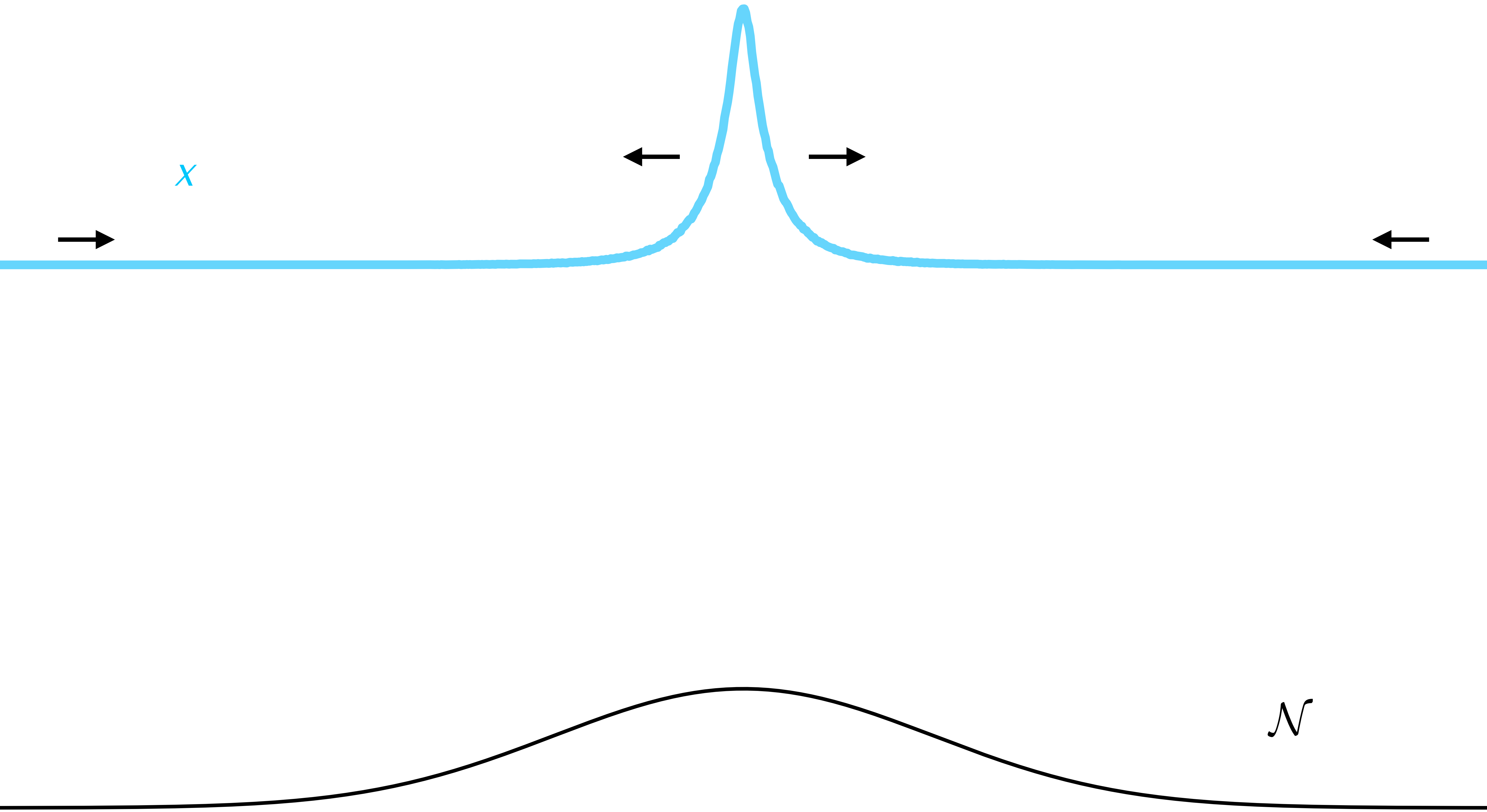
Marginal distribution of linear filter responses



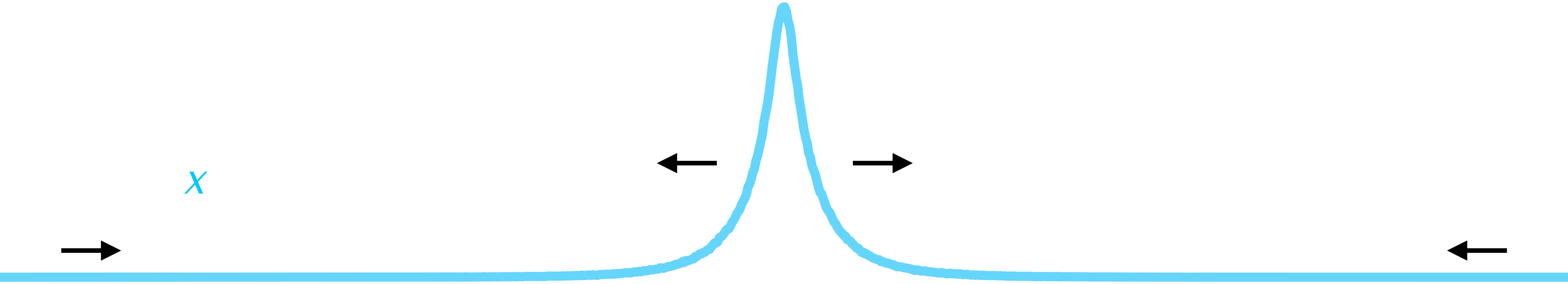


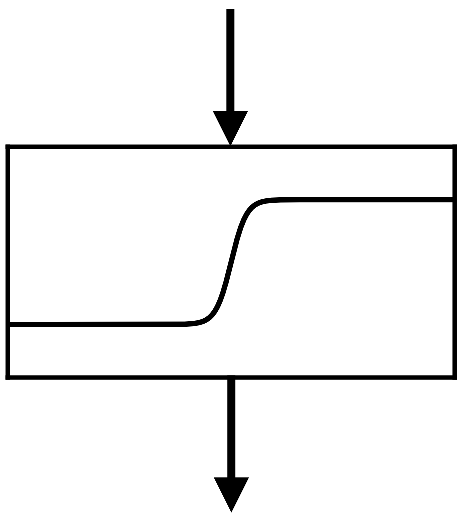
Burt & Adelson, 1981
Field, 1987
Mallat, 1989


x
$\mathcal{N}$

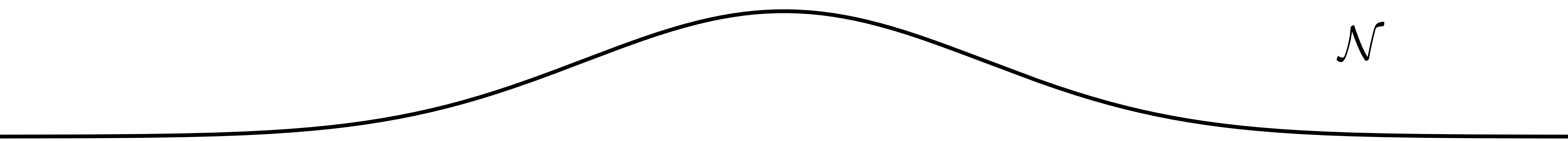


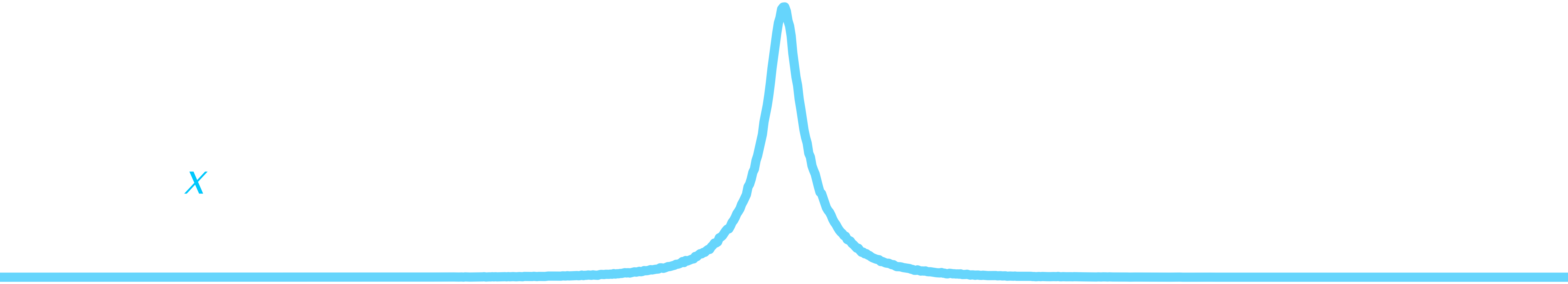
x
𝒩

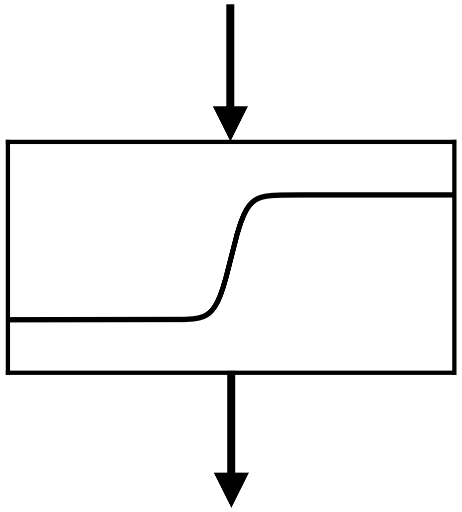




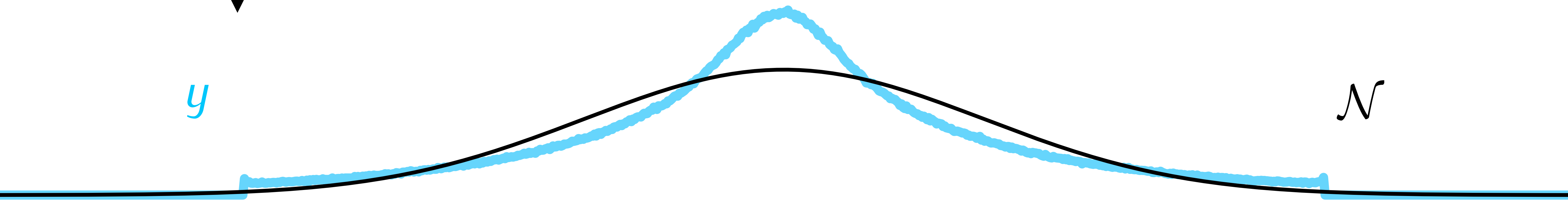
$$y = \frac{c}{1 + \exp(ax + b)}$$

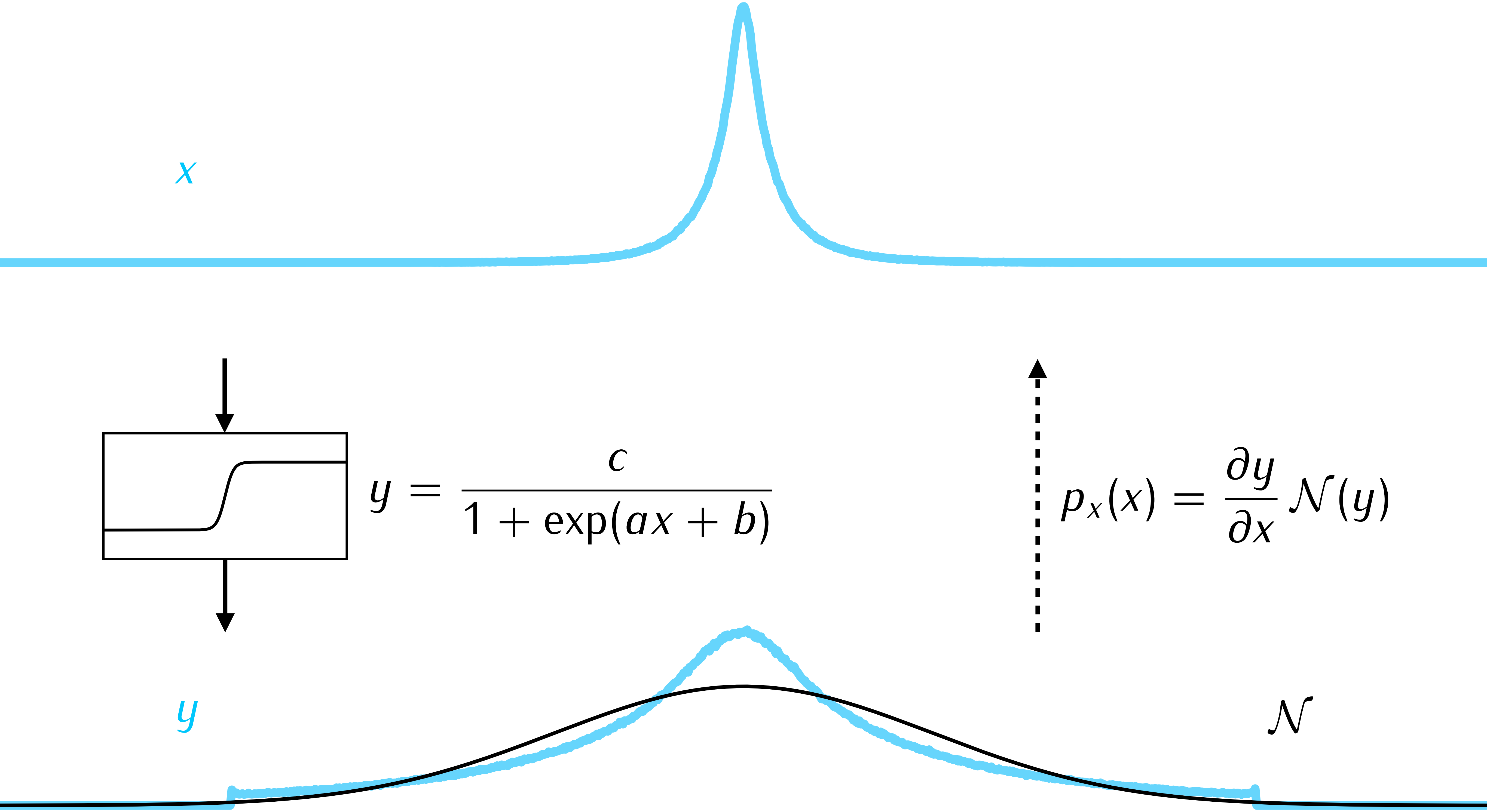


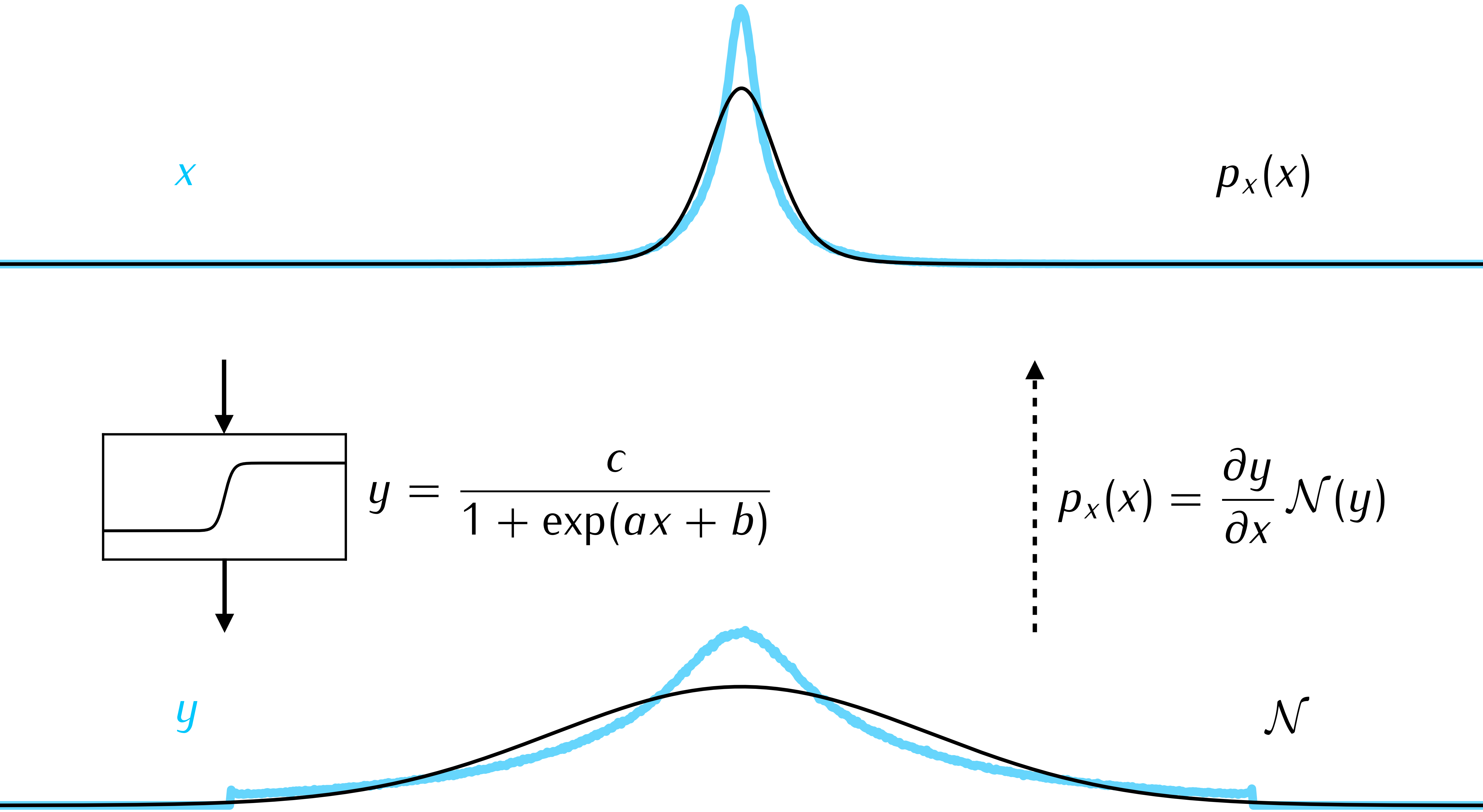


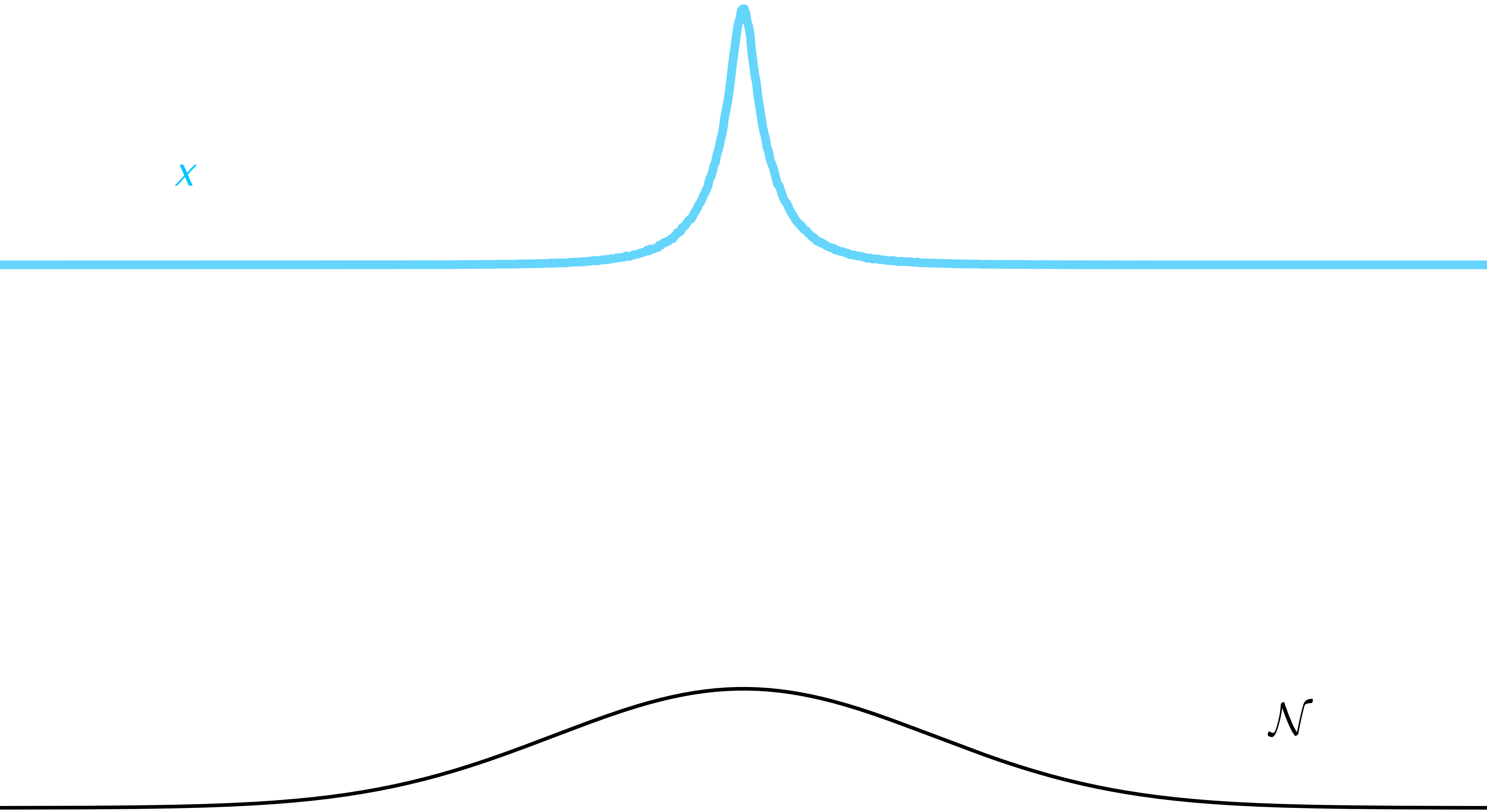


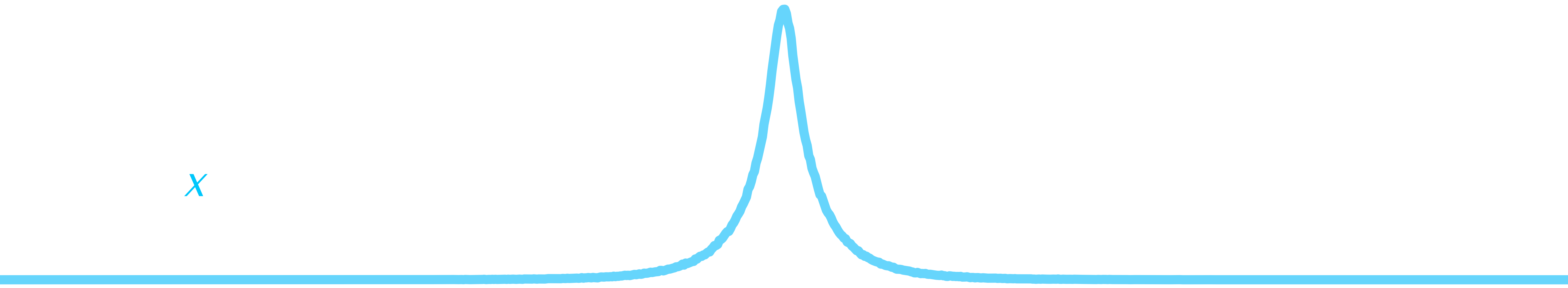
$$y = \frac{c}{1 + \exp(ax + b)}$$

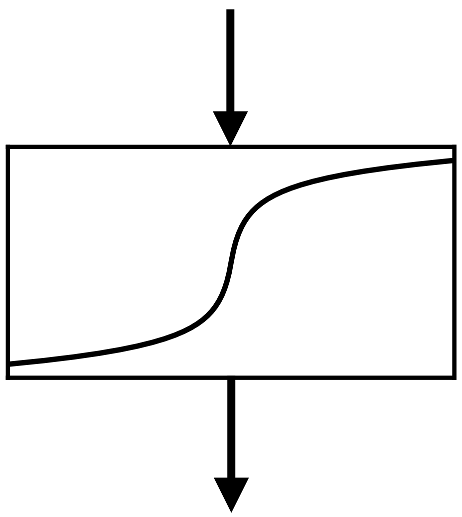



x
$y = \frac{c}{1 + \exp(ax + b)}$
$p_x(x) = \frac{\partial y}{\partial x} \mathcal{N}(y)$
y
$\mathcal{N}$

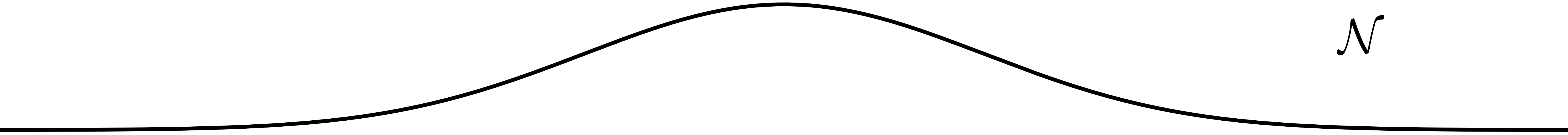

x
$p_x(x)$
$y = \frac{c}{1 + \exp(ax + b)}$
$p_x(x) = \frac{\partial y}{\partial x} \mathcal{N}(y)$
y
$\mathcal{N}$

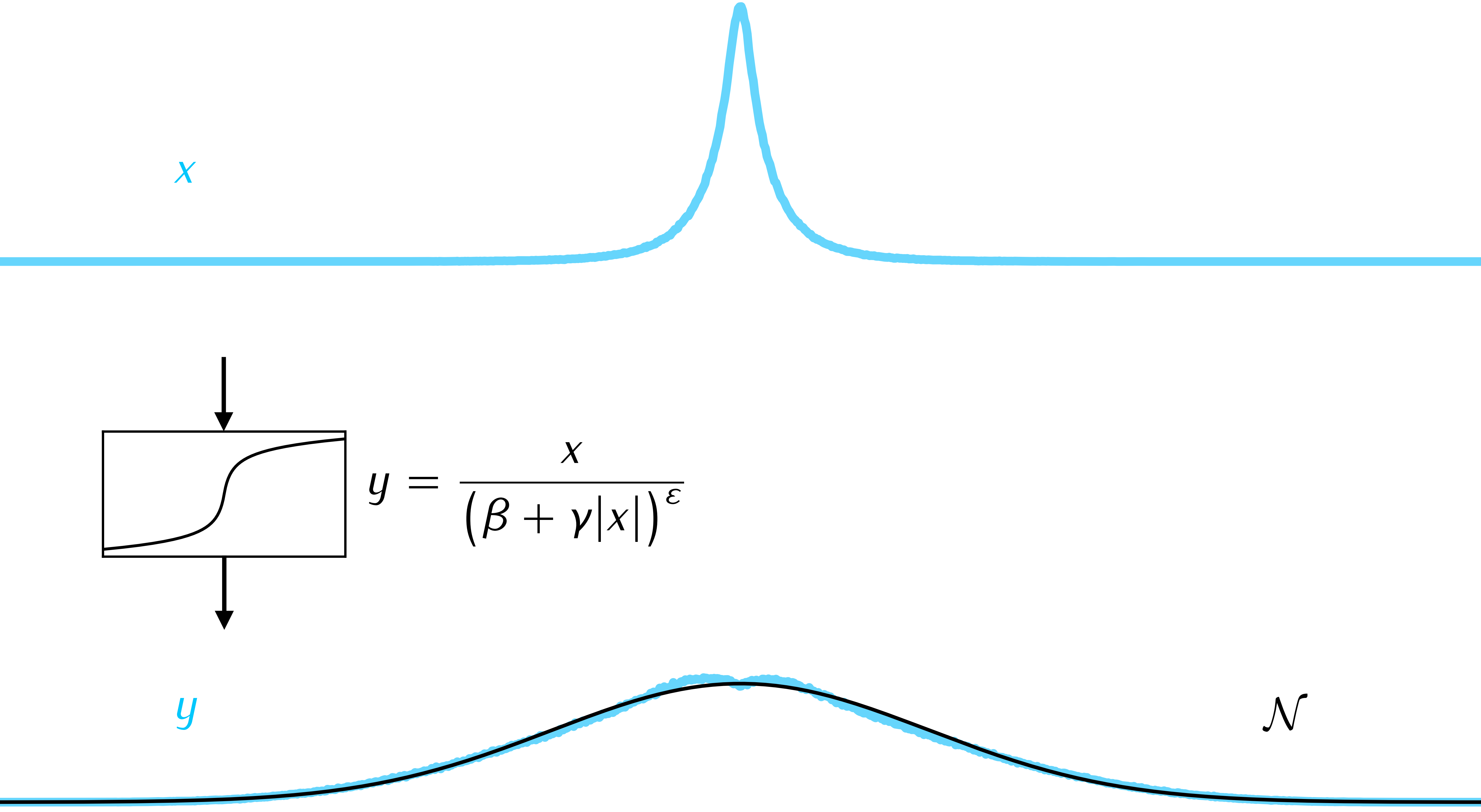

x
$\mathcal{N}$



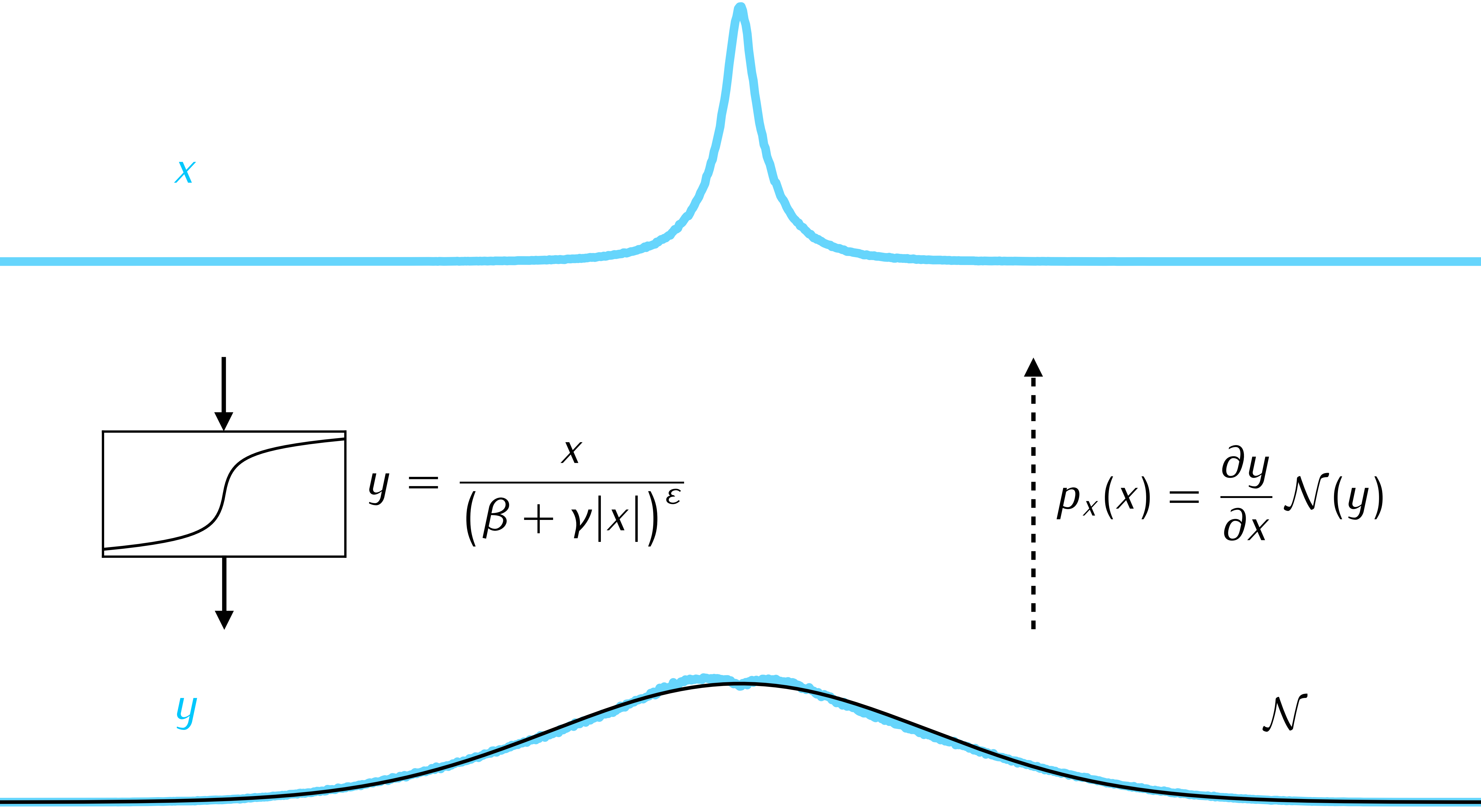


$$y = \frac{x}{\left(\beta + \gamma|x|\right)^{\varepsilon}}$$



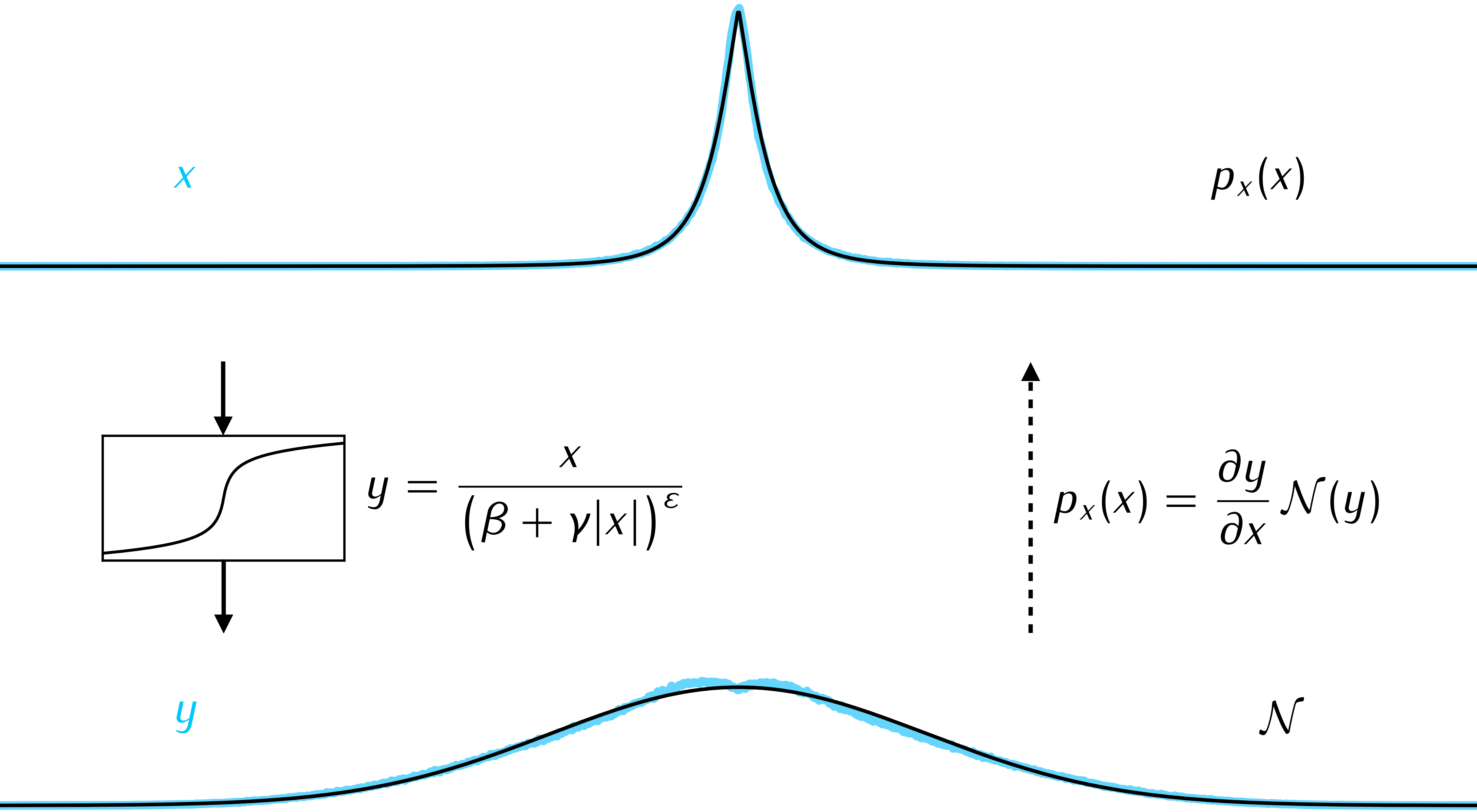


$$y = \frac{x}{\left(\beta + \gamma |x|\right)^{\varepsilon}}$$

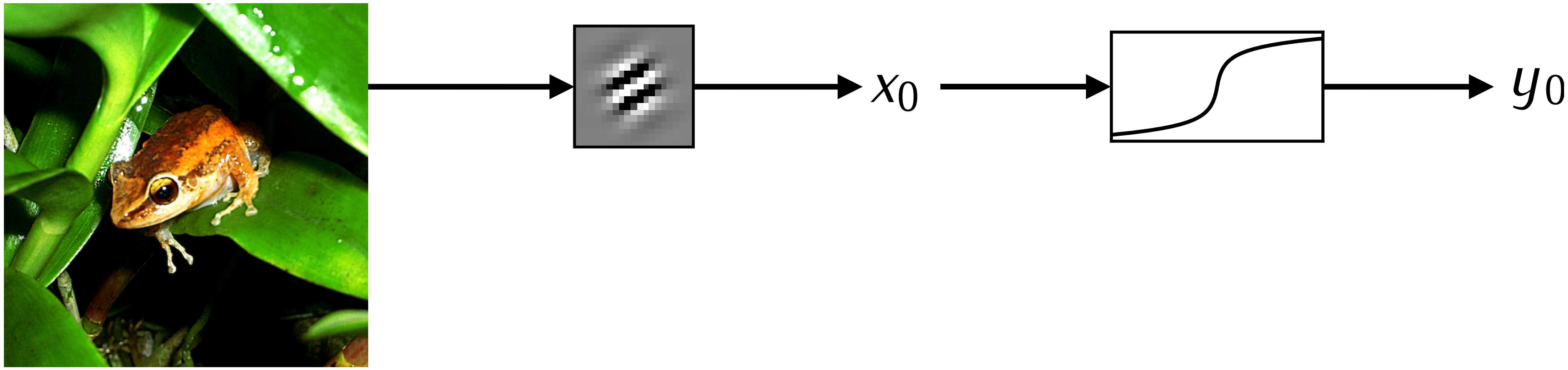


$$y = \frac{x}{\left(\beta + \gamma|x|\right)^{\varepsilon}}$$

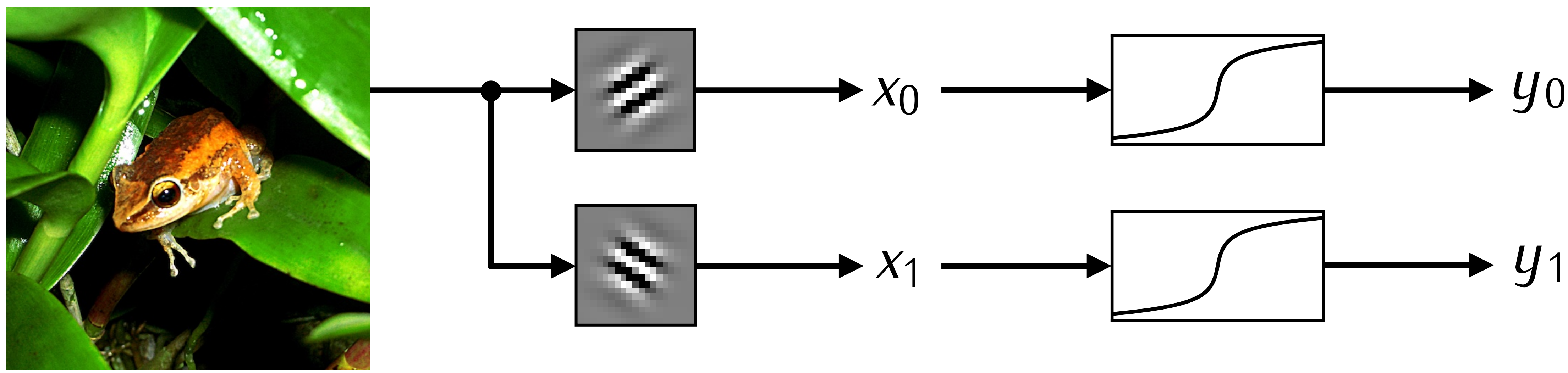
$$p_x(x) = \frac{\partial y}{\partial x}\mathcal{N}(y)$$


x
$p_x(x)$
$y = \frac{x}{(\beta + \gamma|x|)^{\varepsilon}}$
$p_x(x) = \frac{\partial y}{\partial x} \mathcal{N}(y)$
y
$\mathcal{N}$

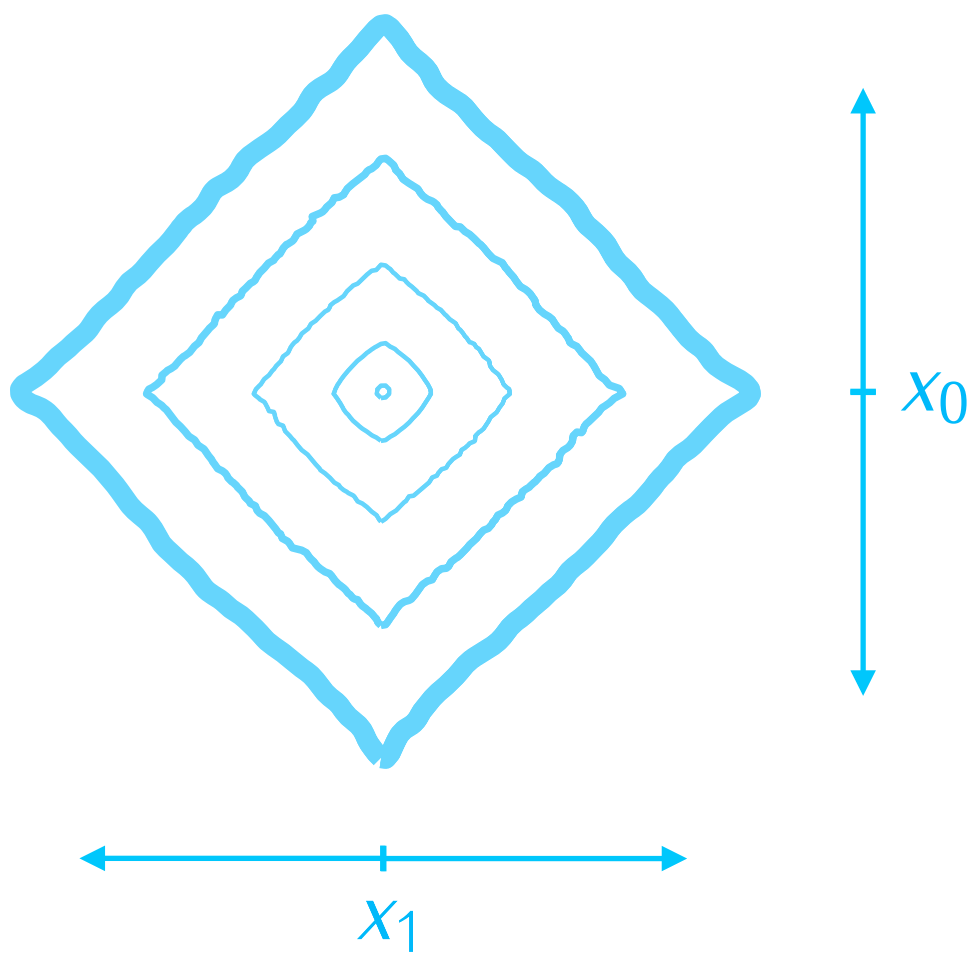
Marginal distribution of linear filter responses

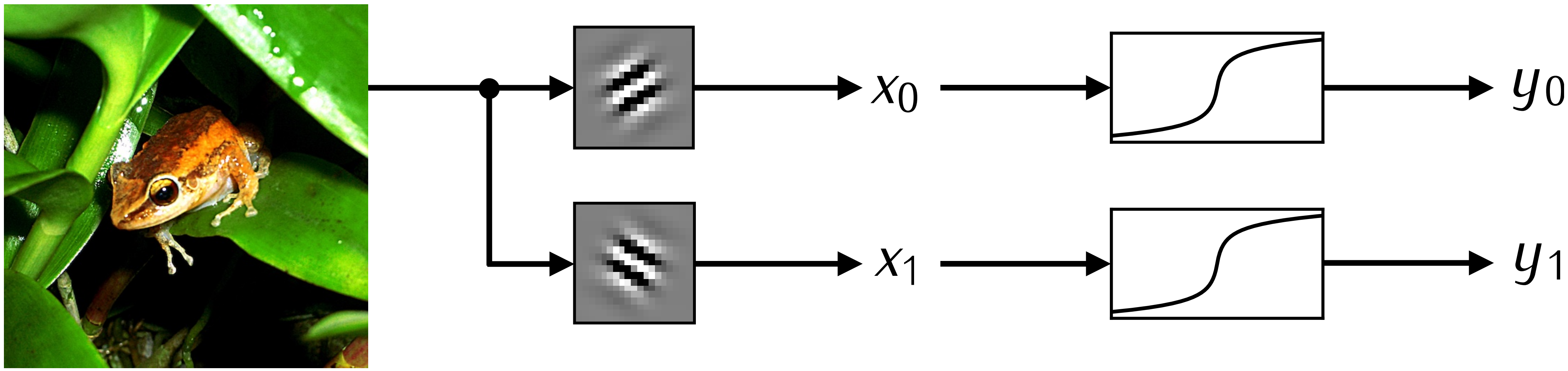


Joint distribution of linear filter responses



contour lines
of joint density

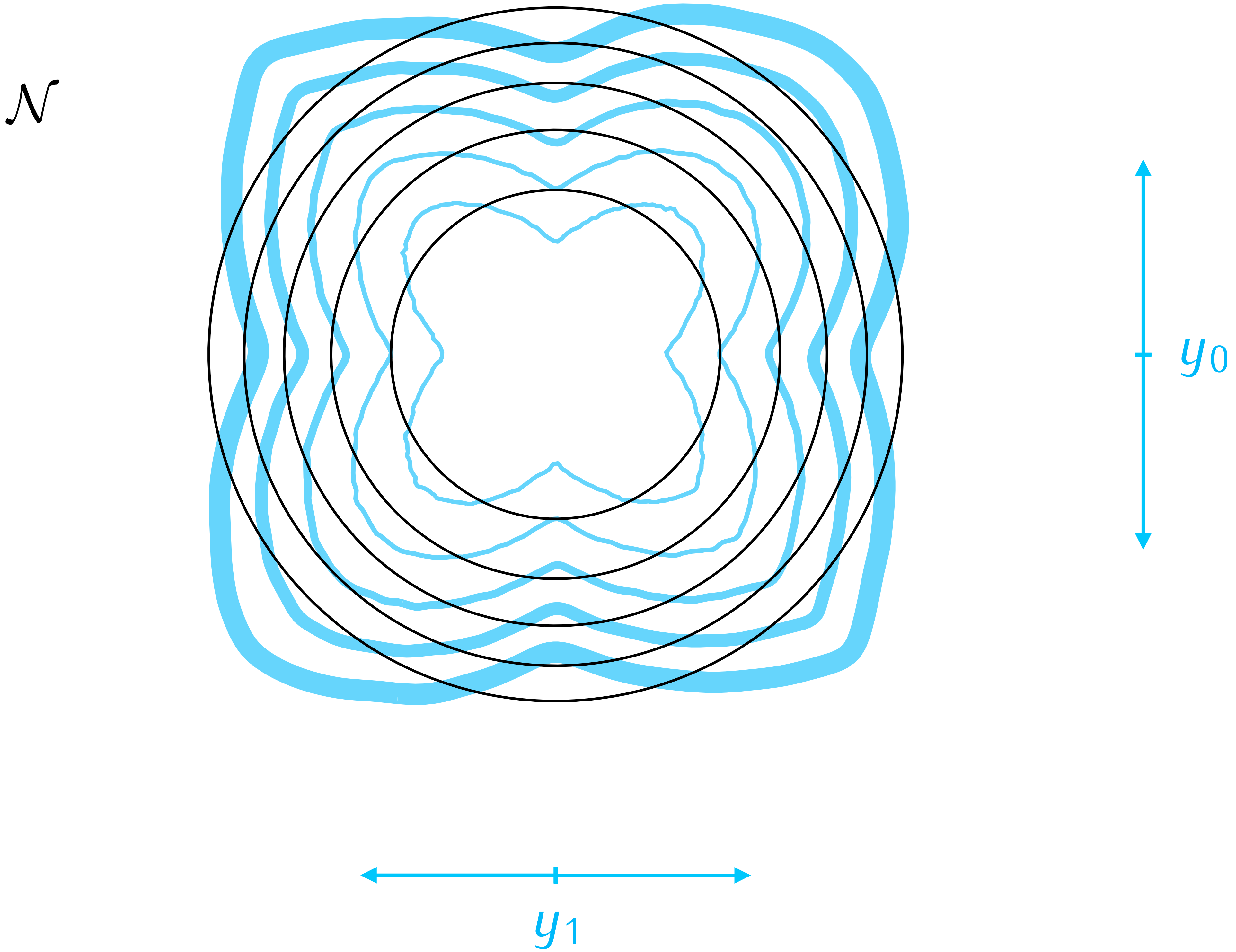




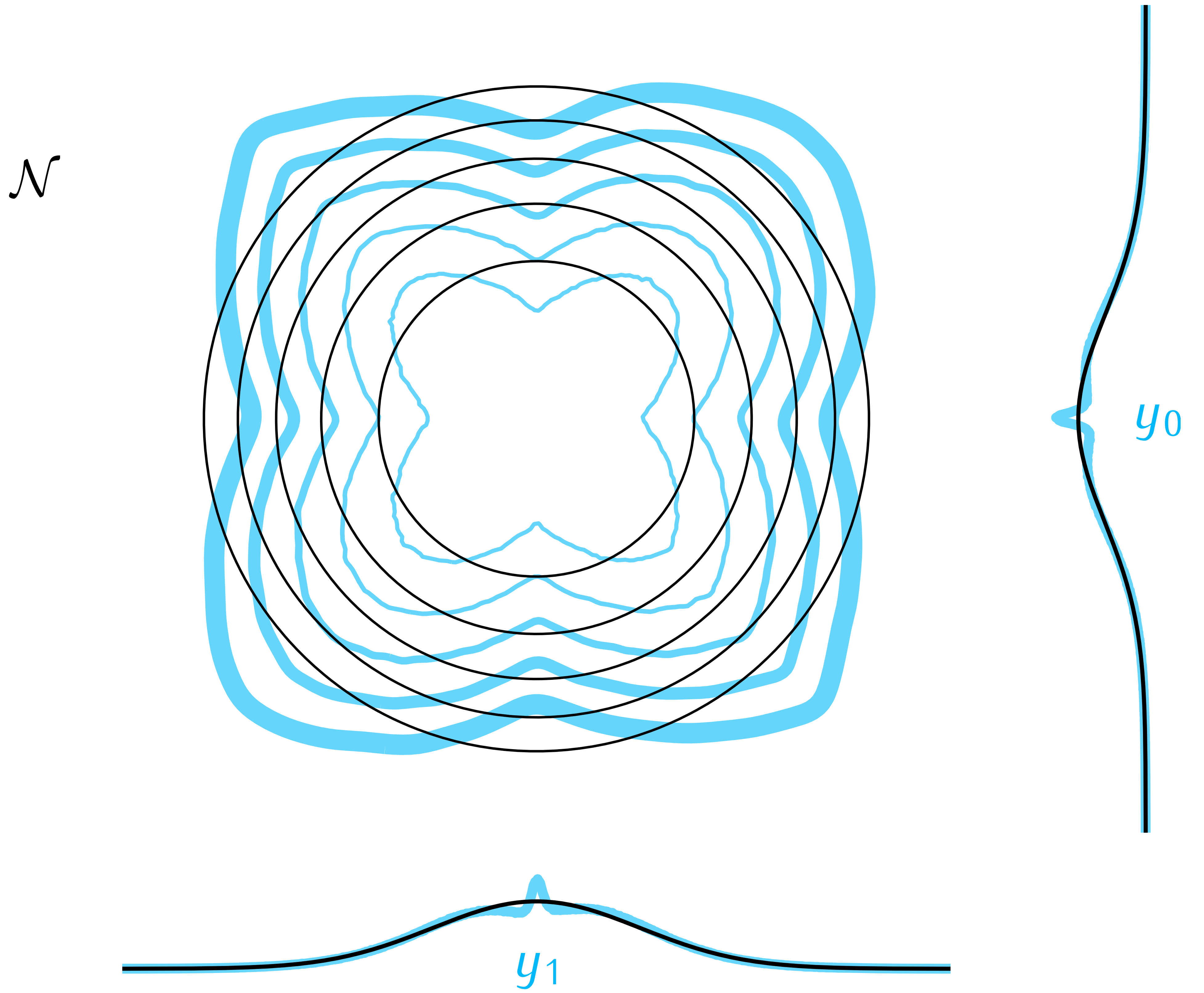
$$y_0 = \frac{x_0}{\left(\beta_0 + \gamma_0 |x_0|^{\alpha_0}\right)^{\varepsilon_0}}$$

$$y_1 = \frac{x_1}{\left(\beta_1 + \gamma_1 |x_1|^{\alpha_1}\right)^{\varepsilon_1}}$$

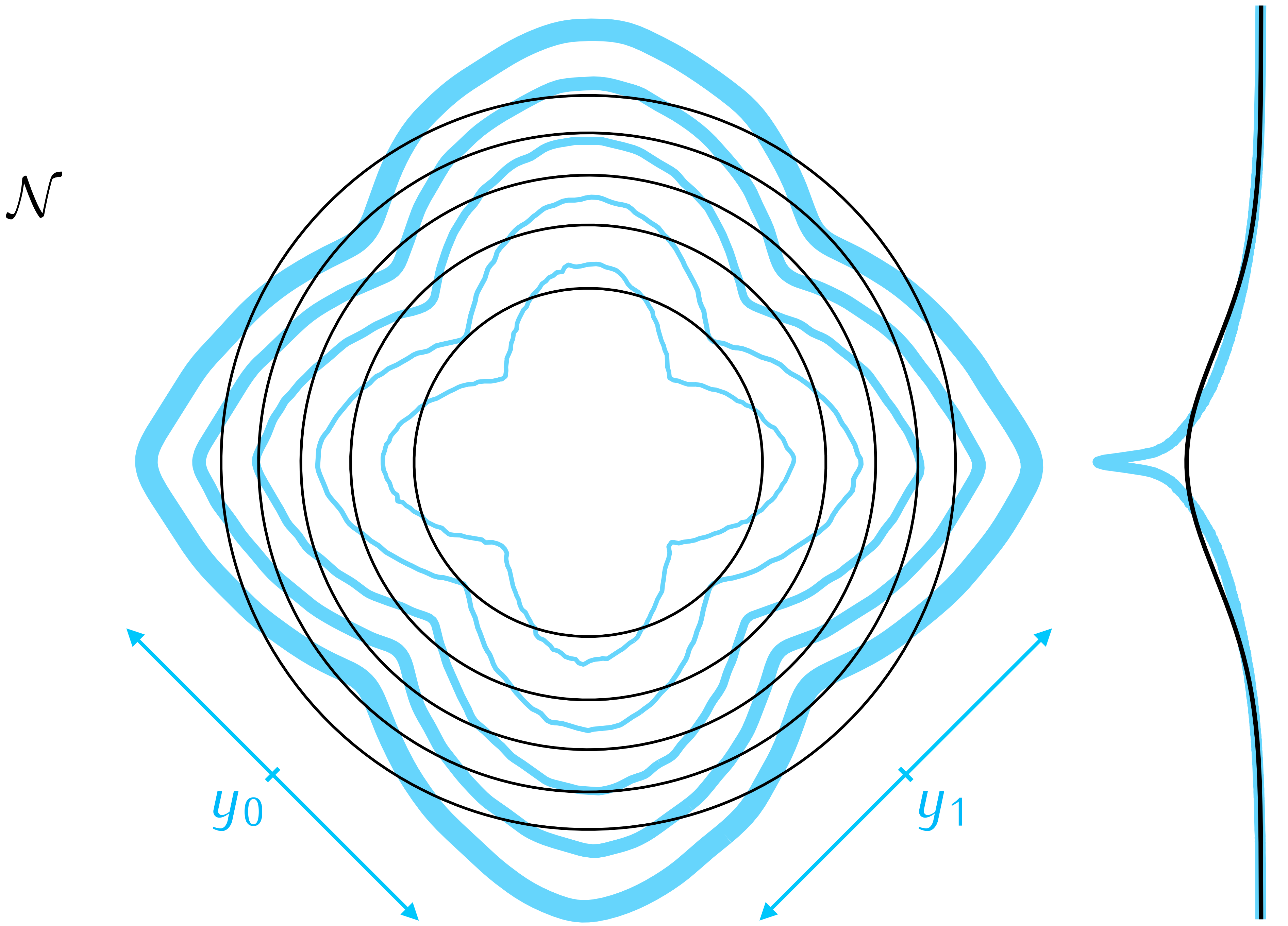
Contour lines, Gaussianized responses



Contour lines, Gaussianized responses

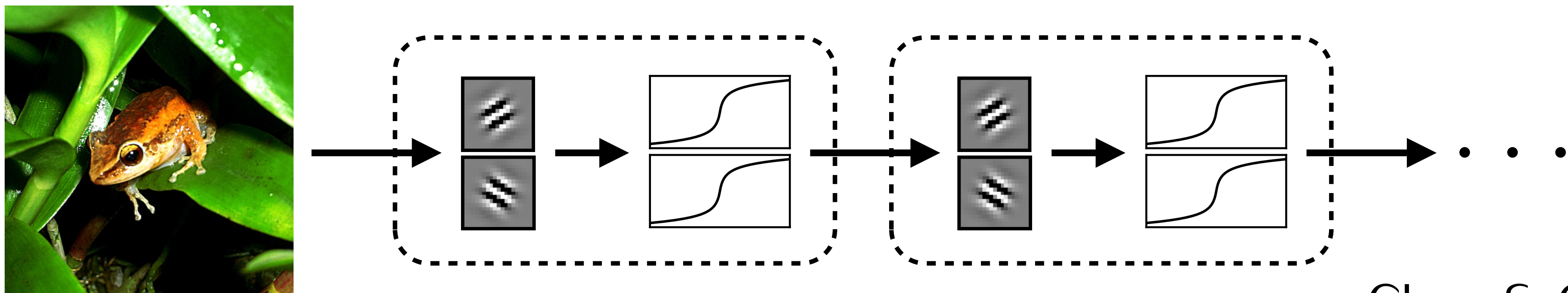


Contour lines, Gaussianized responses



Improved Gaussianization

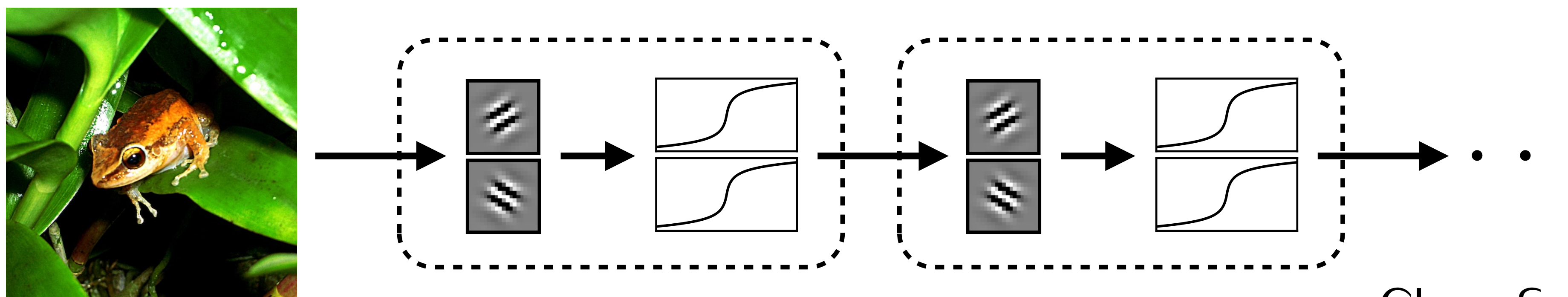
1. Iterated marginal Gaussianization



Chen & Gopinath, 2001
Laparra et al., 2010

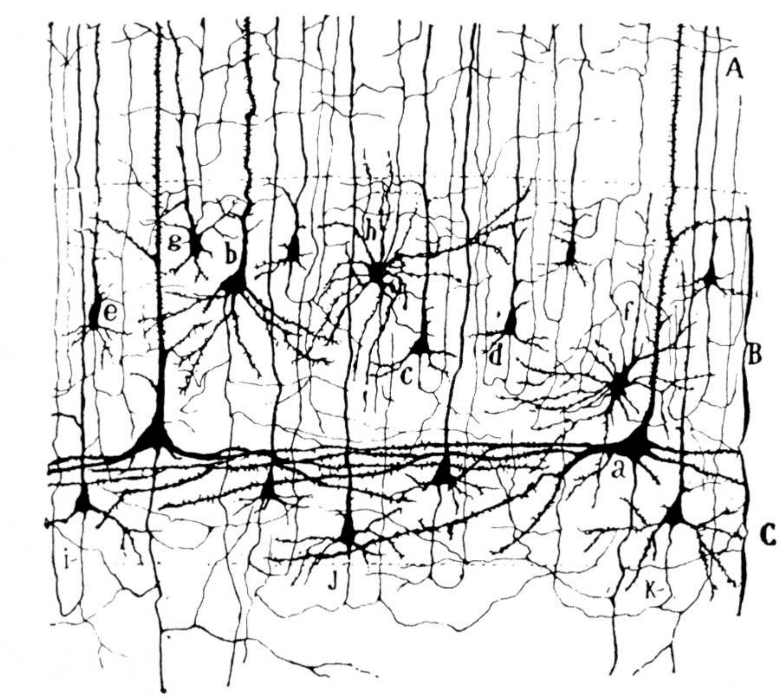
Improved Gaussianization

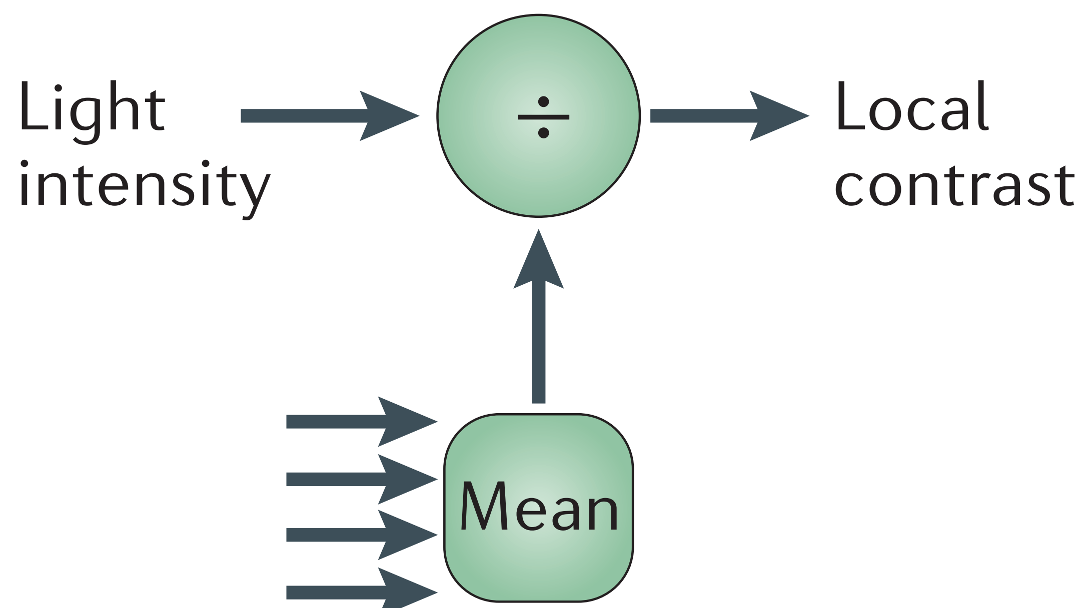
1. Iterated marginal Gaussianization



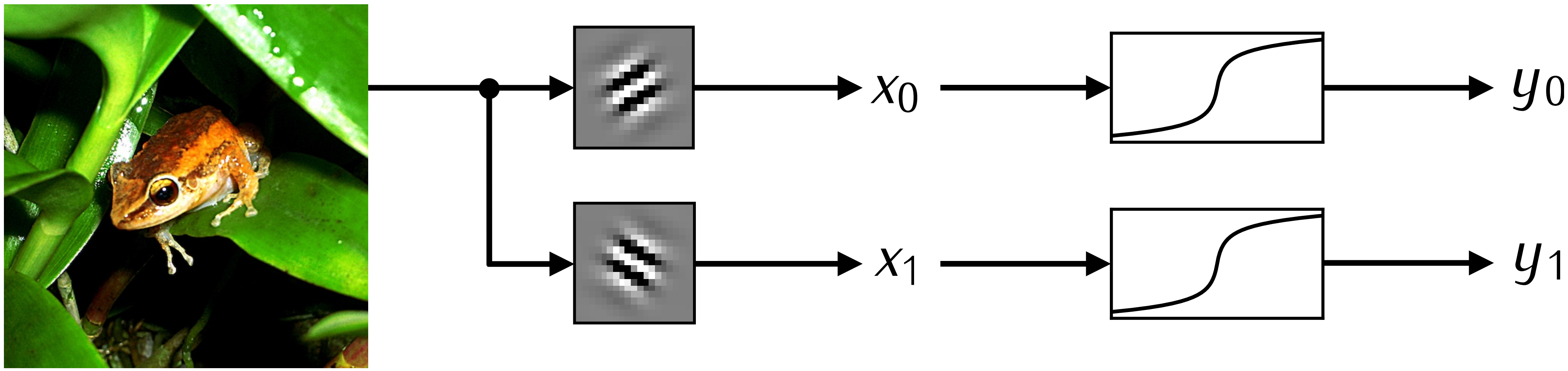
Chen & Gopinath, 2001
Laparra et al., 2010

2. Joint Gaussianization (inspired by biology)



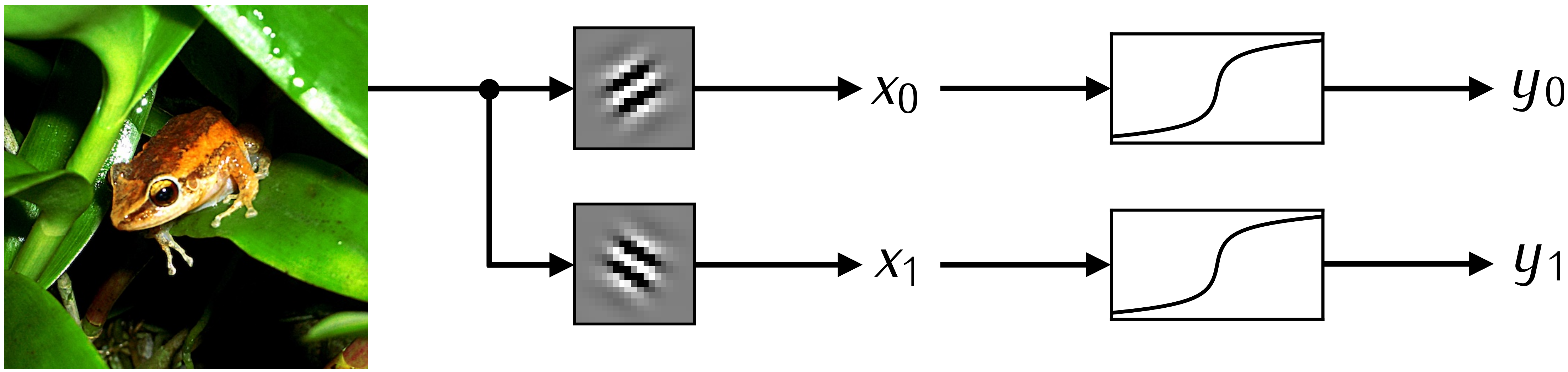


Heeger, 1992
Schwartz & Simoncelli, 2001
figures: Cajal; Carandini & Heeger, 2012



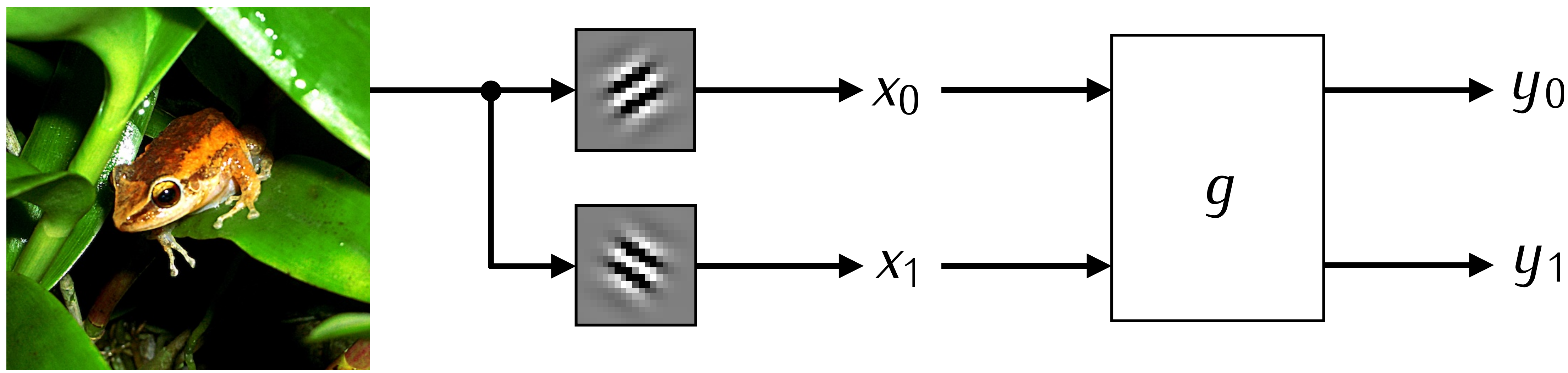
$$y_0 = \frac{x_0}{\left(\beta_0 + \gamma_0 |x_0|^{\alpha_0}\right)^{\varepsilon_0}}$$

$$y_1 = \frac{x_1}{\left(\beta_1 + \gamma_1 |x_1|^{\alpha_1}\right)^{\varepsilon_1}}$$



$$y_0 = \frac{x_0}{\left(\beta_0 + \gamma_0 |x_0|^{\alpha_0}\right)^{\varepsilon_0}}$$

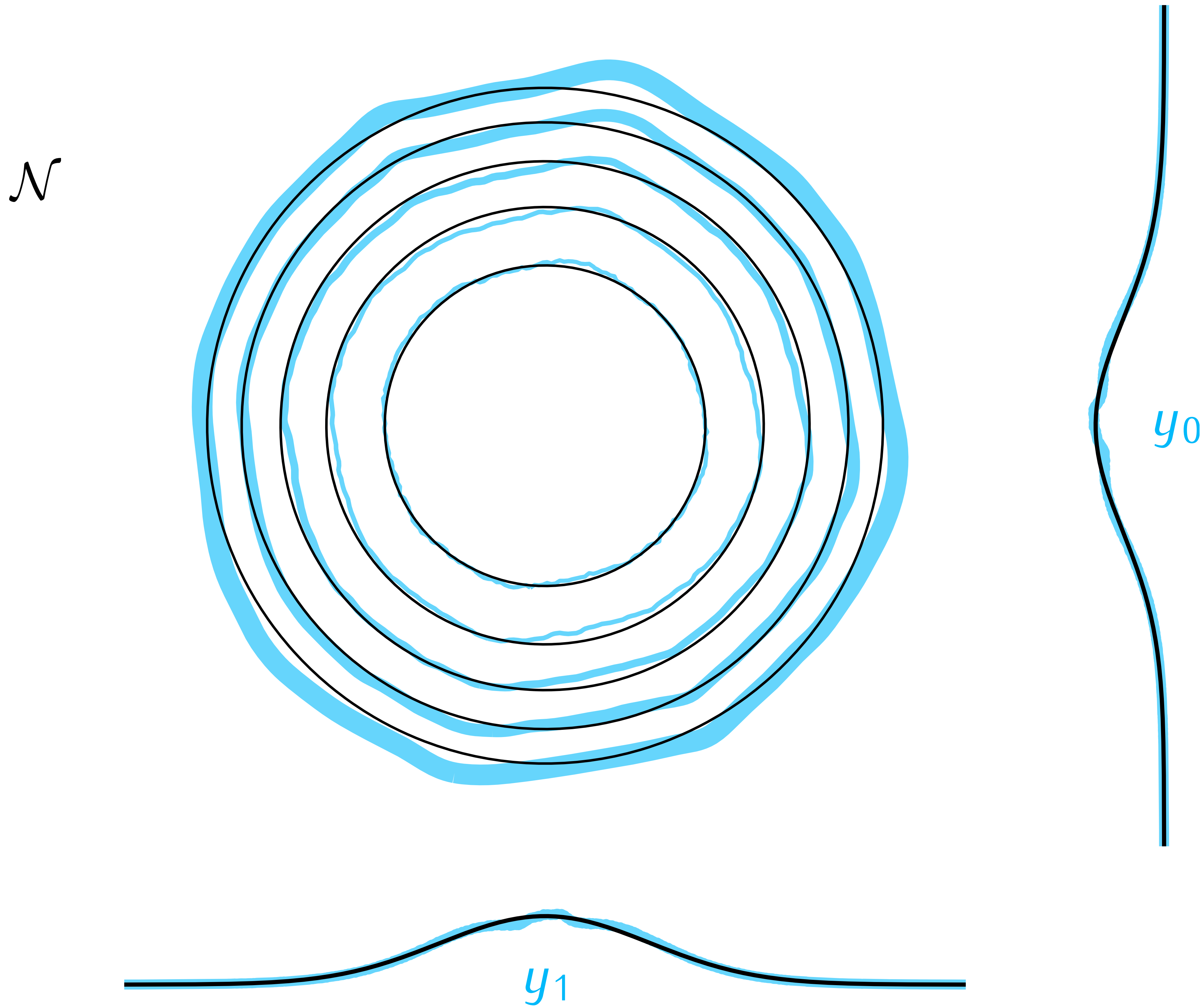
$$y_1 = \frac{x_1}{\left(\beta_1 + \gamma_1 |x_1|^{\alpha_1}\right)^{\varepsilon_1}}$$



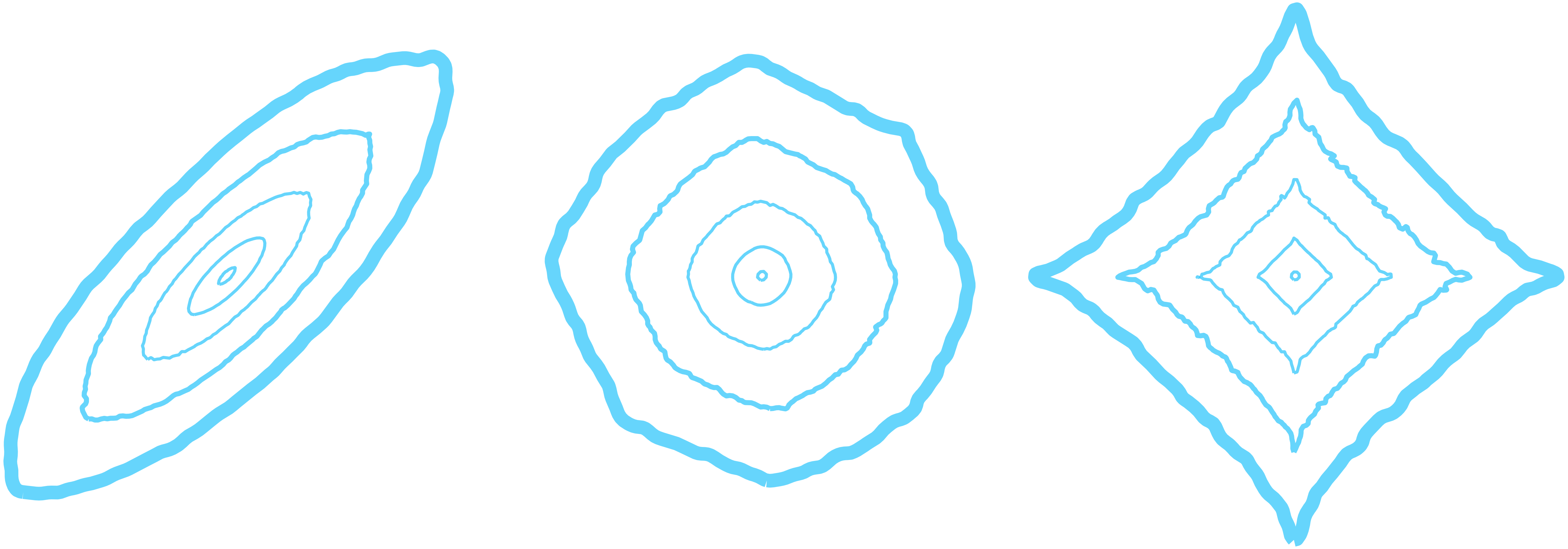
$$y_0 = \frac{x_0}{\left(\beta_0 + \gamma_{01}|x_1|^{\alpha_{01}} + \gamma_{00}|x_0|^{\alpha_{00}}\right)^{\varepsilon_0}}$$

$$y_1 = \frac{x_1}{\left(\beta_1 + \gamma_{10}|x_0|^{\alpha_{10}} + \gamma_{11}|x_1|^{\alpha_{11}}\right)^{\varepsilon_1}}$$

Contour lines, Gaussianized responses



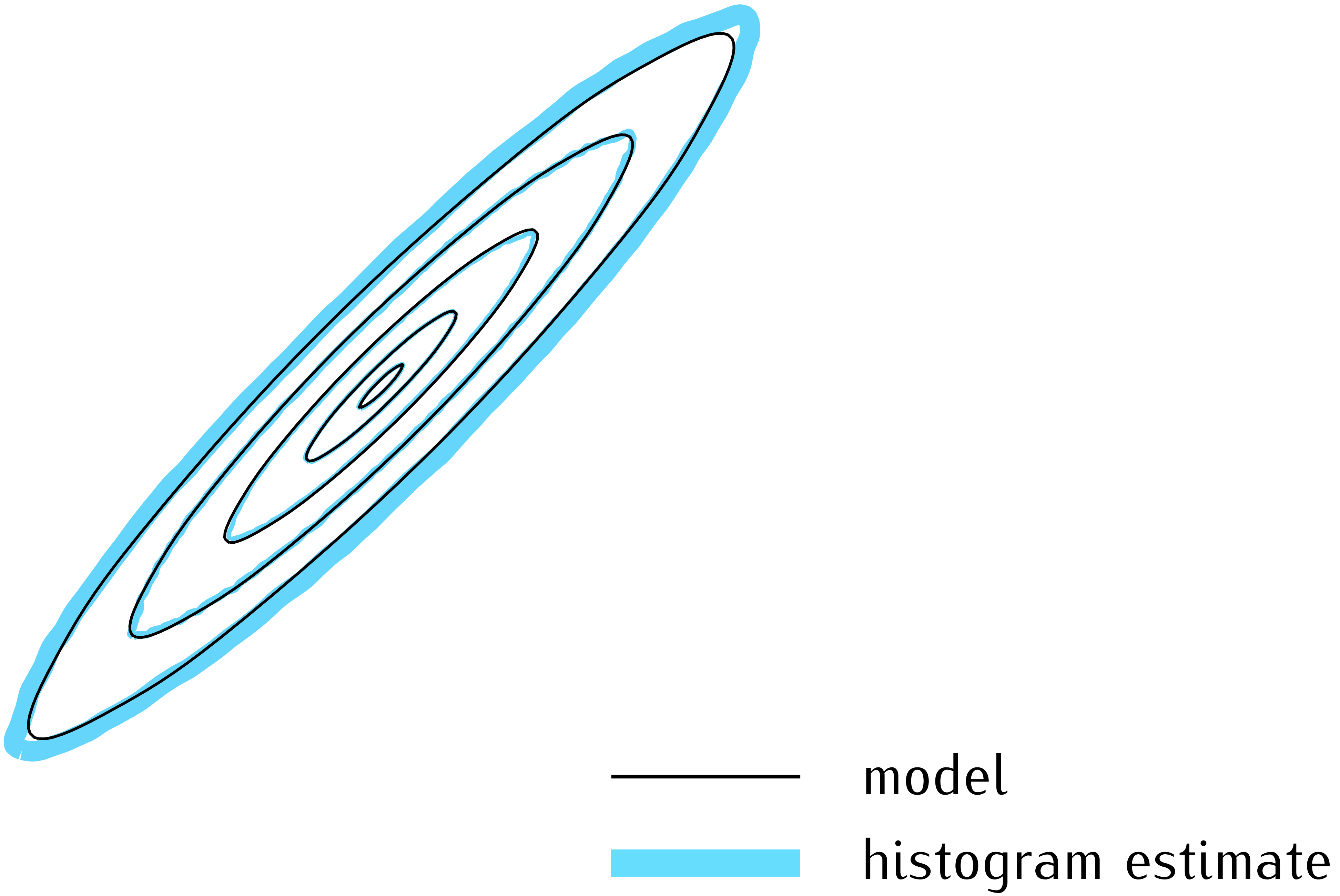
Variety of shapes, joint density of filter responses



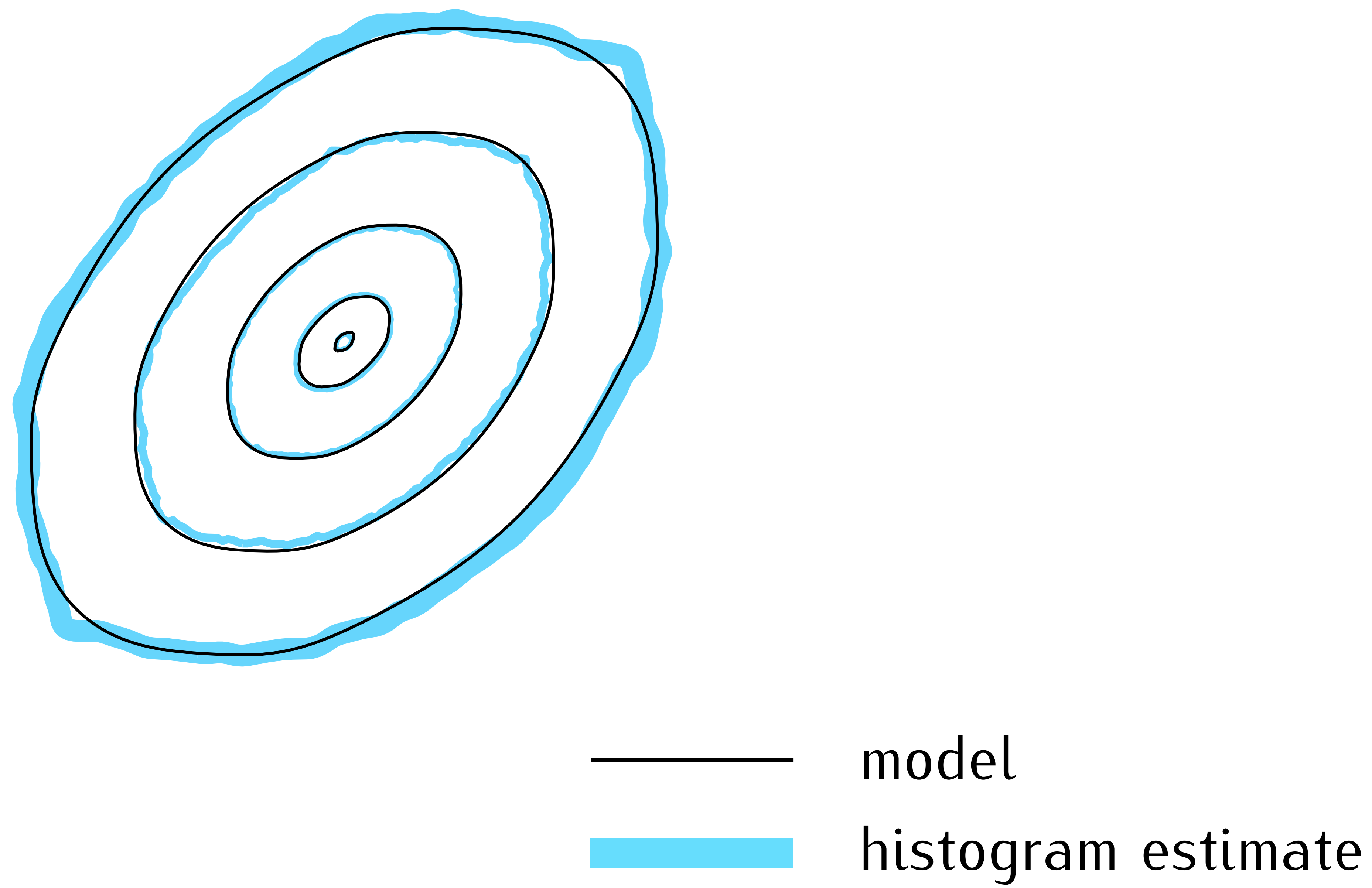
elliptical ? marginally independent

Lyu & Simoncelli, 2009
Sinz et al., 2009

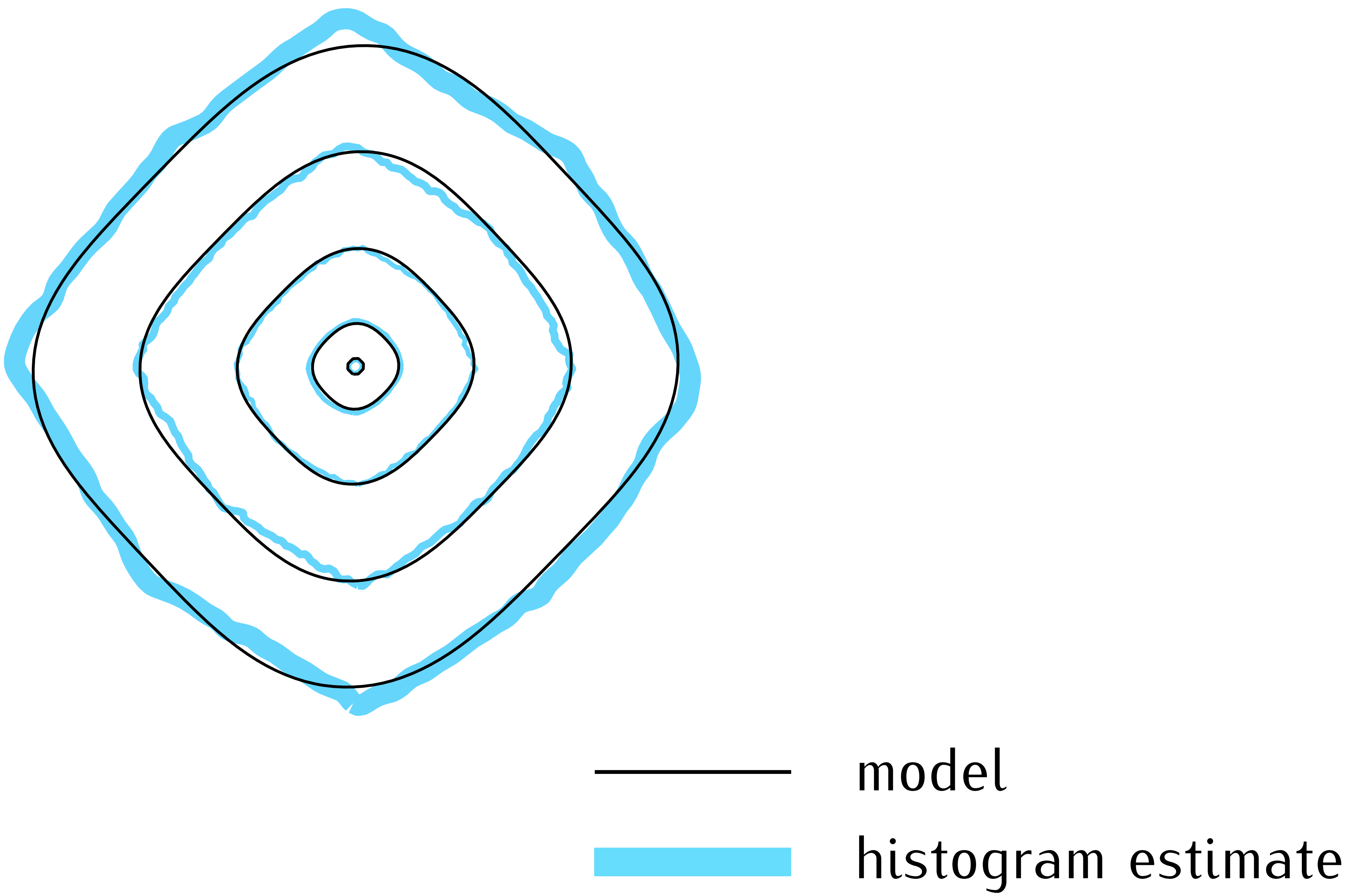
Contour lines, linear filter responses



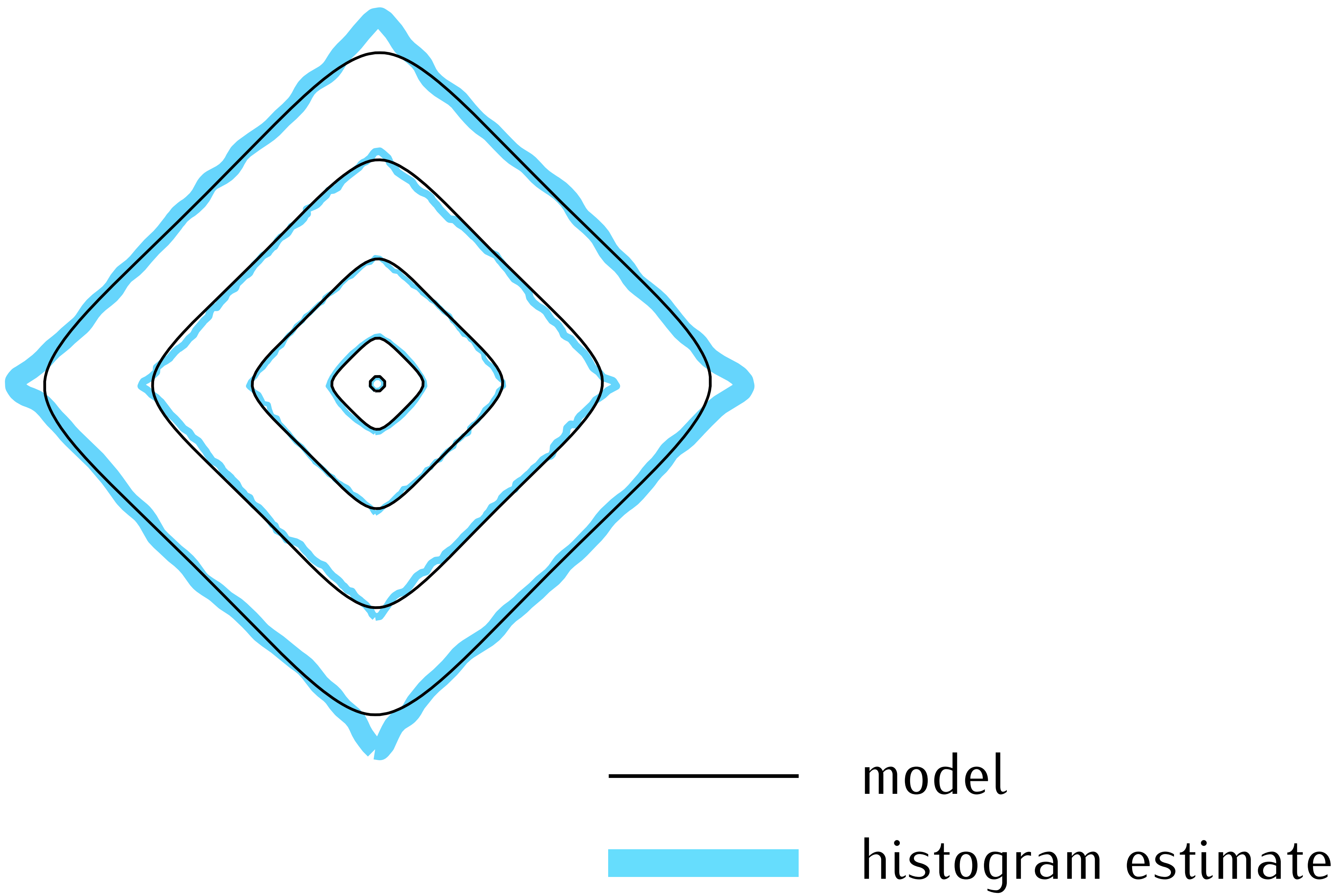
Contour lines, linear filter responses



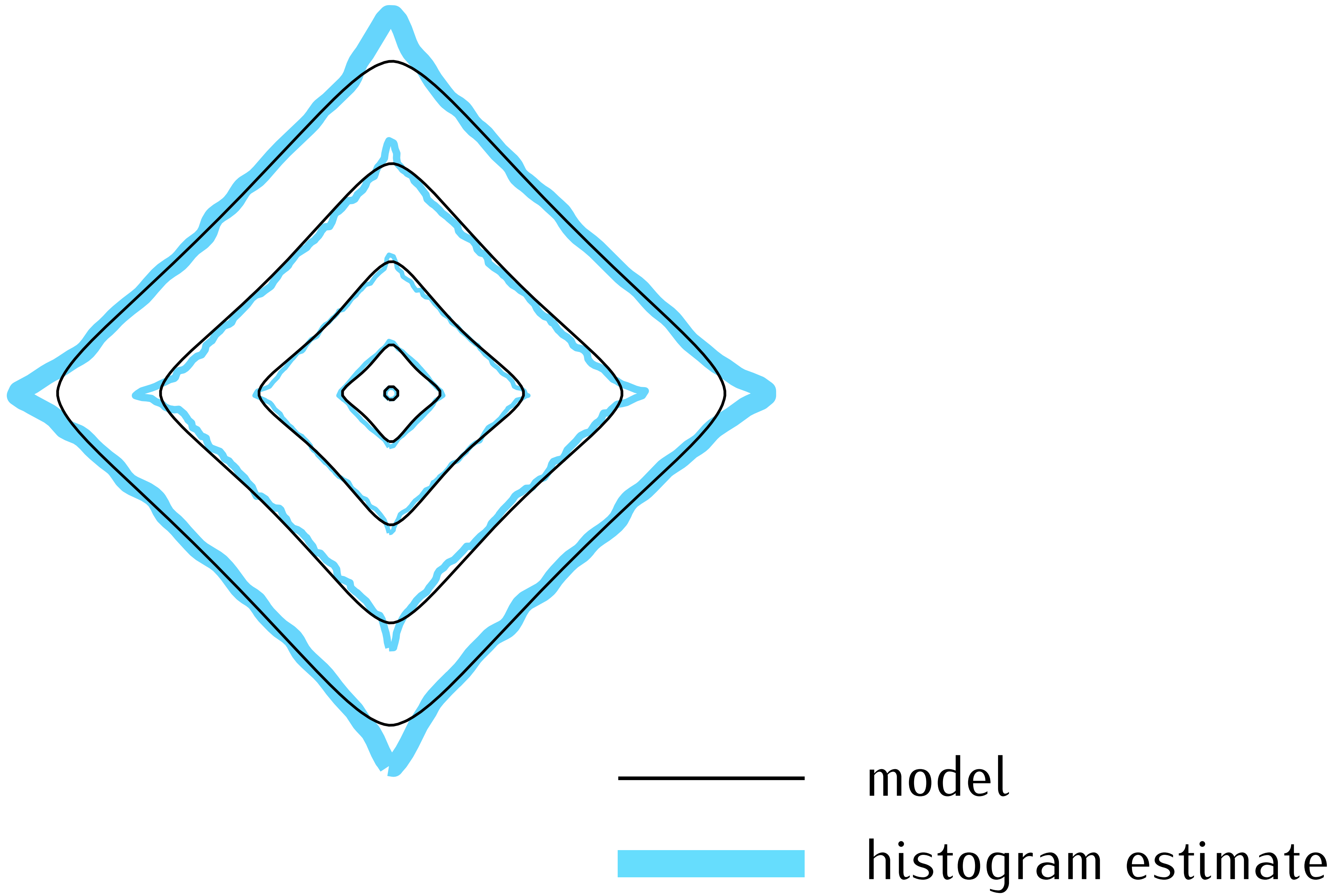
Contour lines, linear filter responses



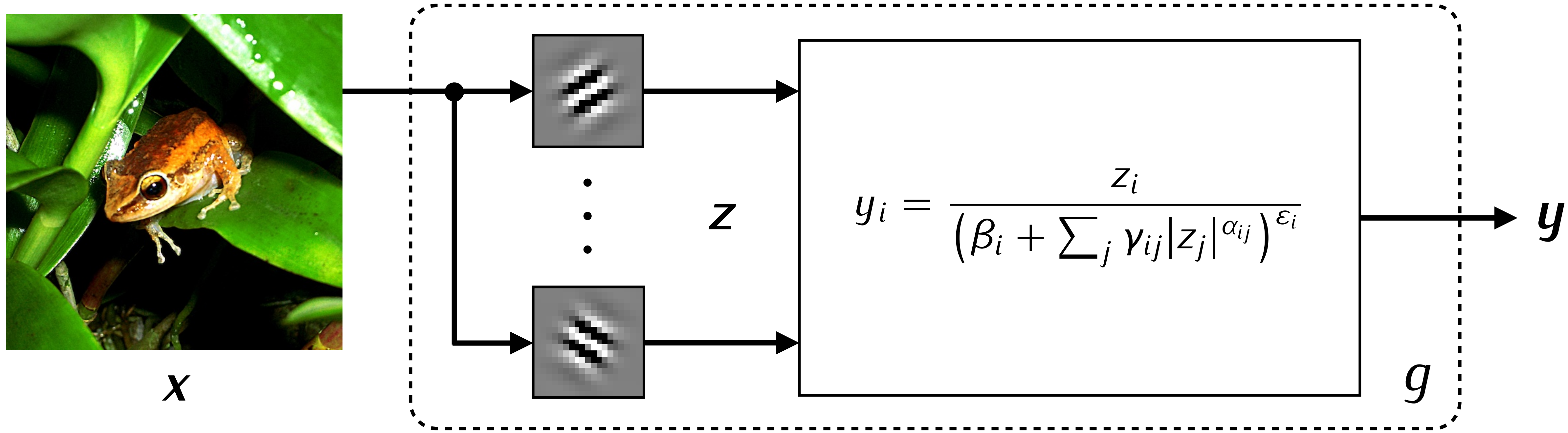
Contour lines, linear filter responses



Contour lines, linear filter responses



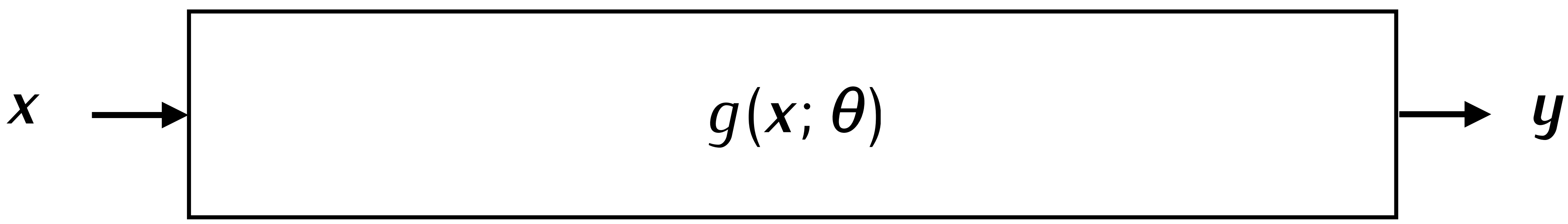
Generalized divisive normalization (GDN)



Special cases/related models:

- Independent Component Analysis, Cardoso, 2003
- Independent Subspace Analysis, Hyvärinen & Hoyer, 2000
- Weighted normalization model, Schwartz & Simoncelli, 2001
- Topographic ICA, Hyvärinen et al., 2001
- Radial Gaussianization, Lyu & Simoncelli, 2009
- L_p-nested symmetric distributions, Sinz & Bethge, 2010
- "Two-layer model", Köster & Hyvärinen, 2010

Parameter estimation (multiple layers)



$$-\log p_x(\boldsymbol{x}) = -\log \left| \frac{\partial g(\boldsymbol{x}; \boldsymbol{\theta})}{\partial \boldsymbol{x}} \right| - \frac{1}{2} \left\| g(\boldsymbol{x}; \boldsymbol{\theta}) \right\|_2^2 + C$$

minimize wrt. $\boldsymbol{\theta}$ using stochastic gradient descent

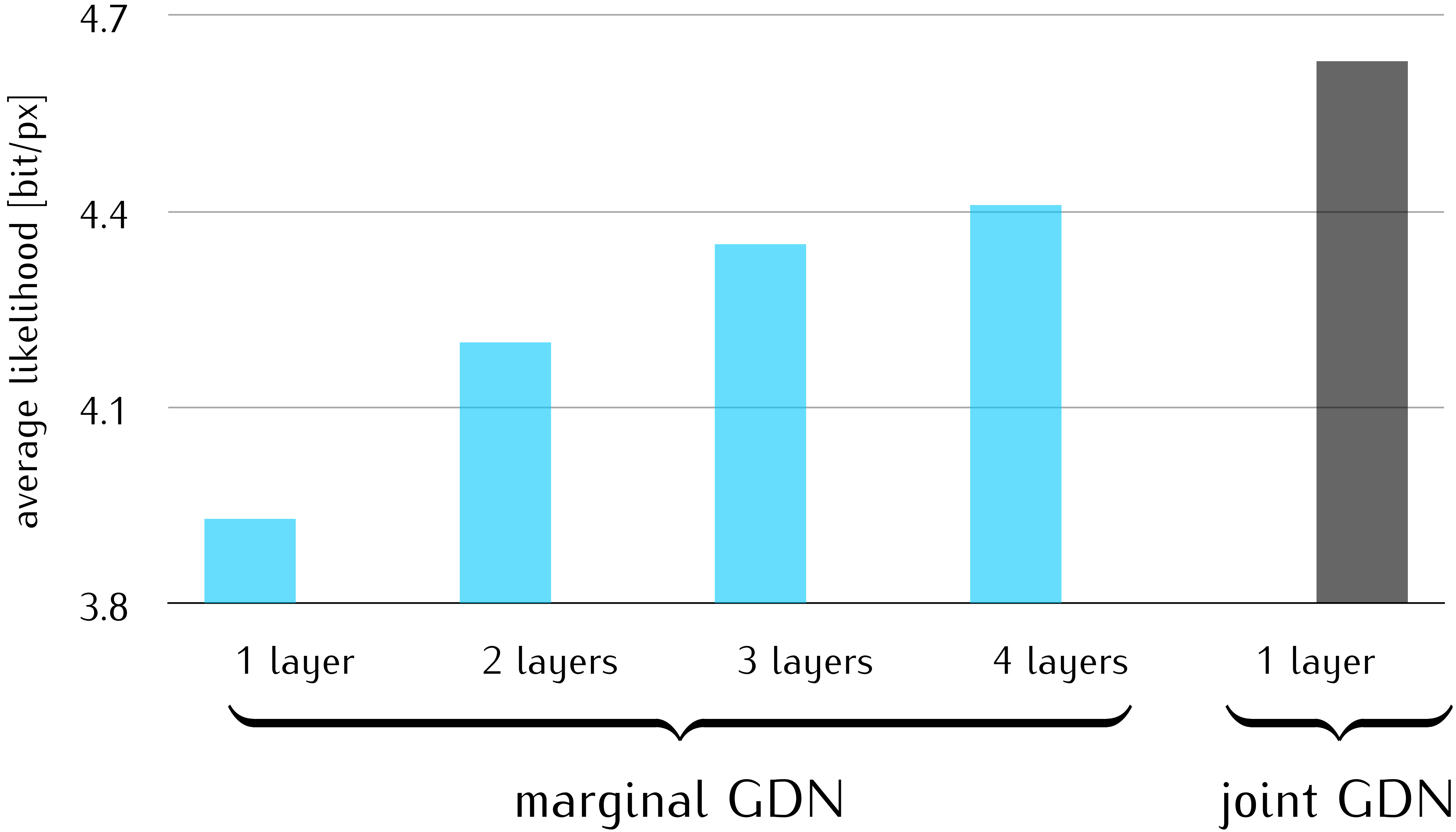
Parameter estimation (multiple layers)

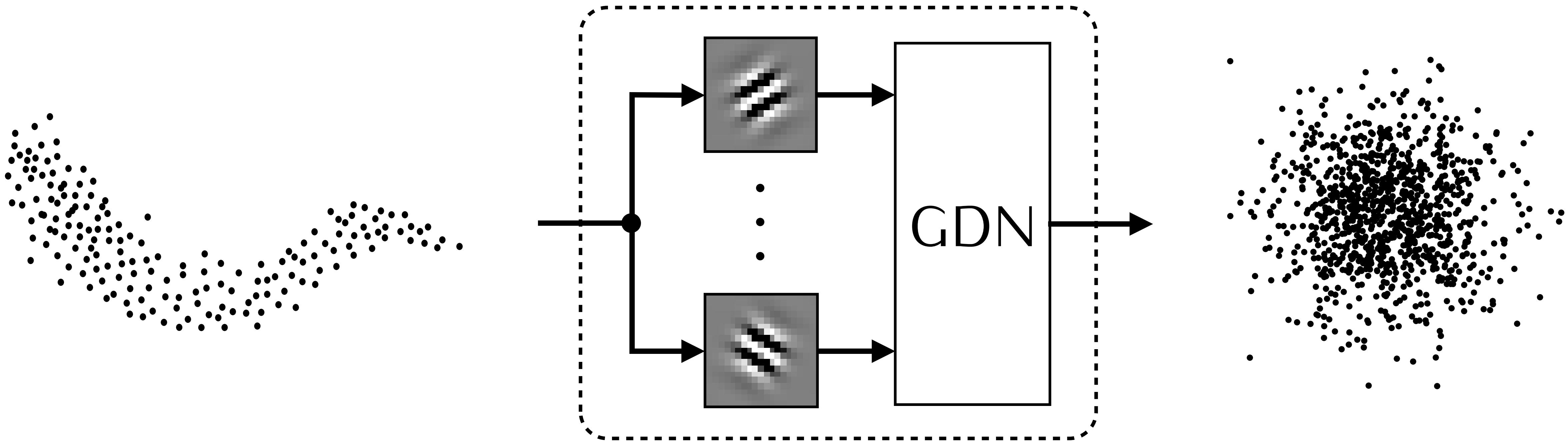
$$\boldsymbol{x} \longrightarrow \boxed{g_0(\boldsymbol{x}_0;\boldsymbol{\theta})} \longrightarrow \boxed{g_1(\boldsymbol{x}_1;\boldsymbol{\theta})} \longrightarrow \boldsymbol{y}$$

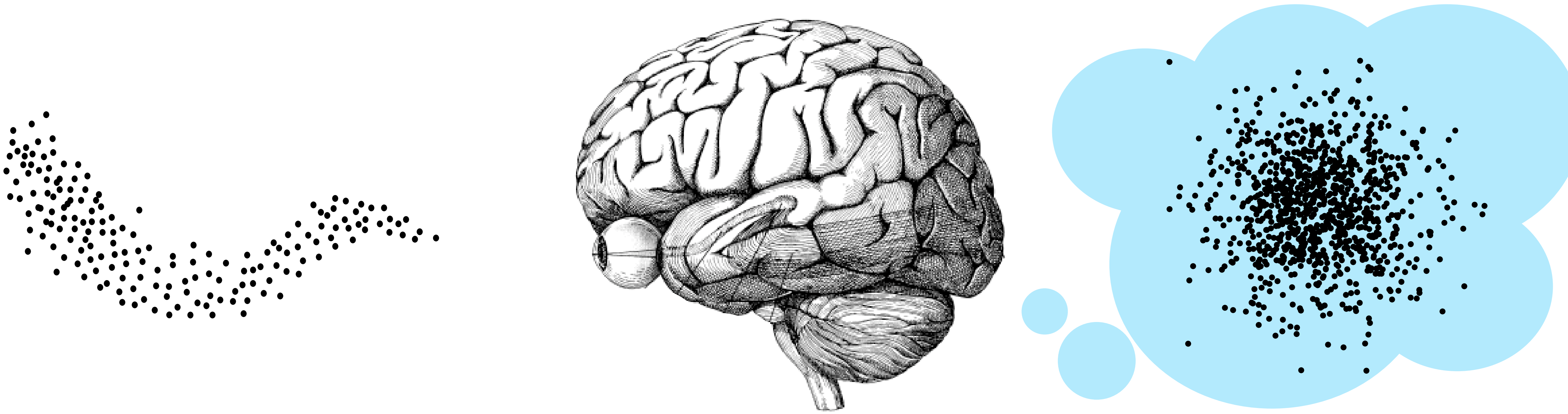
$$-\log p_{\boldsymbol{x}}(\boldsymbol{x}) = \underbrace{-\log\left|\frac{\partial g(\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{x}}\right|}_{-\log\left|\frac{\partial g_0(\boldsymbol{x}_0;\boldsymbol{\theta})}{\partial \boldsymbol{x}_0}\right| - \log\left|\frac{\partial g_1(\boldsymbol{x}_1;\boldsymbol{\theta})}{\partial \boldsymbol{x}_1}\right| - \dots} - \frac{1}{2}\left\|g(\boldsymbol{x};\boldsymbol{\theta})\right\|_2^2 + C$$

minimize wrt. $\boldsymbol{\theta}$ using stochastic gradient descent

One layer of joint GDN > many layers of marginal GDN

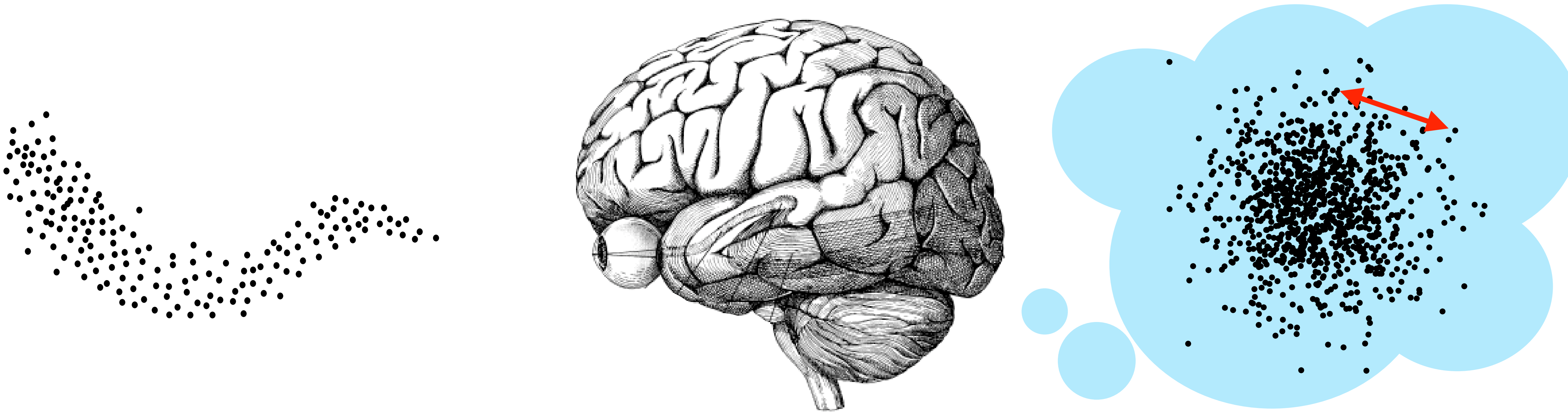



GDN



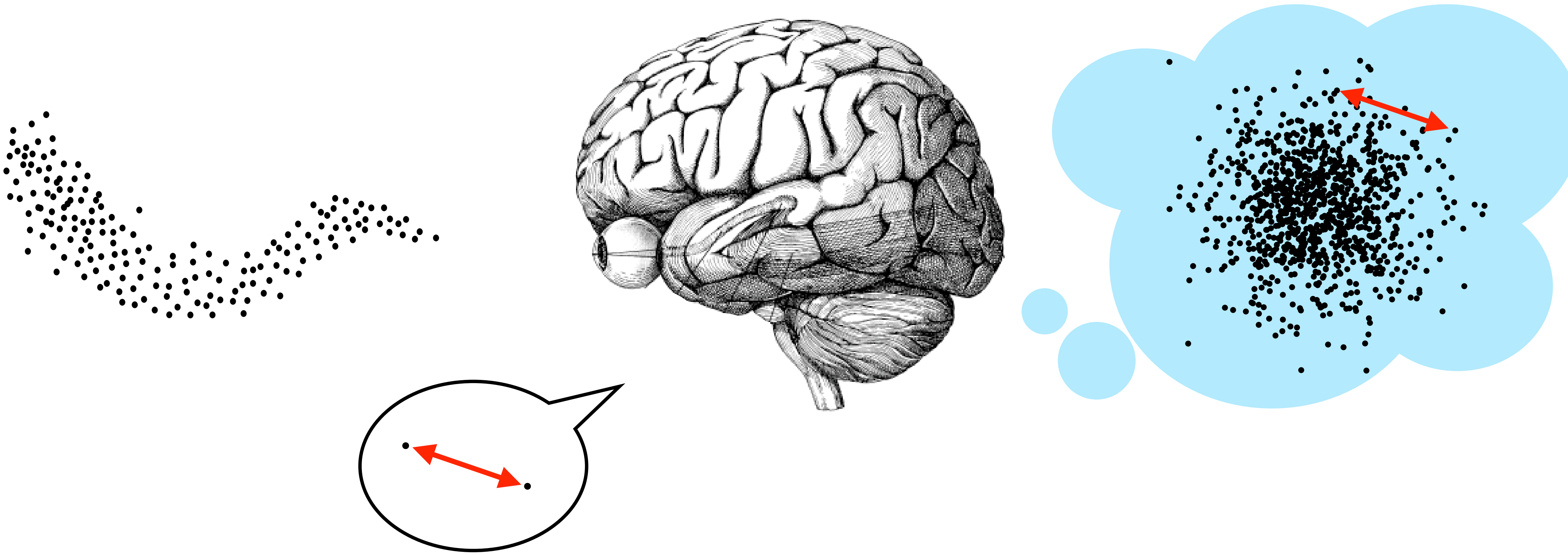
What are the perceptual properties of the representation?

figure: Hubel, 1995



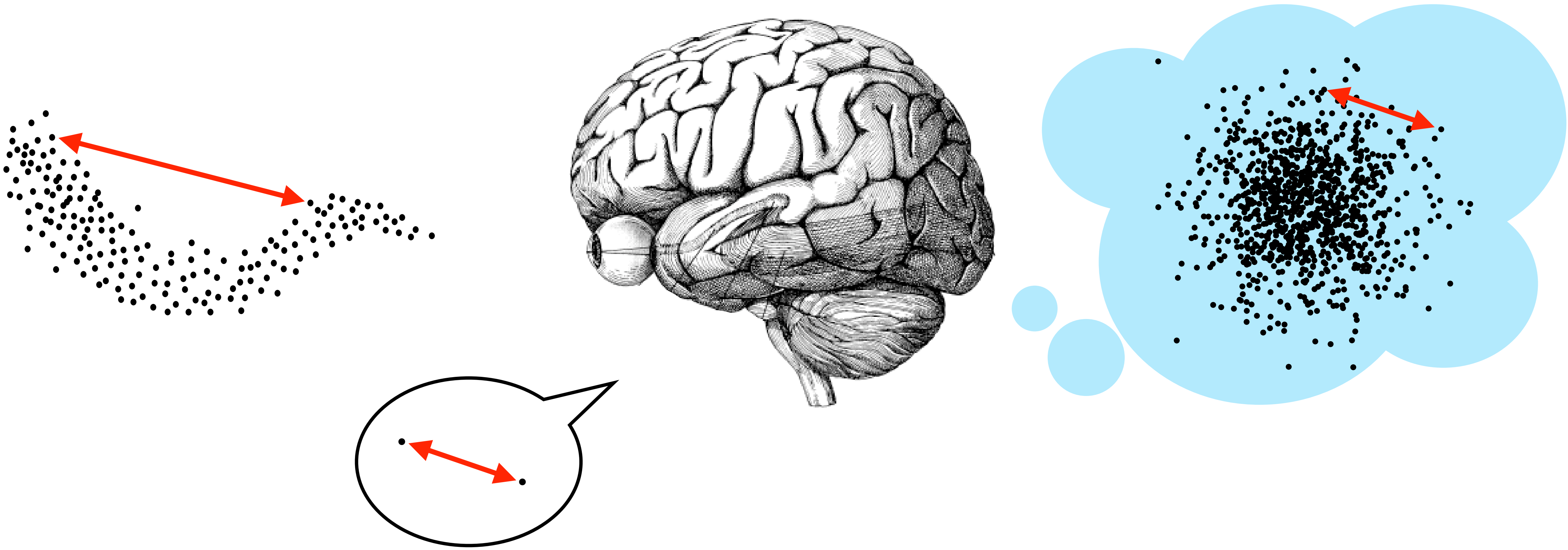
What are the perceptual properties of the representation?

figure: Hubel, 1995



What are the perceptual properties of the representation?

figure: Hubel, 1995

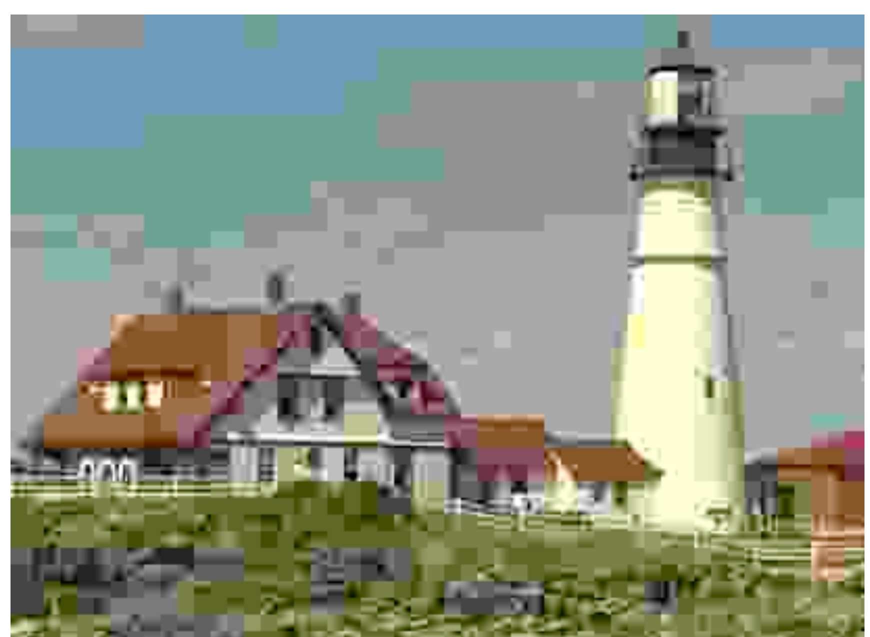


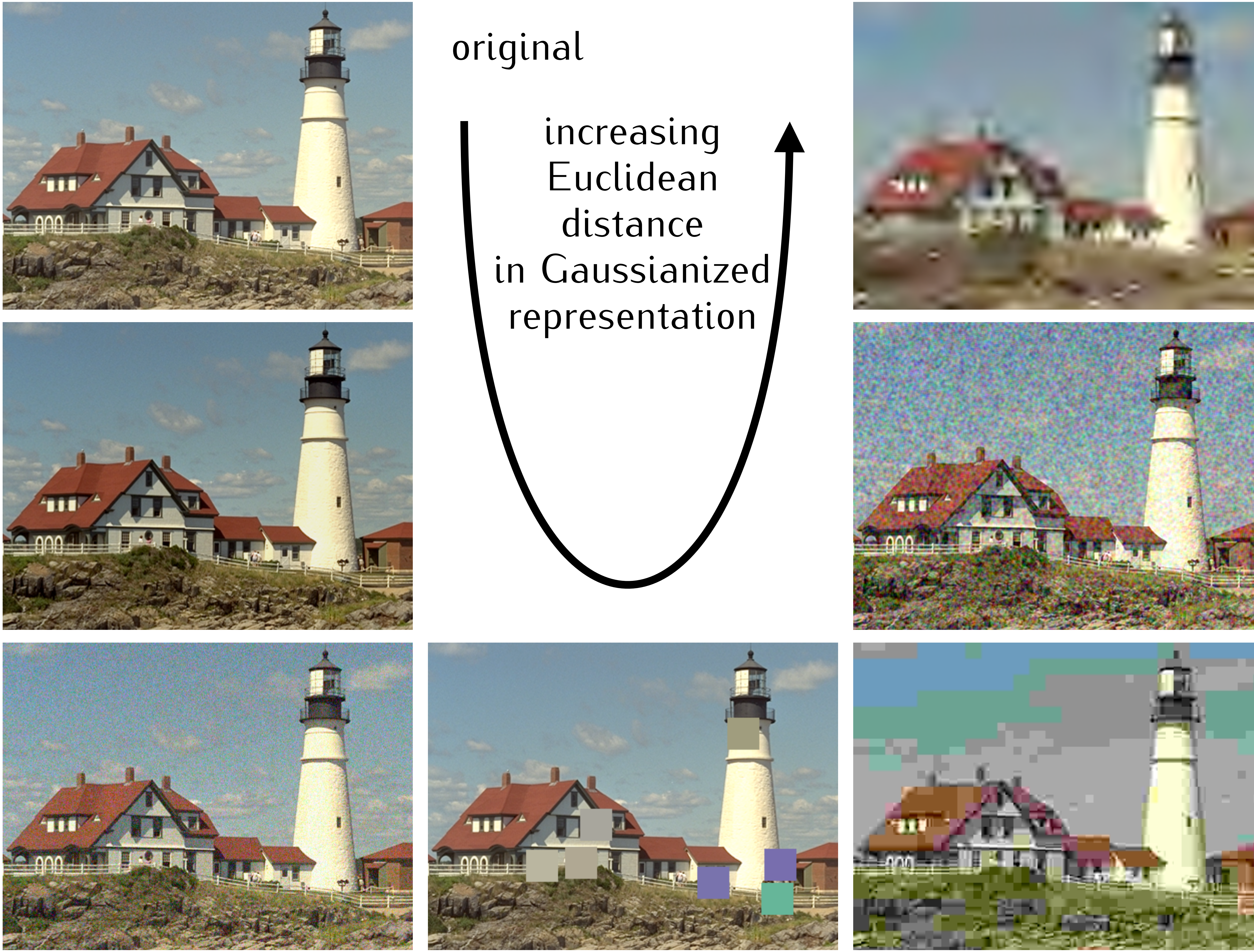
What are the perceptual properties of the representation?

figure: Hubel, 1995

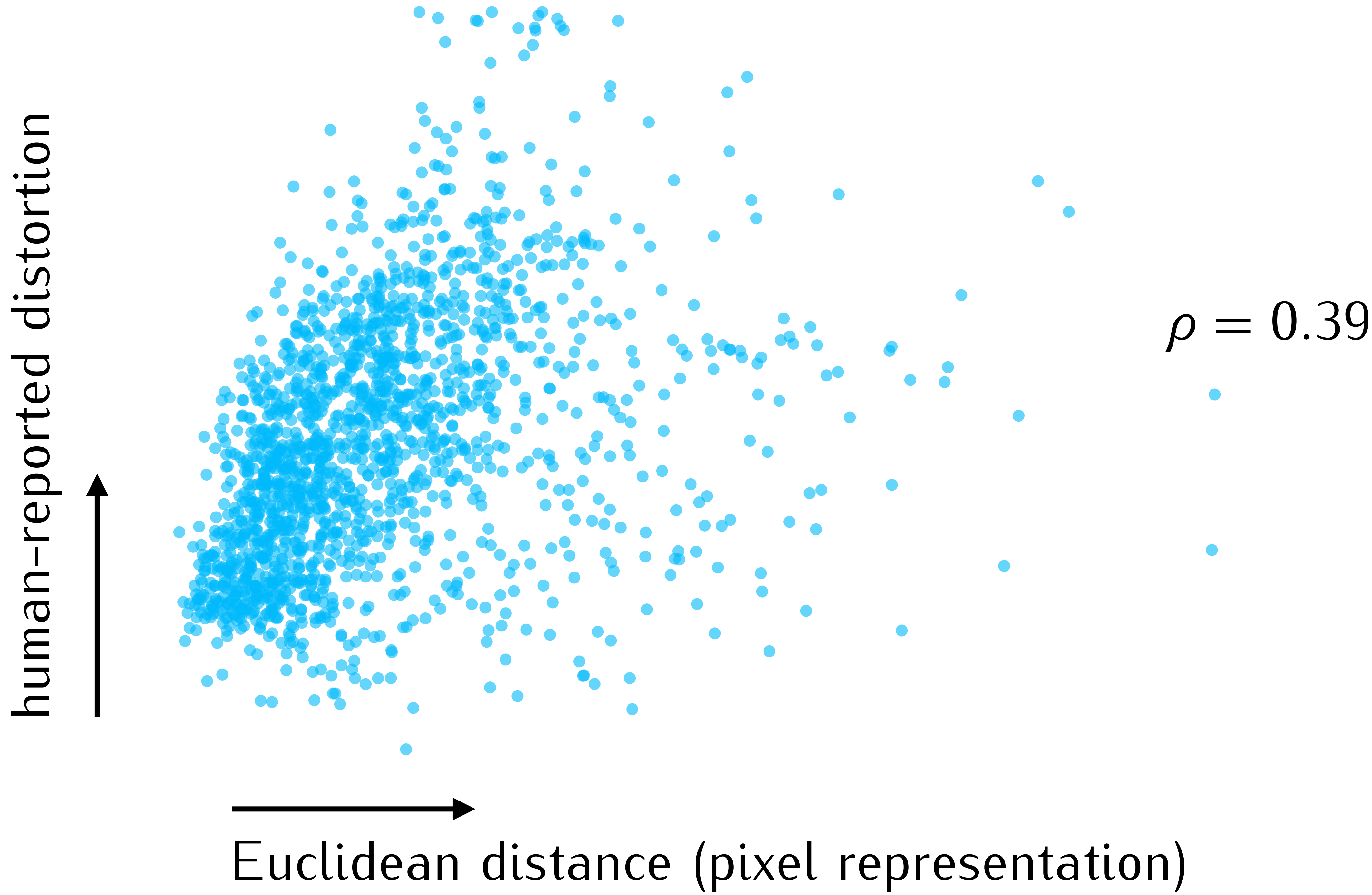
original

increasing
Euclidean
distance
in pixel
representation



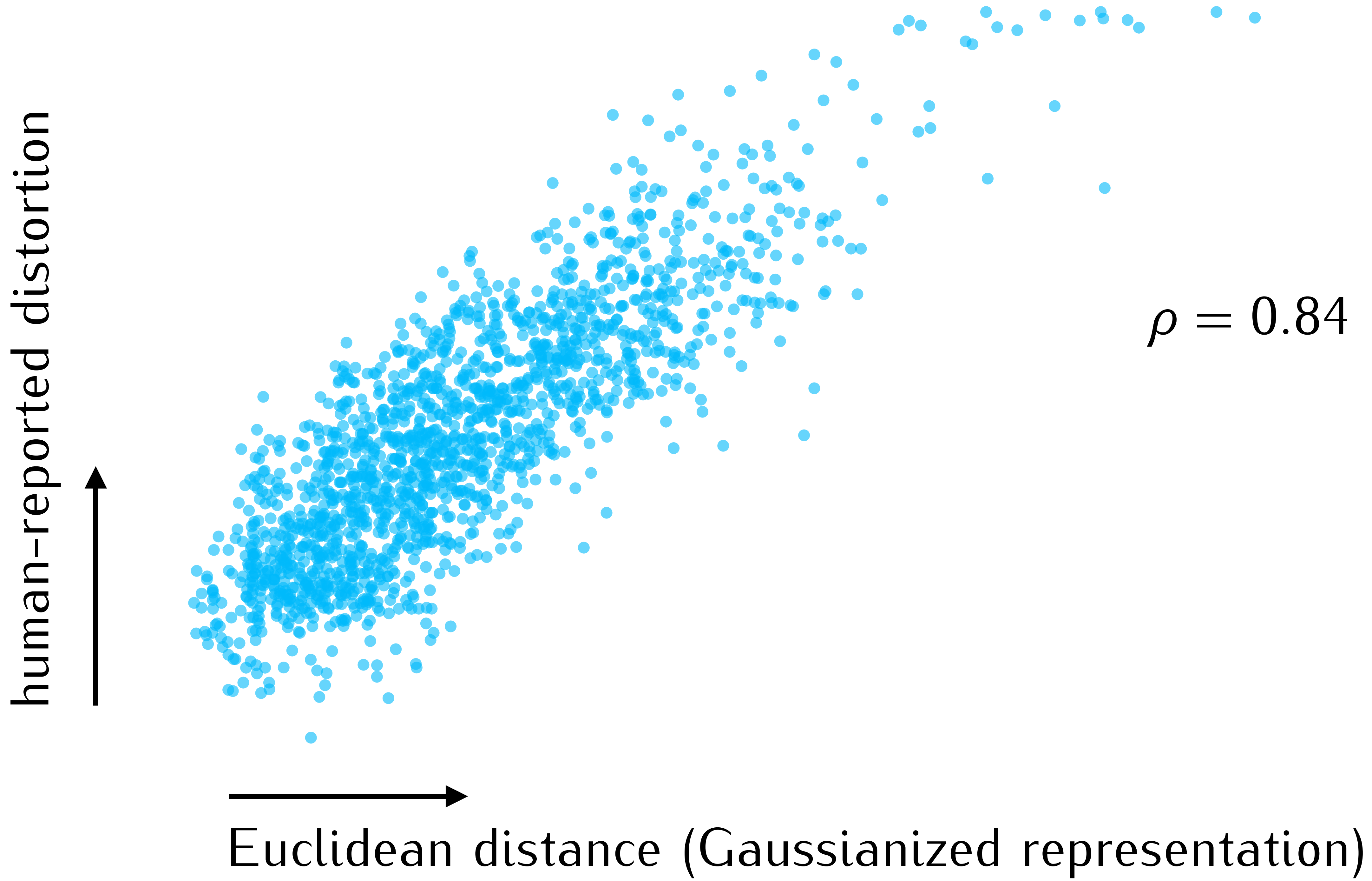

original
increasing
Euclidean
distance
in Gaussianized
representation

Pixel representation



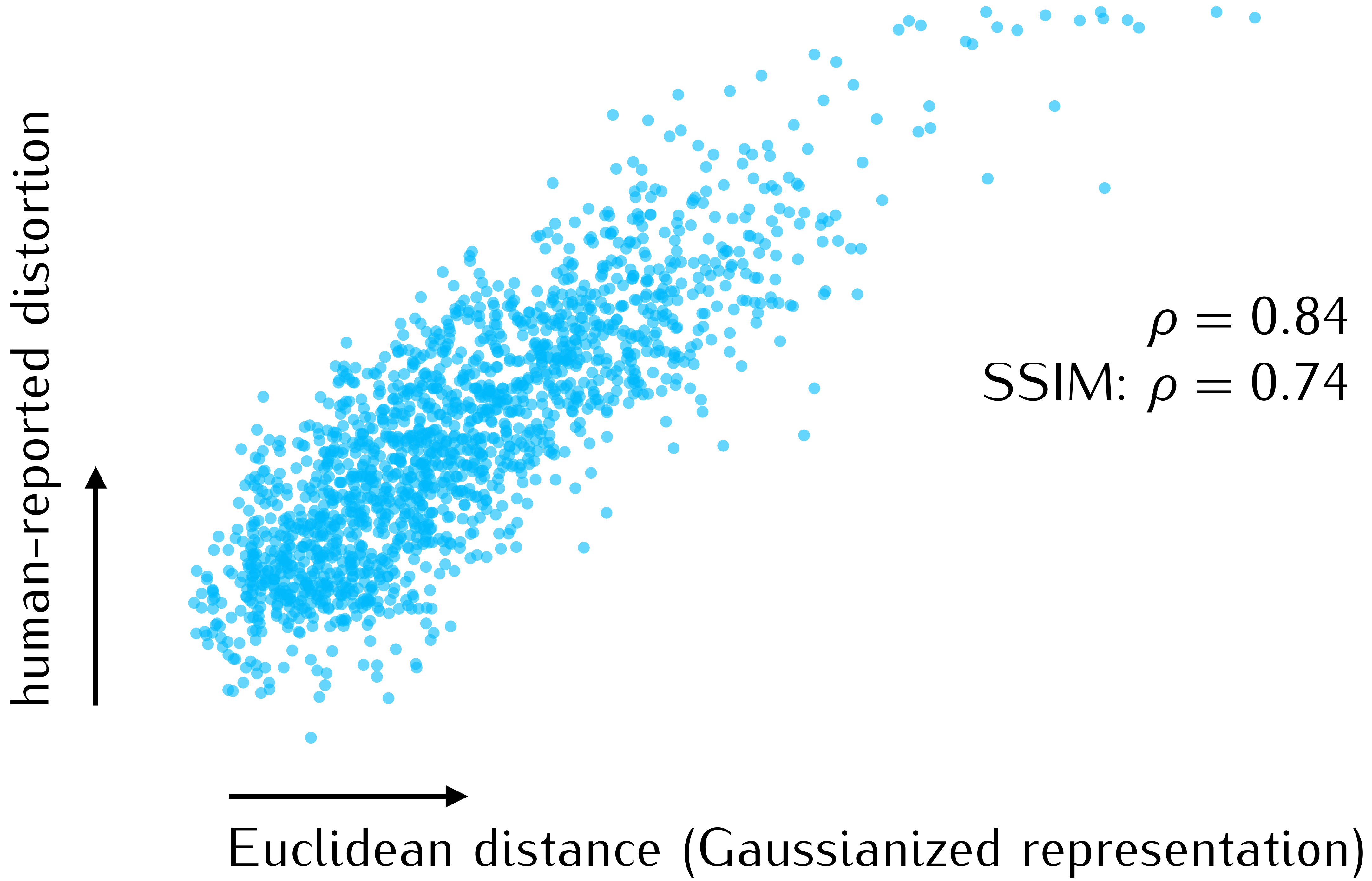
data: TID 2008

Multi-scale GDN representation



data: TID 2008

Multi-scale GDN representation



SSIM: Wang et al., 2004
data: TID 2008

- Gaussianization: Methodology for density estimation and unsupervised learning of a **representation**

- GDN: joint nonlinearity applied across feature maps
  – inspired by nonlinearities of biological neurons
  – generalizes sigmoids used in ANNs
  – capable of Gaussianizing image data

- one layer of GDN $>$ many layers of marginal nonlinearities

- accounts for human judgements of image quality (more so than SSIM, the de facto standard)

Thank you!