

# Closed-Form Information-Theoretic Roughness Measures for Mixture Densities

Uwe D. Hanebeck, **Daniel Frisch**, Dominik Prossel

Intelligent Sensor-Actuator-Systems Laboratory (ISAS)  
Institute for Anthropomatics and Robotics  
Karlsruhe Institute of Technology (KIT)  
Karlsruhe, Germany  
[isas.iar.kit.edu](http://isas.iar.kit.edu)



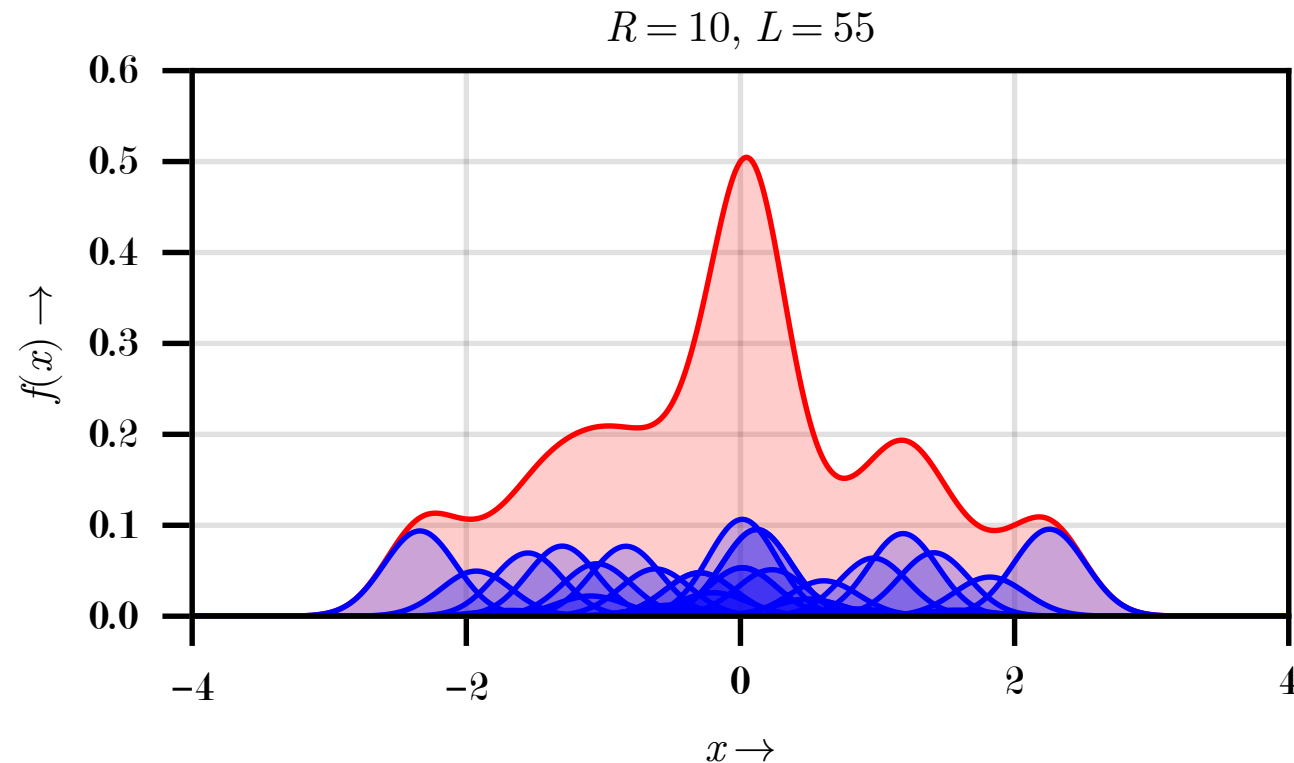
# Context & Motivation

# Context & Motivation

Many applications: Some specifications on probability density function given

**Goal:** Find cont. density that adds as little *information* as possible to specification

Example: Only single constraint  $f(0) = 0.5$

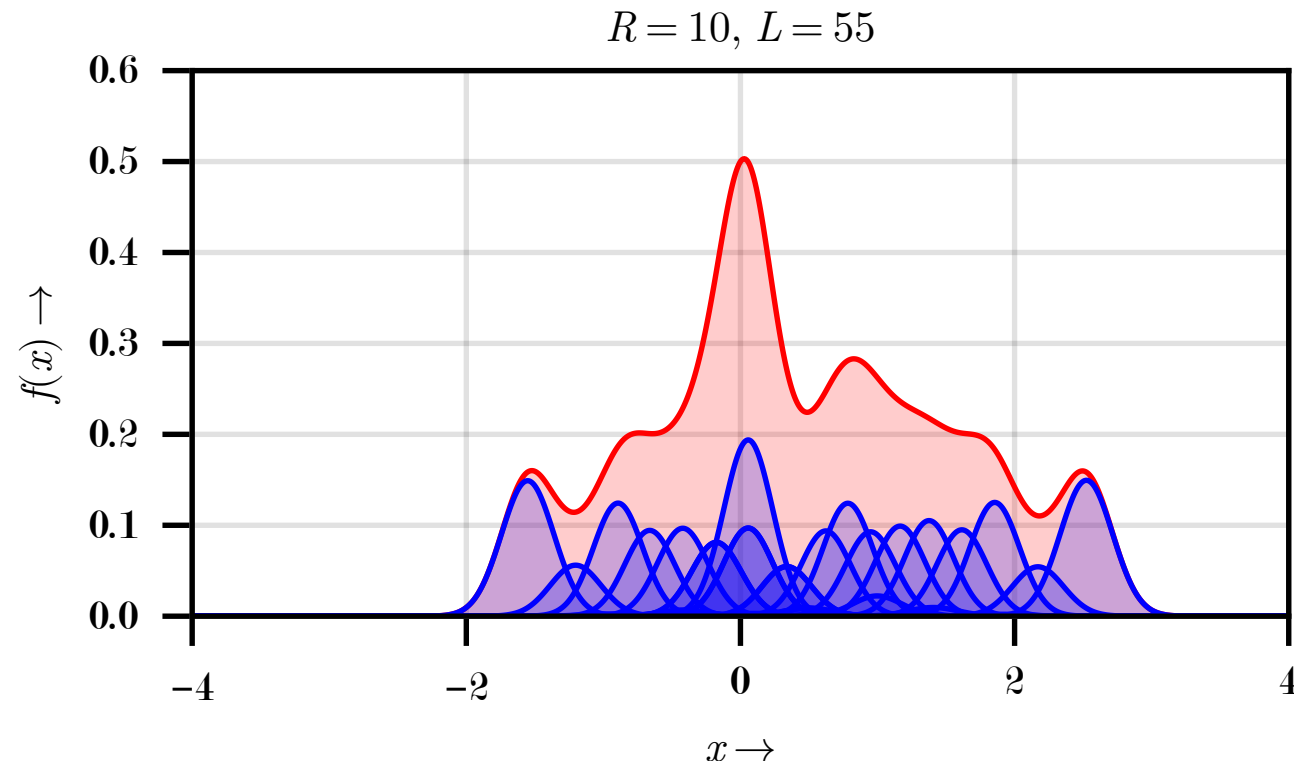


# Context & Motivation

Many applications: Some specifications on probability density function given

**Goal:** Find cont. density that adds as little *information* as possible to specification

Example: Only single constraint  $f(0) = 0.5$

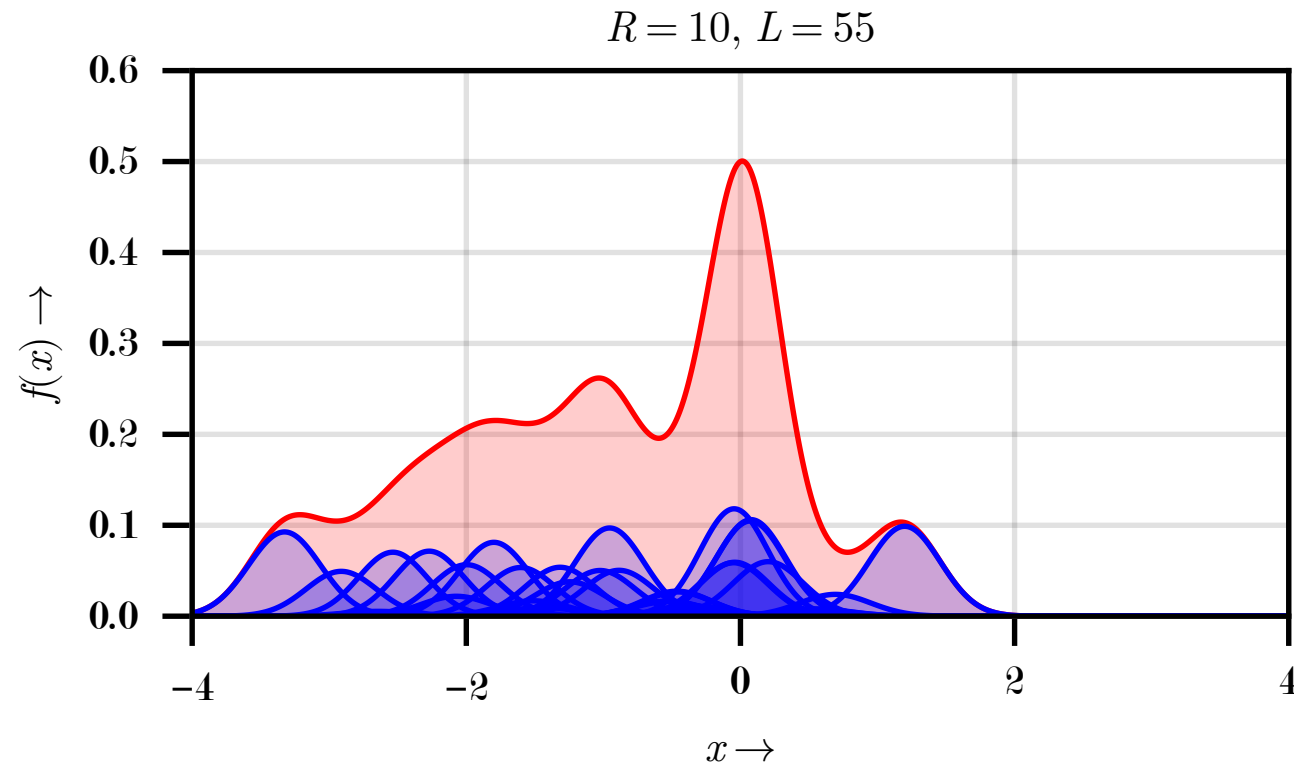


# Context & Motivation

Many applications: Some specifications on probability density function given

**Goal:** Find cont. density that adds as little *information* as possible to specification

Example: Only single constraint  $f(0) = 0.5$

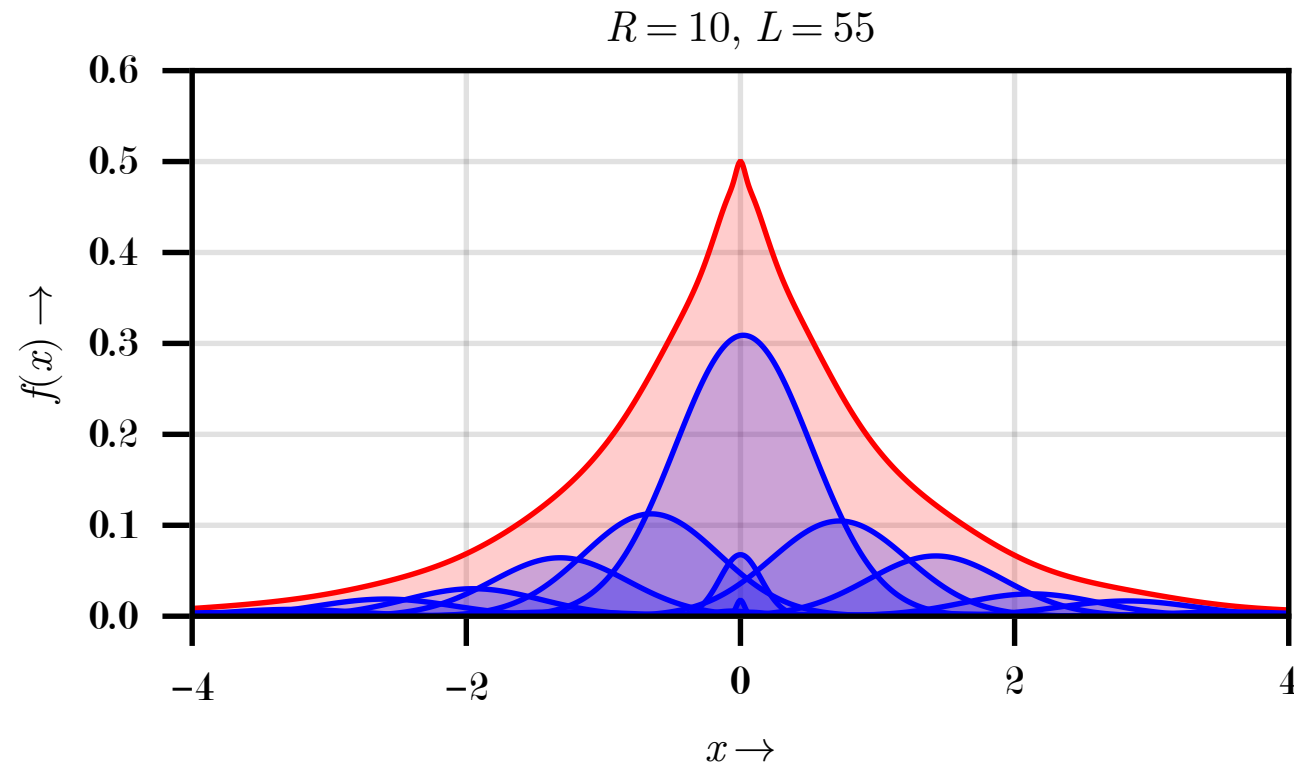


# Context & Motivation

Many applications: Some specifications on probability density function given

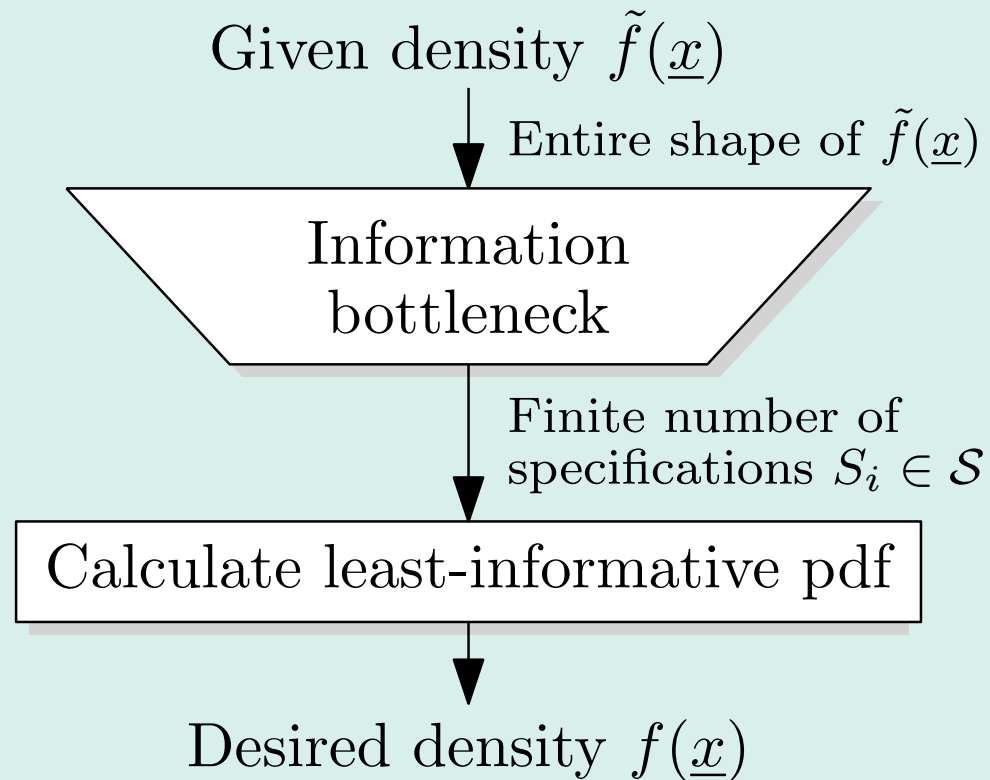
**Goal:** Find cont. density that adds as little *information* as possible to specification

Example: Only single constraint  $f(0) = 0.5$



General setup:

- (Virtual) underlying true density  $\tilde{f}(\underline{x})$
- Discrete set of specifications  $S_i$



# Challenges

Common to applications:

- Constraints do not fully characterize density
- Results in *underdetermined* optimization problem
- Some *regularizer* required

Common to applications:

- Constraints do not fully characterize density
- Results in *underdetermined* optimization problem
- Some *regularizer* required

Want to use information measure as regularizer

→ Find least-informative density under prescribed specifications

Common to applications:

- Constraints do not fully characterize density
- Results in *underdetermined* optimization problem
- Some *regularizer* required

Want to use information measure as regularizer

→ Find least-informative density under prescribed specifications

Least-informative density

- Exhibit certain smoothness
- No bumps and ripples

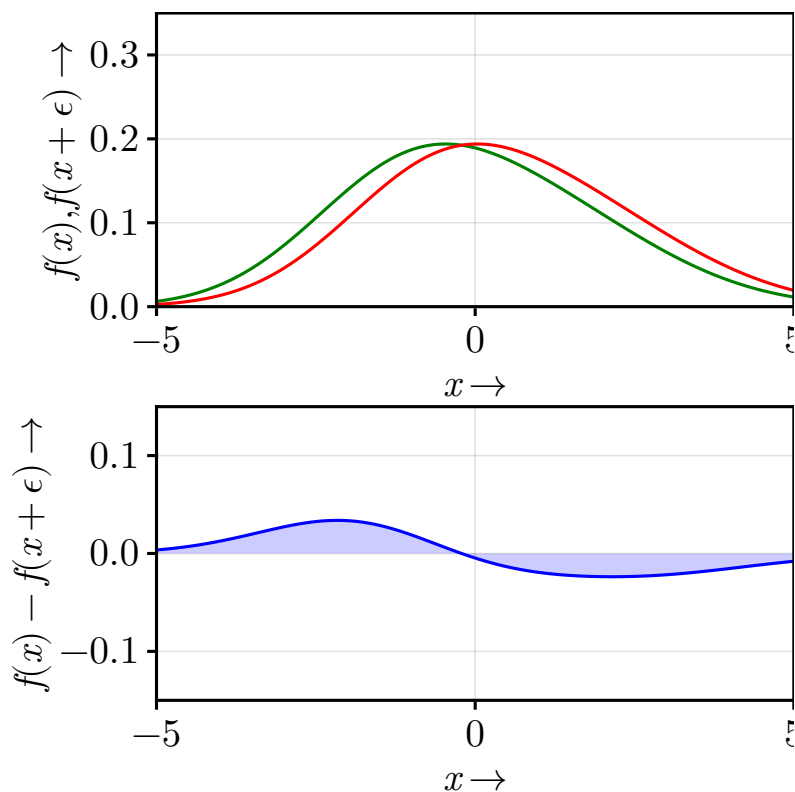
→ Roughness measure

# Derivation of Roughness Measure

# Roughness Measures - Intuition

Compare given density  $f(x)$  with its shifted copy  $f(x + \epsilon)$

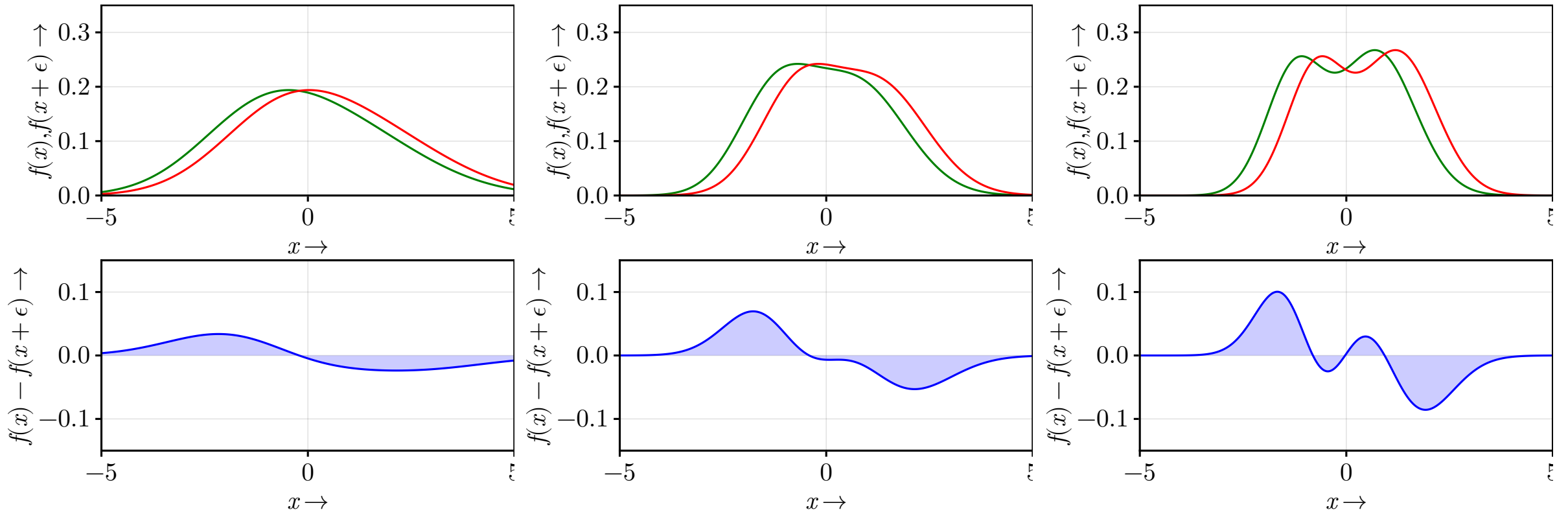
The smoother  $f(x)$ : Deviation between  $f(x)$  and  $f(x + \epsilon)$  smaller



# Roughness Measures - Intuition

Compare given density  $f(x)$  with its shifted copy  $f(x + \epsilon)$

The smoother  $f(x)$ : Deviation between  $f(x)$  and  $f(x + \epsilon)$  smaller



# Roughness Measures - Compare $f(x)$ and $f(x + \epsilon)$

Different measures for comparing  $f(x)$  and  $f(x + \epsilon)$

# Roughness Measures - Compare $f(x)$ and $f(x + \epsilon)$

Different measures for comparing  $f(x)$  and  $f(x + \epsilon)$

Typical (**global !**) measures for comparing two densities  $f(x)$  and  $g(x)$

Squared integral deviation

$$D_{\text{SID}}(f, g) = \int_{\mathbb{R}} (f(x) - g(x))^2 dx$$

# Roughness Measures - Compare $f(x)$ and $f(x + \epsilon)$

Different measures for comparing  $f(x)$  and  $f(x + \epsilon)$

Typical (**global !**) measures for comparing two densities  $f(x)$  and  $g(x)$

Squared integral deviation

$$D_{\text{SID}}(f, g) = \int_{\mathbb{R}} (f(x) - g(x))^2 dx$$

Weighted squared integral deviation

$$D_{\text{WSID}}(f, g) = \int_{\mathbb{R}} \frac{(f(x) - g(x))^2}{f(x)} dx$$

# Roughness Measures - Compare $f(x)$ and $f(x + \epsilon)$

Different measures for comparing  $f(x)$  and  $f(x + \epsilon)$

Typical (**global !**) measures for comparing two densities  $f(x)$  and  $g(x)$

Squared integral deviation

$$D_{\text{SID}}(f, g) = \int_{\mathbb{R}} (f(x) - g(x))^2 dx$$

Weighted squared integral deviation

$$D_{\text{WSID}}(f, g) = \int_{\mathbb{R}} \frac{(f(x) - g(x))^2}{f(x)} dx$$

Kullback-Leibler divergence

$$D_{\text{KL}}(f, g) = \int_{\mathbb{R}} f(x) \log \left( \frac{f(x)}{g(x)} \right) dx$$

# Roughness Measures - Compare $f(x)$ and $f(x + \epsilon)$

Different measures for comparing  $f(x)$  and  $f(x + \epsilon)$

Typical (**global !**) measures for comparing two densities  $f(x)$  and  $g(x)$

Squared integral deviation

$$D_{\text{SID}}(f, g) = \int_{\mathbb{R}} (f(x) - g(x))^2 dx$$

Weighted squared integral deviation

$$D_{\text{WSID}}(f, g) = \int_{\mathbb{R}} \frac{(f(x) - g(x))^2}{f(x)} dx$$

Kullback-Leibler divergence

$$D_{\text{KL}}(f, g) = \int_{\mathbb{R}} f(x) \log \left( \frac{f(x)}{g(x)} \right) dx$$

Hellinger distance

$$D_{\text{HELL}}(f, g) = \int_{\mathbb{R}} \left( \sqrt{f(x)} - \sqrt{g(x)} \right)^2 dx$$

$$D(\epsilon) = D_{\text{WSID}}(f(x), f(x + \epsilon)) = \int_{\mathbb{R}} \frac{(f(x) - f(x + \epsilon))^2}{2 f(x)} dx$$

$$D(\epsilon) = D_{\text{WSID}}(f(x), f(x + \epsilon)) = \int_{\mathbb{R}} \frac{(f(x) - f(x + \epsilon))^2}{2 f(x)} dx$$

For function  $h(\epsilon): \mathbb{R} \rightarrow \mathbb{R}$ , Taylor series expansion around  $\epsilon_0$

$$h(\epsilon) \approx h(\epsilon_0) + \frac{dh(\epsilon)}{d\epsilon} \cdot (\epsilon - \epsilon_0) + \frac{d^2h(\epsilon)}{d\epsilon^2} \cdot (\epsilon - \epsilon_0)^2$$

$$D(\epsilon) = D_{\text{WSID}}(f(x), f(x + \epsilon)) = \int_{\mathbb{R}} \frac{(f(x) - f(x + \epsilon))^2}{2f(x)} dx$$

For function  $h(\epsilon): \mathbb{R} \rightarrow \mathbb{R}$ , Taylor series expansion around  $\epsilon_0$

$$h(\epsilon) \approx h(\epsilon_0) + \frac{dh(\epsilon)}{d\epsilon} \cdot (\epsilon - \epsilon_0) + \frac{d^2h(\epsilon)}{d\epsilon^2} \cdot (\epsilon - \epsilon_0)^2$$

Taylor series of  $D(\epsilon)$  around  $\epsilon_0 = 0$  (assume regularity)

$$D(\epsilon) \approx D(0) + \left. \frac{dD(\epsilon)}{d\epsilon} \right|_{\epsilon=0} \cdot \epsilon + \left. \frac{d^2D(\epsilon)}{d\epsilon^2} \right|_{\epsilon=0} \cdot \epsilon^2 = \int_{\mathbb{R}} \frac{[f'(x)]^2}{f(x)} dx \cdot \epsilon^2$$

$$D(\epsilon) = D_{\text{KL}}(f(x), f(x + \epsilon)) = \int_{\mathbb{R}} f(x) \log \left( \frac{f(x)}{f(x + \epsilon)} \right) dx$$

$$D(\epsilon) = D_{\text{KL}}(f(x), f(x + \epsilon)) = \int_{\mathbb{R}} f(x) \log \left( \frac{f(x)}{f(x + \epsilon)} \right) dx$$

Taylor series of  $D(\epsilon)$  around  $\epsilon_0 = 0$  (assume regularity)

$$\begin{aligned} D(\epsilon) &\approx D(0) + \left. \frac{dD(\epsilon)}{d\epsilon} \right|_{\epsilon=0} \cdot \epsilon + \left. \frac{d^2 D(\epsilon)}{d\epsilon^2} \right|_{\epsilon=0} \cdot \epsilon^2 \\ &= \underbrace{\int_{\mathbb{R}} f'(x) dx}_0 \cdot \epsilon + \int_{\mathbb{R}} \left\{ \underbrace{\frac{[f'(x)]^2}{f(x)} - f''(x)}_0 \right\} dx \cdot \epsilon^2 \end{aligned}$$

$$= \int_{\mathbb{R}} \frac{[f'(x)]^2}{f(x)} dx \cdot \epsilon^2$$

$$D(\epsilon) = D_{\text{HELL}}(f(x), f(x + \epsilon)) = \int_{\mathbb{R}} \left( \sqrt{f(x)} - \sqrt{f(x + \epsilon)} \right)^2 dx$$

$$D(\epsilon) = D_{\text{HELL}}(f(x), f(x + \epsilon)) = \int_{\mathbb{R}} \left( \sqrt{f(x)} - \sqrt{f(x + \epsilon)} \right)^2 dx$$

Taylor series of  $D(\epsilon)$  around  $\epsilon_0 = 0$  (assume regularity)

$$D(\epsilon) \approx D(0) + \left. \frac{dD(\epsilon)}{d\epsilon} \right|_{\epsilon=0} \cdot \epsilon + \left. \frac{d^2 D(\epsilon)}{d\epsilon^2} \right|_{\epsilon=0} \cdot \epsilon^2 = \int_{\mathbb{R}} \frac{[f'(x)]^2}{f(x)} dx \cdot \epsilon^2$$

$$D(\epsilon) = D_{\text{HELL}}(f(x), f(x + \epsilon)) = \int_{\mathbb{R}} \left( \sqrt{f(x)} - \sqrt{f(x + \epsilon)} \right)^2 dx$$

Taylor series of  $D(\epsilon)$  around  $\epsilon_0 = 0$  (assume regularity)

$$D(\epsilon) \approx D(0) + \left. \frac{dD(\epsilon)}{d\epsilon} \right|_{\epsilon=0} \cdot \epsilon + \left. \frac{d^2 D(\epsilon)}{d\epsilon^2} \right|_{\epsilon=0} \cdot \epsilon^2 = \int_{\mathbb{R}} \frac{[f'(x)]^2}{f(x)} dx \cdot \epsilon^2$$

$$D(\epsilon) = D_{\text{HELL}}(f(x), f(x + \epsilon)) = \int_{\mathbb{R}} (r(x) - r(x + \epsilon))^2 dx$$

$$D(\epsilon) \approx D(0) + \left. \frac{dD(\epsilon)}{d\epsilon} \right|_{\epsilon=0} \cdot \epsilon + \left. \frac{d^2 D(\epsilon)}{d\epsilon^2} \right|_{\epsilon=0} \cdot \epsilon^2 = \int_{\mathbb{R}} [r'(x)]^2 dx \cdot \epsilon^2$$

# 2nd Derivation Based on Fisher Information

For **location parameter**  $\theta$ , the likelihood is given by  $f(x|\theta) = f(x + \theta)$ , the score is

$$s(x|\theta) = \frac{\partial}{\partial \theta} \log (f(x + \theta)) = \frac{\partial f(x + \theta)}{\partial \theta} \frac{1}{f(x + \theta)} = \frac{\partial f(x + \theta)}{\partial x} \frac{1}{f(x + \theta)}$$

For **location parameter**  $\theta$ , the likelihood is given by  $f(x|\theta) = f(x + \theta)$ , the score is

$$s(x|\theta) = \frac{\partial}{\partial \theta} \log (f(x + \theta)) = \frac{\partial f(x + \theta)}{\partial \theta} \frac{1}{f(x + \theta)} = \frac{\partial f(x + \theta)}{\partial x} \frac{1}{f(x + \theta)}$$

The Fisher information can be rewritten as

$$I_F(\theta) = \int_{\mathbb{R}} \left[ \frac{\partial f(x + \theta)}{\partial x} \frac{1}{f(x + \theta)} \right]^2 f(x + \theta) dx = \int_{\mathbb{R}} \left[ \frac{\partial f(x + \theta)}{\partial x} \right]^2 \frac{1}{f(x + \theta)} dx$$

For **location parameter**  $\theta$ , the likelihood is given by  $f(x|\theta) = f(x + \theta)$ , the score is

$$s(x|\theta) = \frac{\partial}{\partial \theta} \log (f(x + \theta)) = \frac{\partial f(x + \theta)}{\partial \theta} \frac{1}{f(x + \theta)} = \frac{\partial f(x + \theta)}{\partial x} \frac{1}{f(x + \theta)}$$

The Fisher information can be rewritten as

$$I_F(\theta) = \int_{\mathbb{R}} \left[ \frac{\partial f(x + \theta)}{\partial x} \frac{1}{f(x + \theta)} \right]^2 f(x + \theta) dx = \int_{\mathbb{R}} \left[ \frac{\partial f(x + \theta)}{\partial x} \right]^2 \frac{1}{f(x + \theta)} dx$$

As  $I_F(\theta)$  does not depend on  $\theta$  anymore, we have

$$I_F = \int_{\mathbb{R}} \left[ \frac{df(x)}{dx} \right]^2 \frac{1}{f(x)} dx = \int_{\mathbb{R}} \frac{[f'(x)]^2}{f(x)} dx$$

# Review: Fisher Information with Respect to Parameter

For **location parameter**  $\theta$ , the likelihood is given by  $f(x|\theta) = f(x + \theta)$ , the score is

$$s(x|\theta) = \frac{\partial}{\partial \theta} \log (f(x + \theta)) = \frac{\partial f(x + \theta)}{\partial \theta} \frac{1}{f(x + \theta)} = \frac{\partial f(x + \theta)}{\partial x} \frac{1}{f(x + \theta)}$$

The Fisher information can be rewritten as

$$I_F(\theta) = \int_{\mathbb{R}} \left[ \frac{\partial f(x + \theta)}{\partial x} \frac{1}{f(x + \theta)} \right]^2 f(x + \theta) dx = \int_{\mathbb{R}} \left[ \frac{\partial f(x + \theta)}{\partial x} \right]^2 \frac{1}{f(x + \theta)} dx$$

As  $I_F(\theta)$  does not depend on  $\theta$  anymore, we have

$$I_F = \int_{\mathbb{R}} \left[ \frac{df(x)}{dx} \right]^2 \frac{1}{f(x)} dx = \int_{\mathbb{R}} \frac{[f'(x)]^2}{f(x)} dx$$

That looks familiar!

All derivations give **Fisher Information Number (FIN)**

$$I_F = \int_{\mathbb{R}} \frac{[f'(x)]^2}{f(x)} dx$$

All derivations give **Fisher Information Number (FIN)**

$$I_F = \int_{\mathbb{R}} \frac{[f'(x)]^2}{f(x)} dx$$

penalizes changes



All derivations give **Fisher Information Number (FIN)**

$$I_F = \int_{\mathbb{R}} \frac{[f'(x)]^2}{f(x)} dx$$

penalizes changes  
more sensitive  
for small values

All derivations give **Fisher Information Number (FIN)**

$$I_F = \int_{\mathbb{R}} \frac{[f'(x)]^2}{f(x)} dx$$

penalizes changes  
more sensitive  
for small values

Generalization to  $D$  dimensions

$$I_F^{\mathcal{M}}(f) = \int_{\{\mathbb{R}^D, f > 0\}} \frac{|\nabla f(\underline{x})|^2}{f(\underline{x})} d\underline{x}$$

# Root Mixtures

Define square root mixtures as

$$r(\underline{x}) = \sum_{i=1}^R v_i r_i(\underline{x}) \quad v_i \text{ arbitrary for now}$$

Define square root mixtures as

$$r(\underline{x}) = \sum_{i=1}^R v_i r_i(\underline{x}) \quad v_i \text{ arbitrary for now}$$

Conversion of square root mixture to mixture in original mixture space

$$f(\underline{x}) = r^2(\underline{x})$$

Define square root mixtures as

$$r(\underline{x}) = \sum_{i=1}^R v_i r_i(\underline{x}) \quad v_i \text{ arbitrary for now}$$

Conversion of square root mixture to mixture in original mixture space

$$\begin{aligned} f(\underline{x}) &= r^2(\underline{x}) \\ &= \left( \sum_{i=1}^R v_i r_i(\underline{x}) \right) \left( \sum_{i=1}^R v_i r_i(\underline{x}) \right) \end{aligned}$$

Define square root mixtures as

$$r(\underline{x}) = \sum_{i=1}^R v_i r_i(\underline{x}) \quad v_i \text{ arbitrary for now}$$

Conversion of square root mixture to mixture in original mixture space

$$\begin{aligned} f(\underline{x}) &= r^2(\underline{x}) \\ &= \left( \sum_{i=1}^R v_i r_i(\underline{x}) \right) \left( \sum_{i=1}^R v_i r_i(\underline{x}) \right) \\ &= \sum_{i=1}^R \sum_{j=1}^R v_i v_j r_i(\underline{x}) r_j(\underline{x}) \end{aligned}$$

Define square root mixtures as

$$r(\underline{x}) = \sum_{i=1}^R v_i r_i(\underline{x}) \quad v_i \text{ arbitrary for now}$$

Conversion of square root mixture to mixture in original mixture space

$$\begin{aligned} f(\underline{x}) &= r^2(\underline{x}) \\ &= \left( \sum_{i=1}^R v_i r_i(\underline{x}) \right) \left( \sum_{i=1}^R v_i r_i(\underline{x}) \right) \\ &= \sum_{i=1}^R \sum_{j=1}^R v_i v_j r_i(\underline{x}) r_j(\underline{x}) \end{aligned}$$

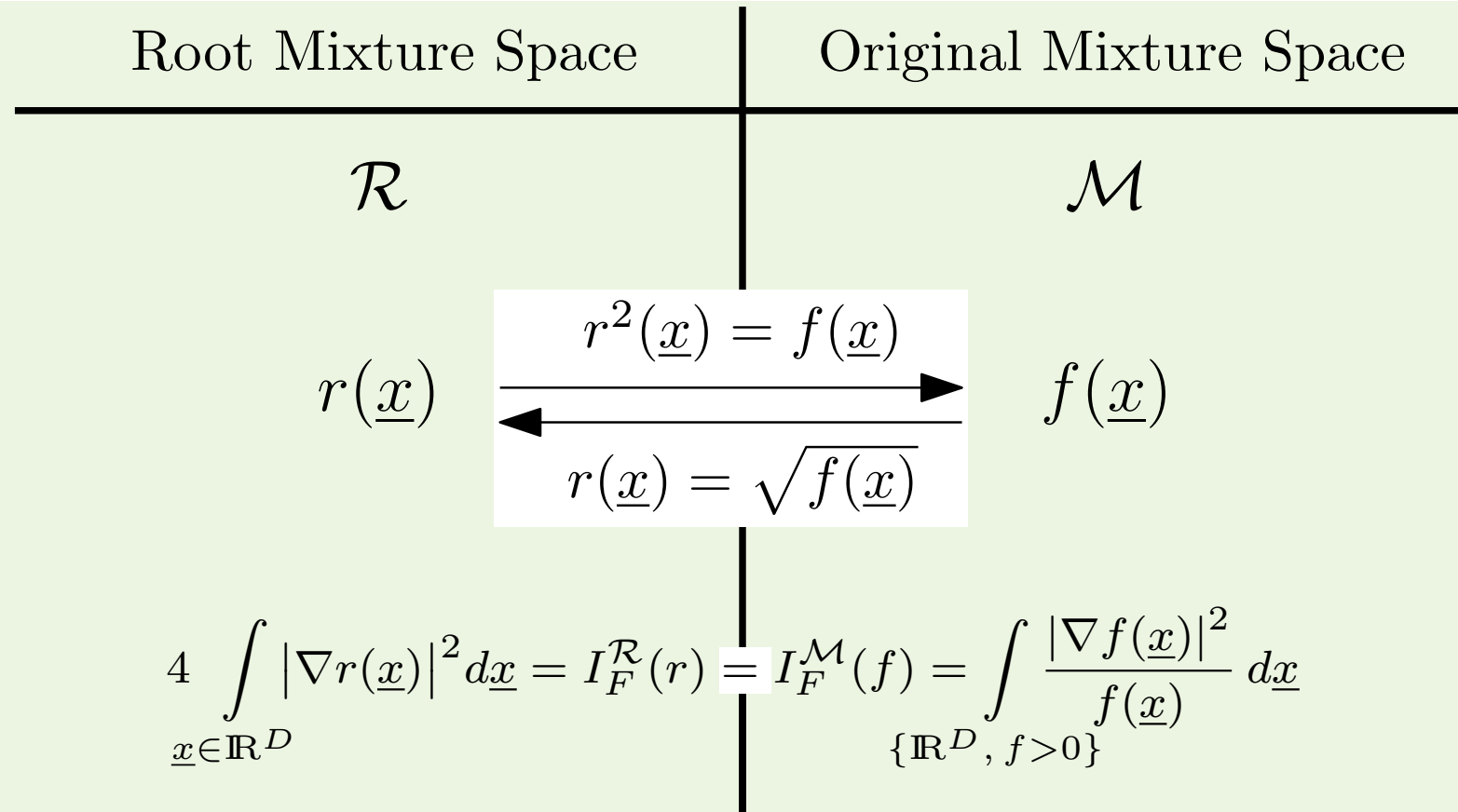
Expression contains redundant terms as  $r_i(\underline{x}) r_j(\underline{x}) = r_j(\underline{x}) r_i(\underline{x})$  for  $i \neq j$

# Fisher Information Number - Challenges & Key Idea

# Fisher Information Number - Challenges & Key Idea

Define two density spaces:

- Original space of mixture densities  $\mathcal{M}$
- Space of root mixture densities  $\mathcal{R}$



# Optimization

Given:

- Set of specifications  $S_i(f) = 0$ ,  $S_i \in \mathcal{S}$  on  $f(\underline{x})$
- Basic constraints on  $f(\underline{x})$ , i.e., unit integral constraint

Desired:

- Find parameters of least-informative root mixture under constraints

Given:

- Set of specifications  $S_i(f) = 0$ ,  $S_i \in \mathcal{S}$  on  $f(\underline{x})$
- Basic constraints on  $f(\underline{x})$ , i.e., unit integral constraint

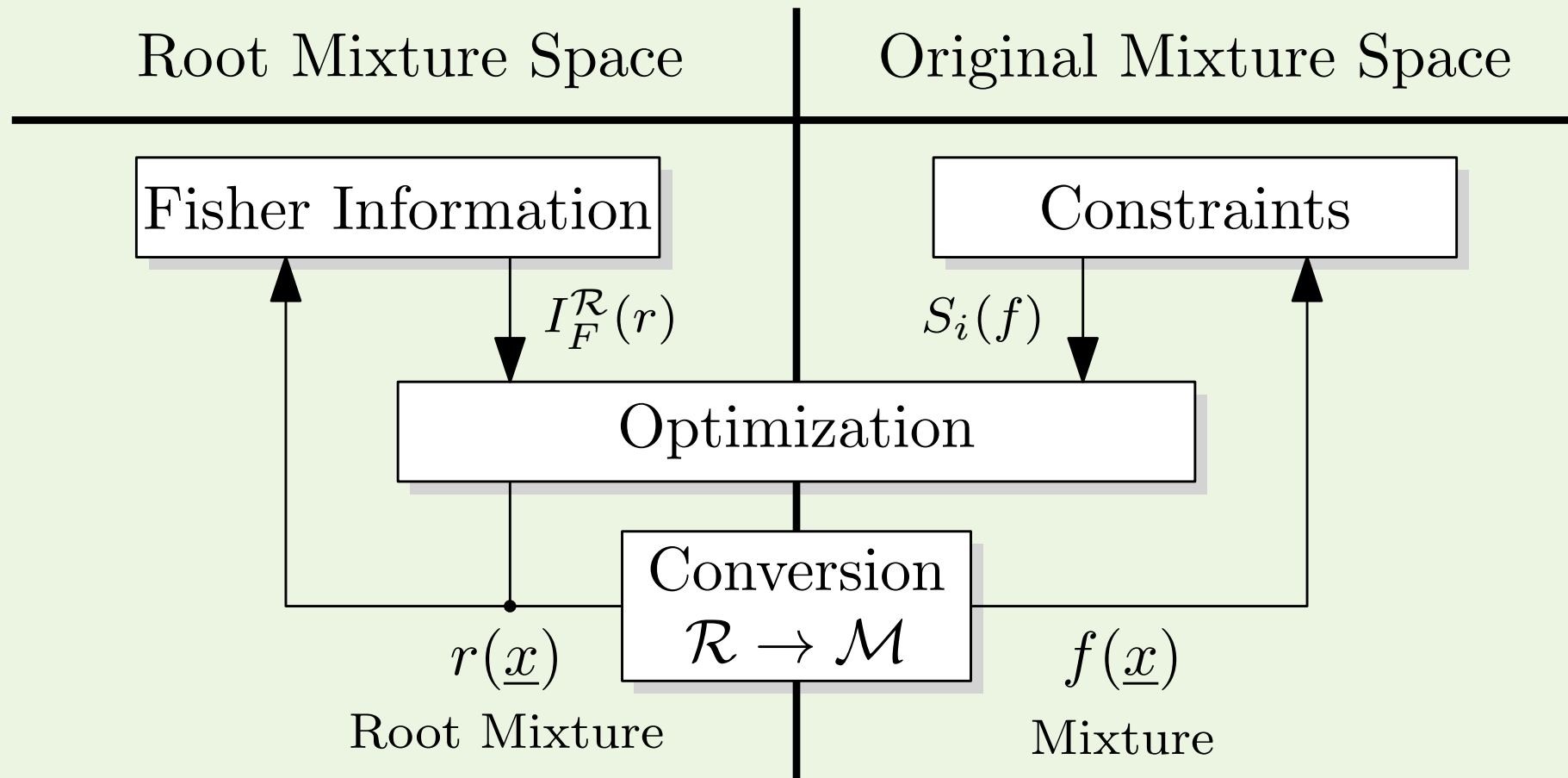
Desired:

- Find parameters of least-informative root mixture under constraints

Optimization problem:

$$\begin{aligned} \min_{r \in \mathcal{R}} \quad & I_F^{\mathcal{R}}(r) \\ \text{s.t.} \quad & f = r^2 \\ & S_i(f) = 0, S_i \in \mathcal{S} \end{aligned}$$

Tandem processing in root mixture space space  $\mathcal{R}$  and original mixture space  $\mathcal{M}$

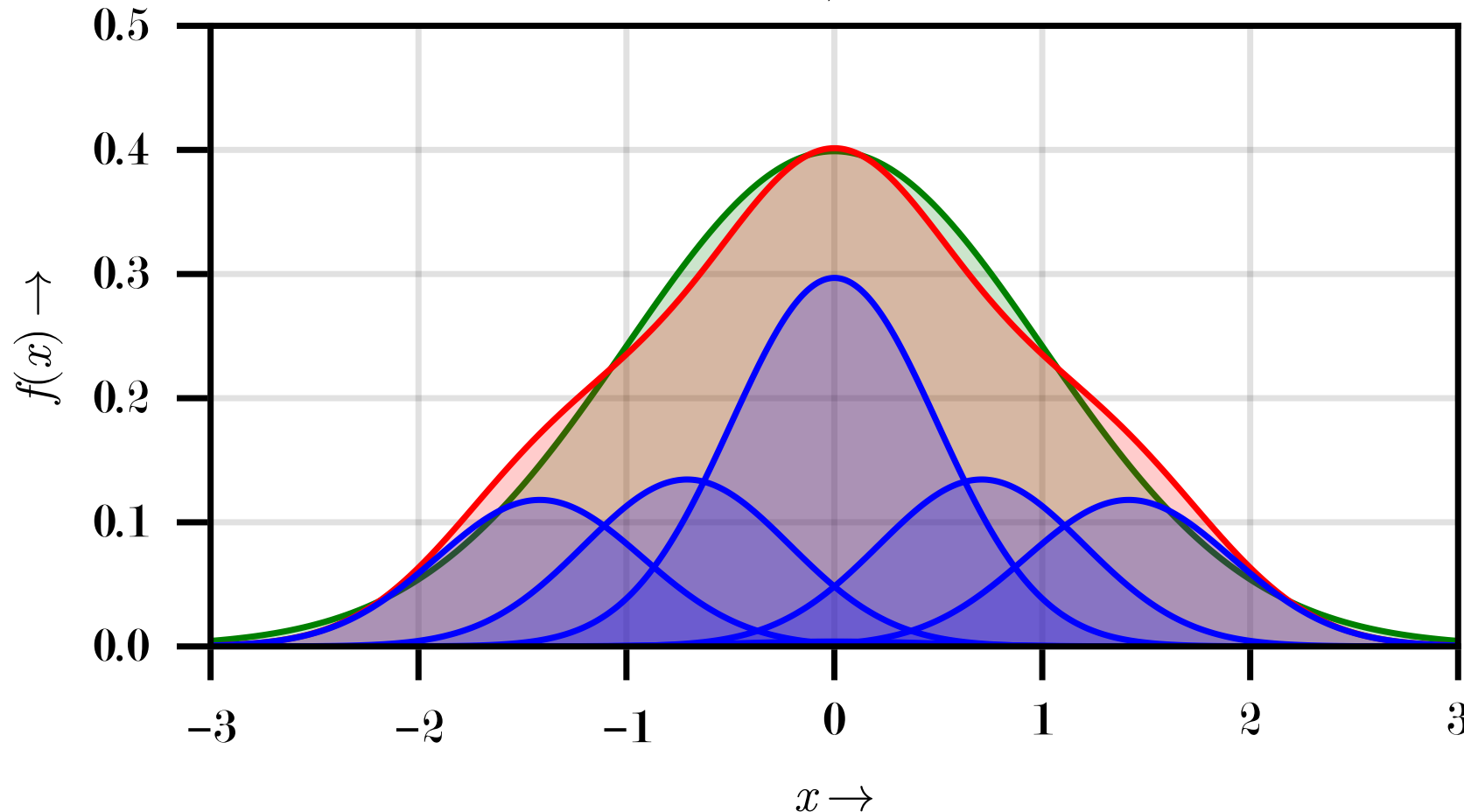


# Examples

# Example - Gaussian Mixture with zero mean, unit variance

$R \in \{3, 4, 5\}$  root mixture components,  $L = R \cdot (R + 1)/2 \in \{6, 10, 15\}$  mixture components

$R = 3, L = 6$

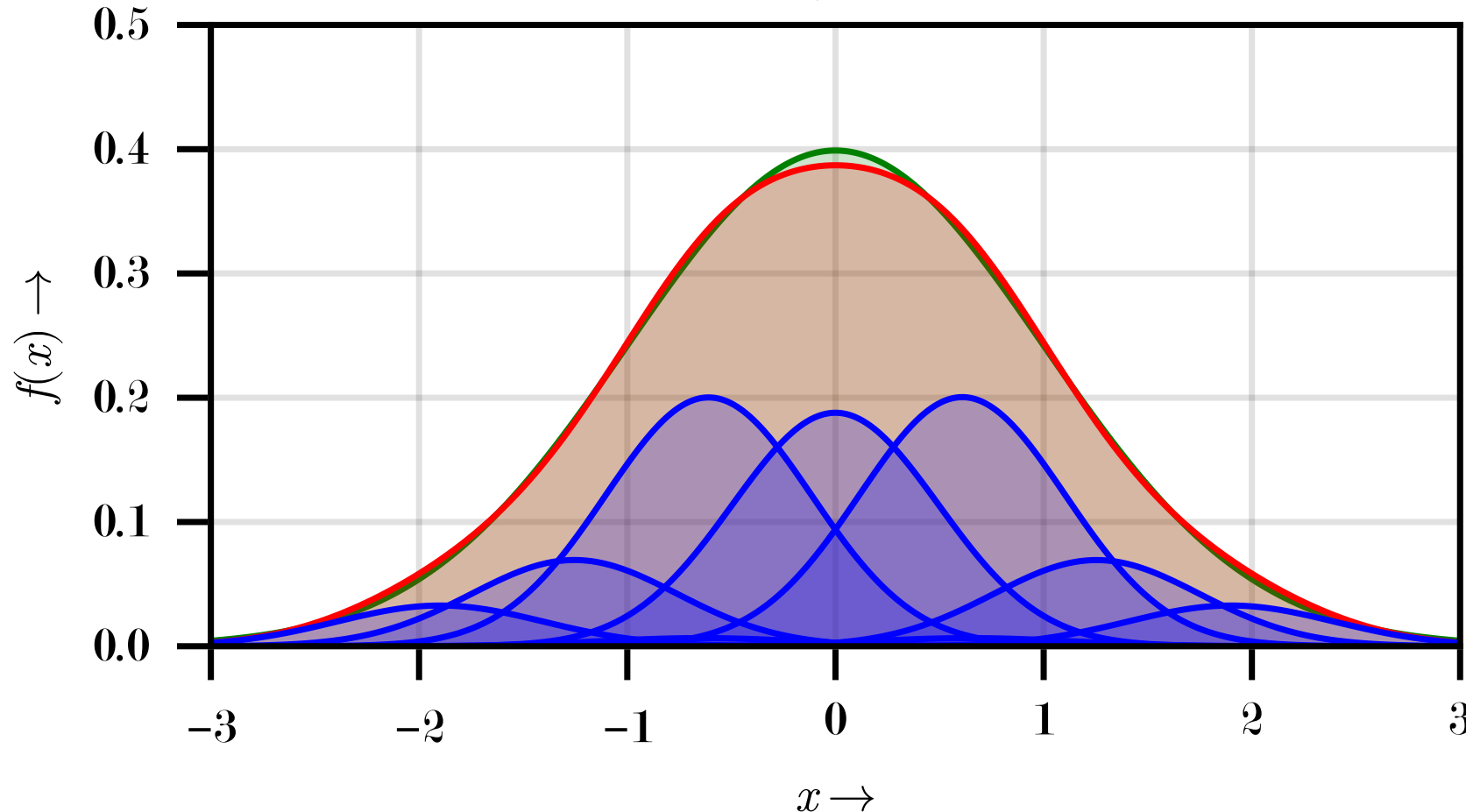


avoid trivial solution  
by bounding variance  
of components

# Example - Gaussian Mixture with zero mean, unit variance

$R \in \{3, 4, 5\}$  root mixture components,  $L = R \cdot (R + 1)/2 \in \{6, 10, 15\}$  mixture components

$R = 4, L = 10$

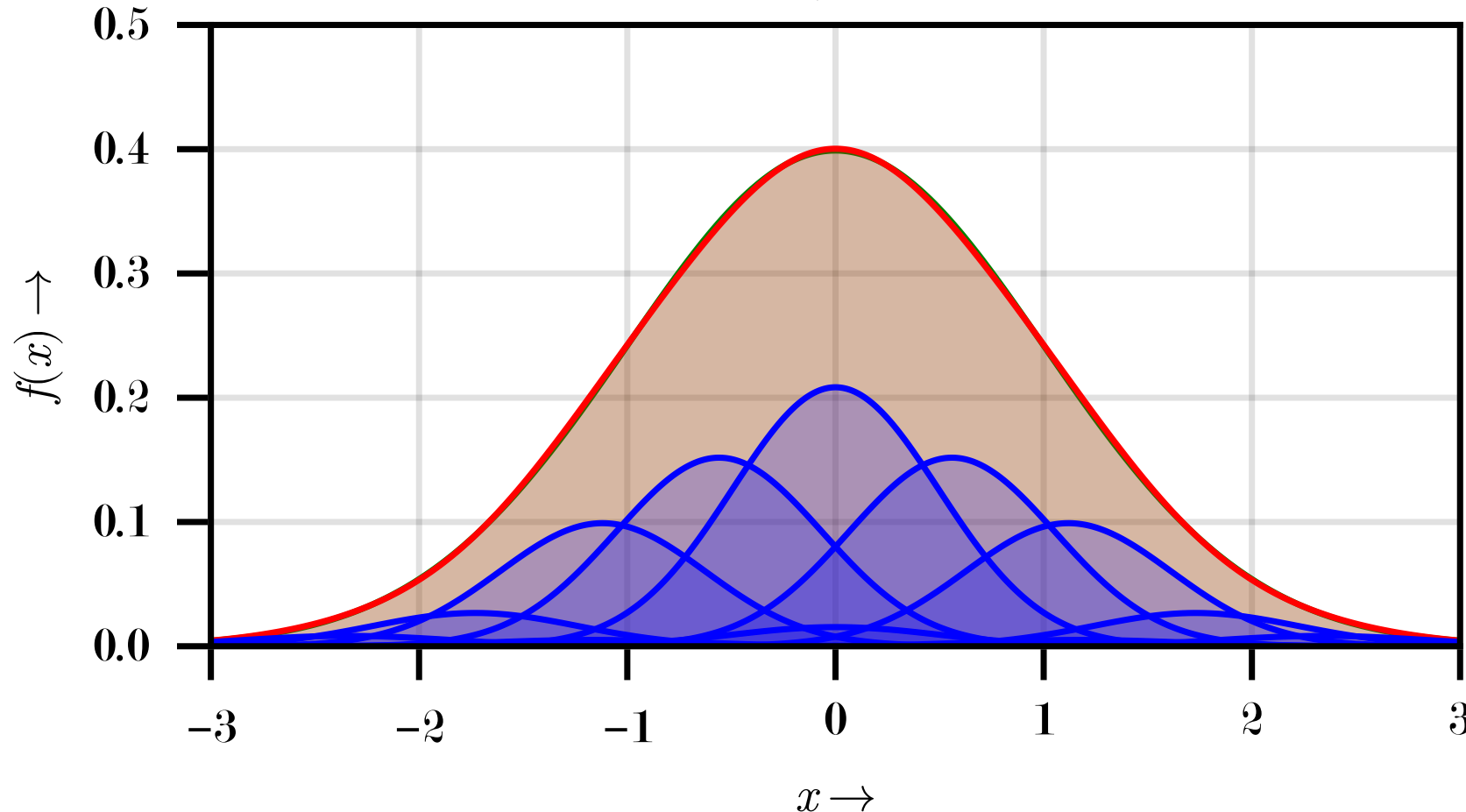


avoid trivial solution  
by bounding variance  
of components

# Example - Gaussian Mixture with zero mean, unit variance

$R \in \{3, 4, 5\}$  root mixture components,  $L = R \cdot (R + 1)/2 \in \{6, 10, 15\}$  mixture components

$R = 5, L = 15$

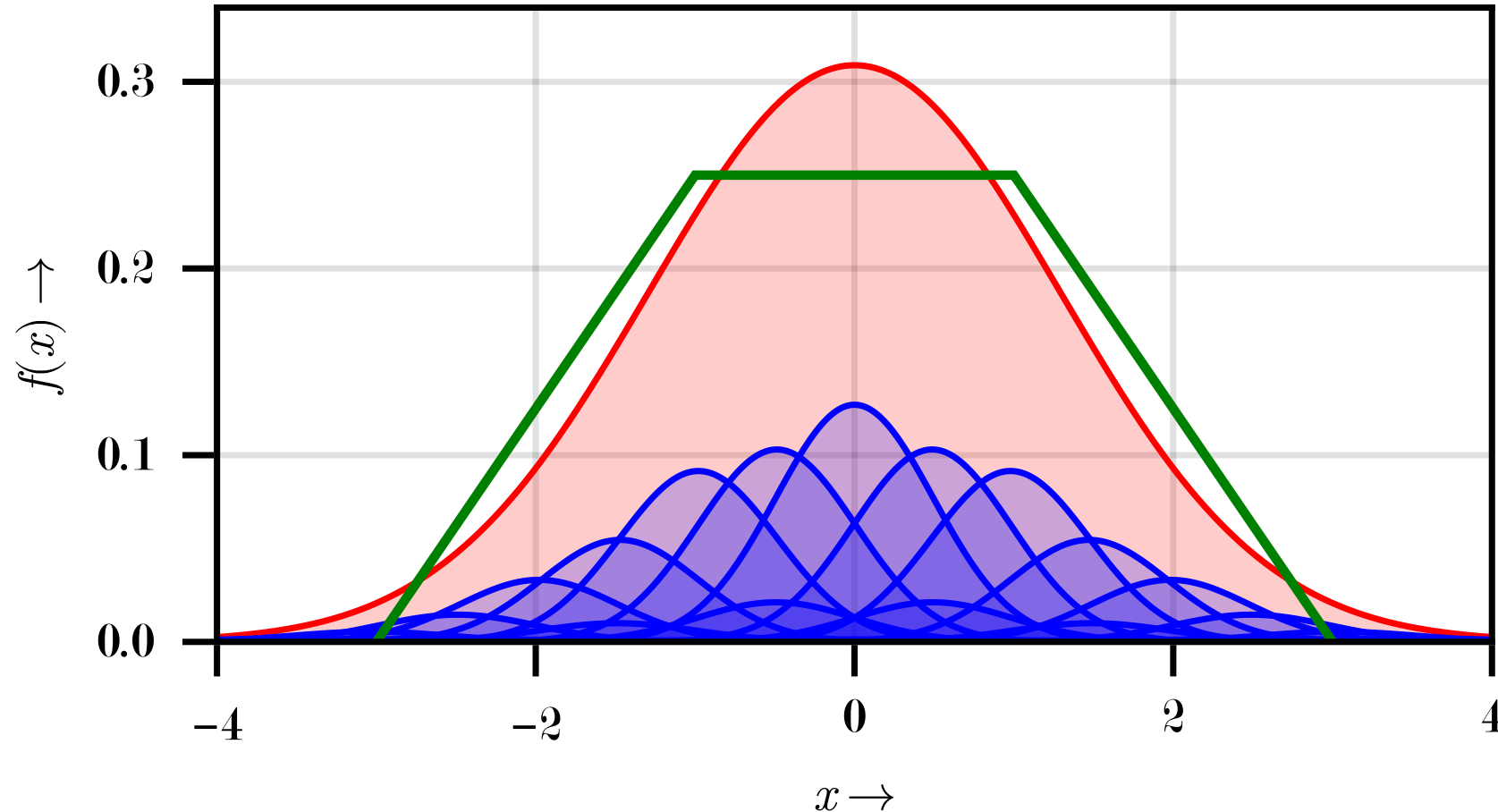


avoid trivial solution  
by bounding variance  
of components

# Example - Gaussian Mixture with Higher-Order Moments

$R = 10$  root mixture components,  $L = R \cdot (R + 1)/2 = 55$  mixture components

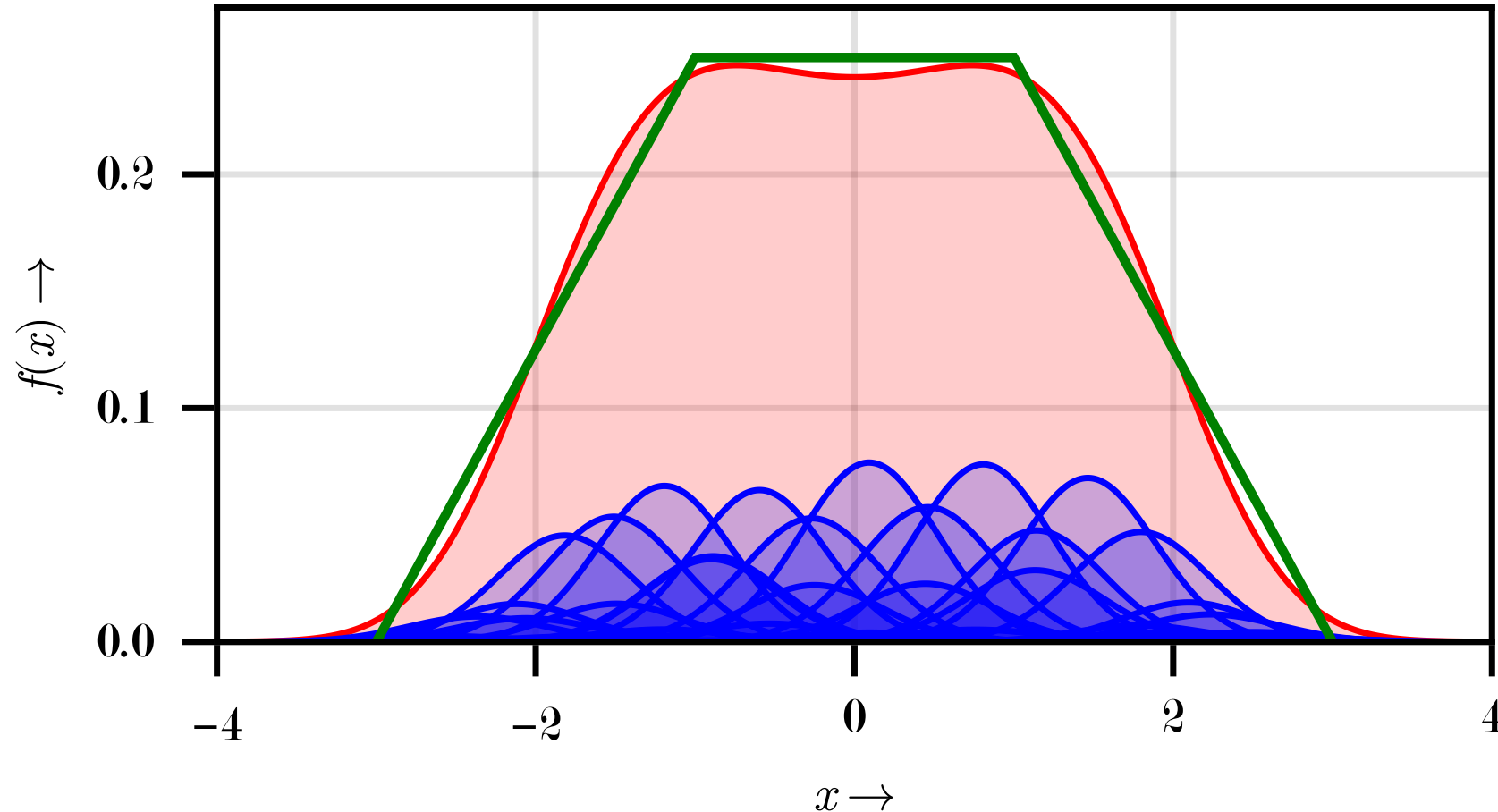
$R = 10, L = 55, \text{order} = 4$



# Example - Gaussian Mixture with Higher-Order Moments

$R = 10$  root mixture components,  $L = R \cdot (R + 1)/2 = 55$  mixture components

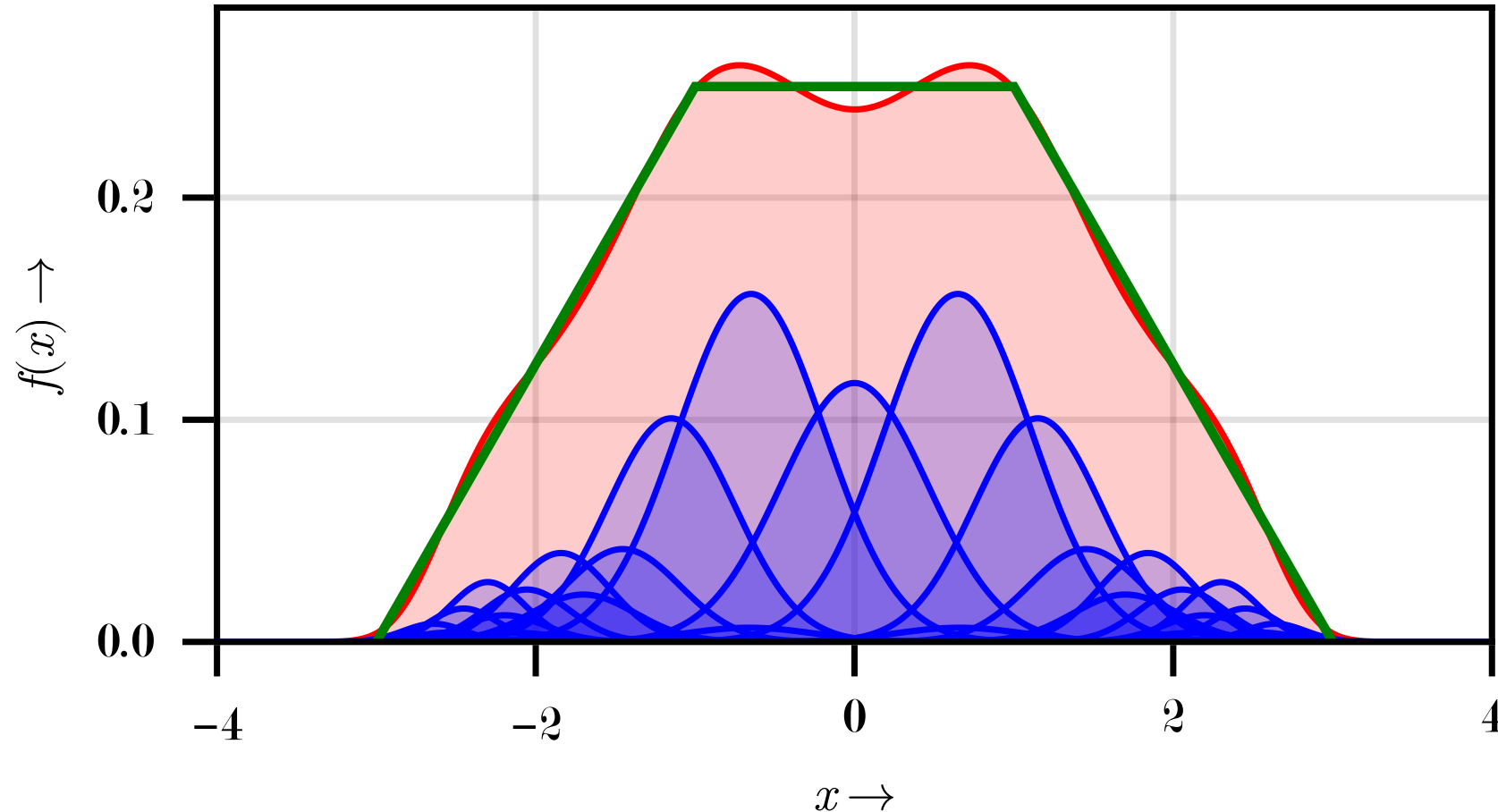
$R = 10, L = 55, \text{order} = 6$



# Example - Gaussian Mixture with Higher-Order Moments

$R = 10$  root mixture components,  $L = R \cdot (R + 1)/2 = 55$  mixture components

$R = 10, L = 55, \text{order} = 10$



# Conclusion & Outlook

## Fisher Information Number

Fisher Information Number



Closed-form solution for root mixtures

Fisher Information Number



Closed-form solution for root mixtures



Versatile tool for optimal reconstruction  
of continuous densities  
from sparse specifications