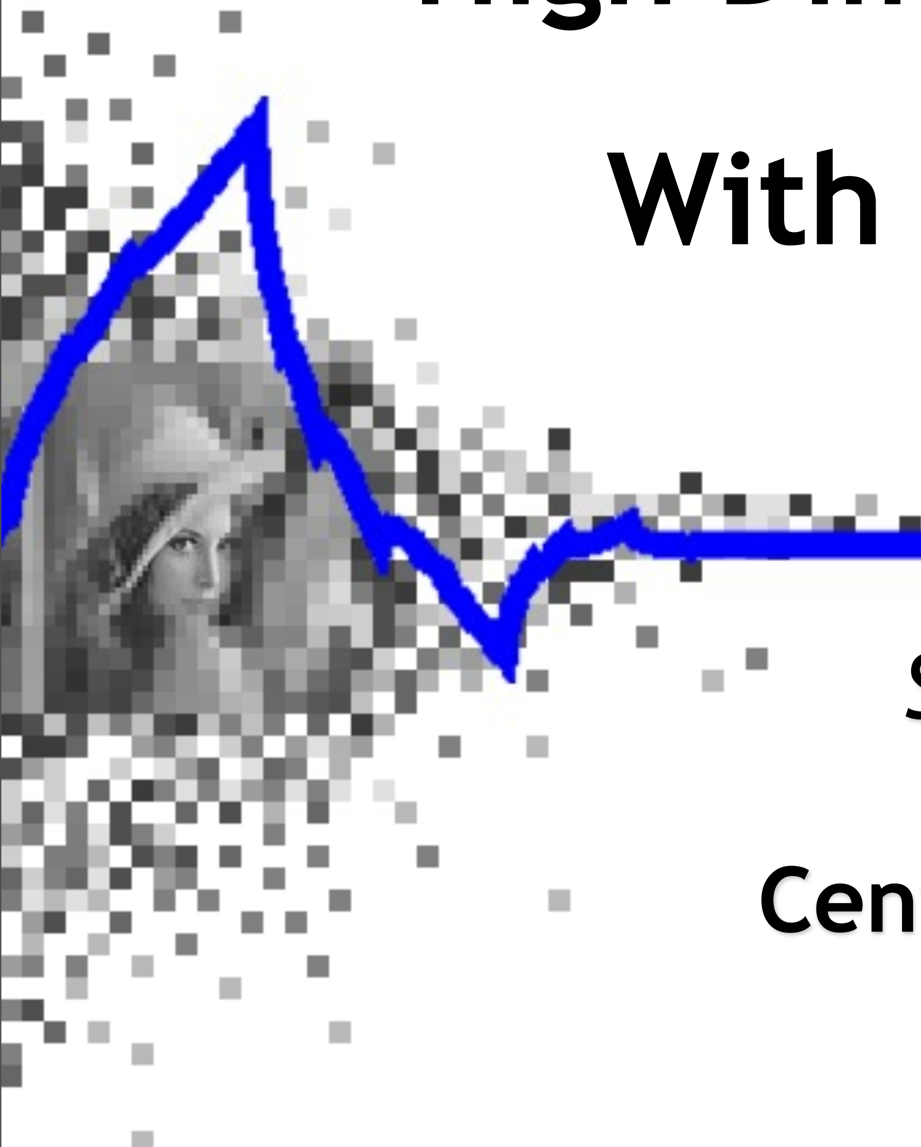


High Dimensional Signal Classification With Invariant Representations

Stéphane Mallat and Joan Bruna

Centre de Mathématiques Appliquées
Ecole Polytechnique

Low-Level Signal Representation

- Low-level statistical signal processing:
  - compression/information theory for storage and transmission
  - estimation from partial and degraded measurements

- A key idea: find **sparse** accurate representations with few parameters.

- Mathematical tools: Fourier transform, wavelet/cosine bases, adaptive representations...

- A relatively well understood framework.

Representation for Classification

Face retrieval: 

Representation for Classification

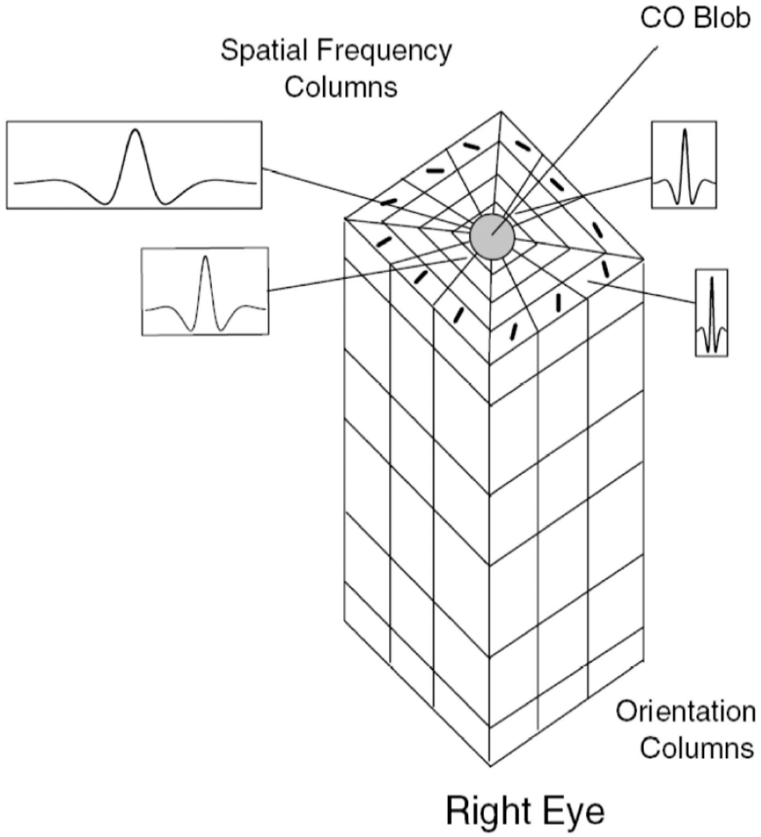
Face retrieval:

Representation for Classification

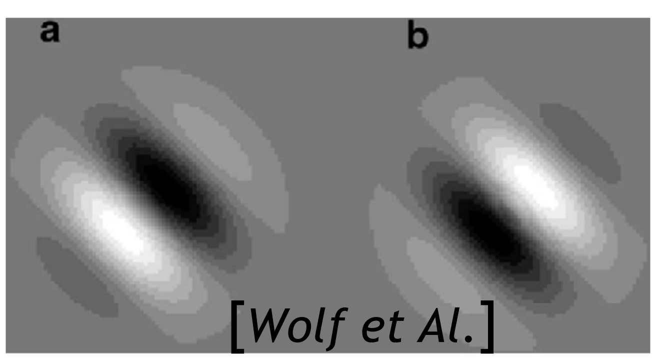
Face retrieval:

- Difficult but some algorithms work: sparsity is not key.
- Key concept: **informative stable invariants**.

Hypercolumns in V1:
directional wavelets



Simple cells Gabor linear models

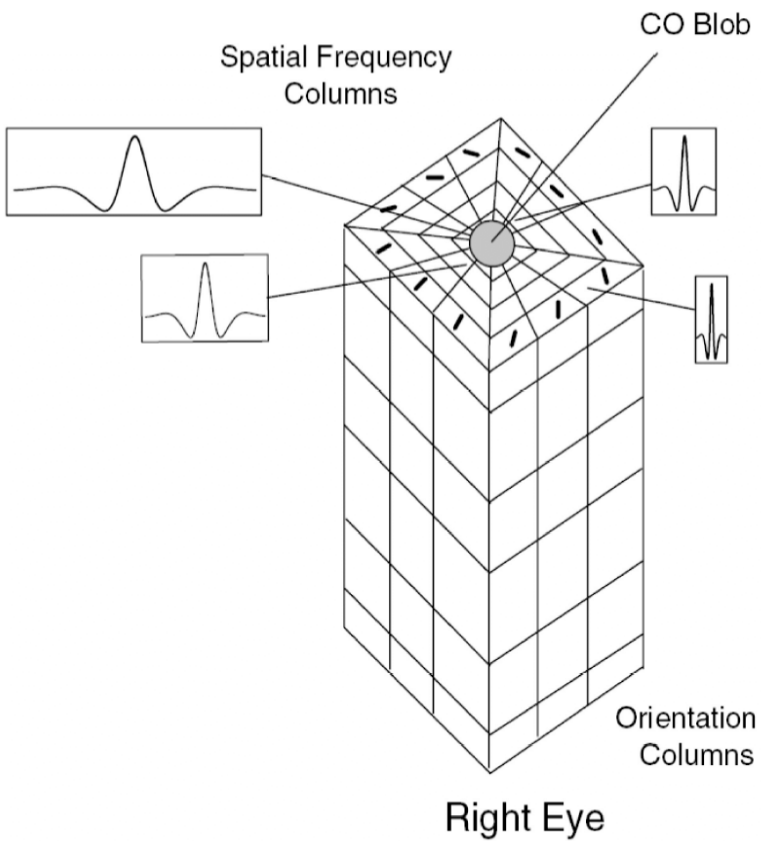


[Wolf et Al.]

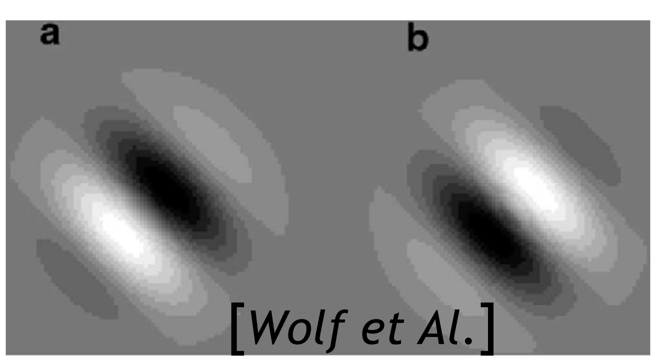
$$\psi(x) = \theta(x)e^{i\xi x}$$

Psychophysics of Vision

Hypercolumns in V1:
directional wavelets



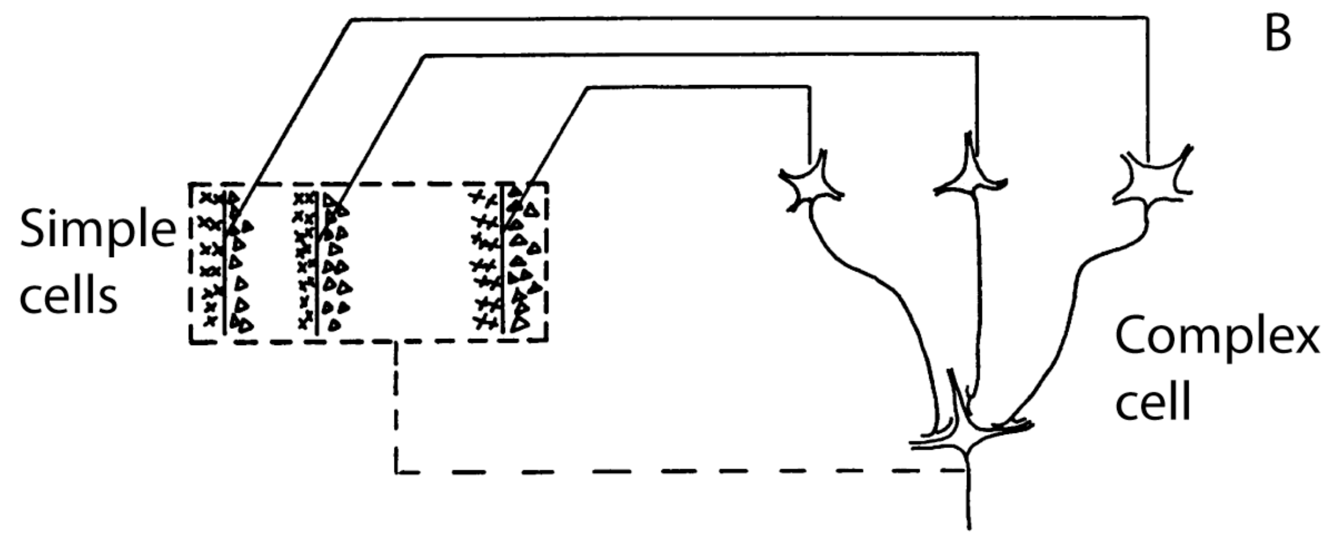
Simple cells Gabor linear models



$$\psi(x) = \theta(x) e^{i\xi x}$$

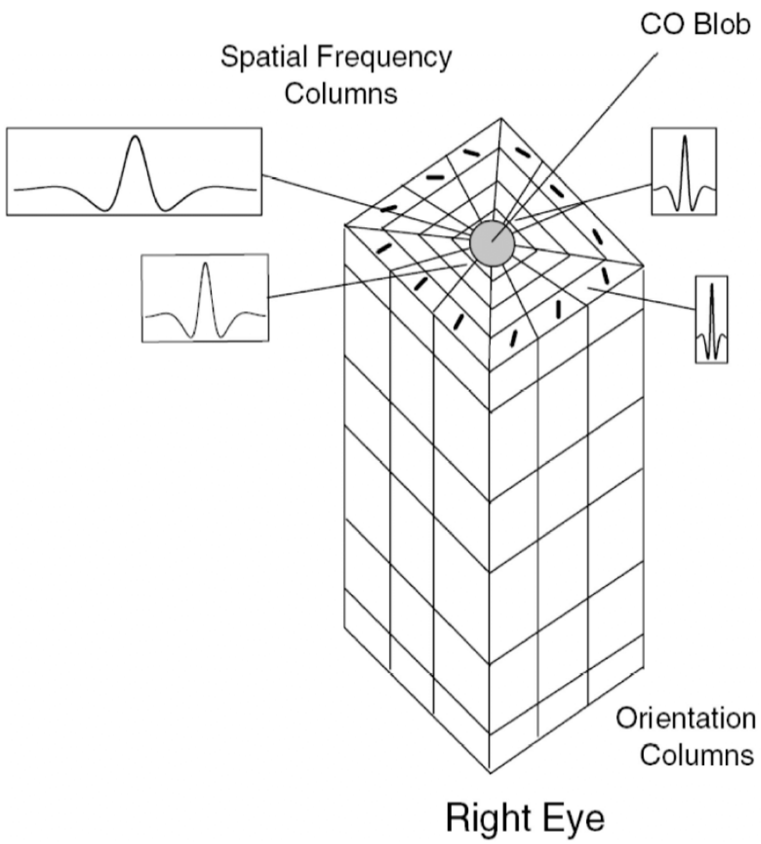
Complex Cells

- Non-linear
- Large receptive fields
- Some forms of invariance

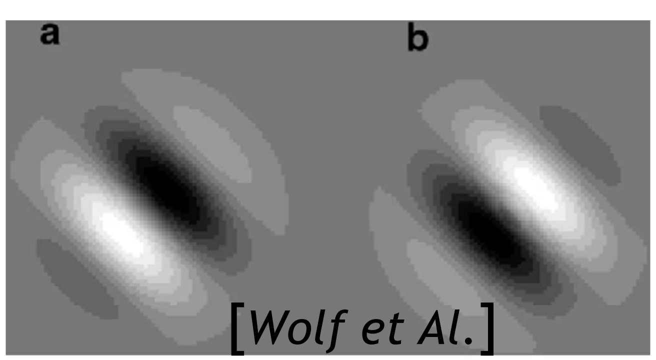


Psychophysics of Vision

Hypercolumns in V1:
directional wavelets



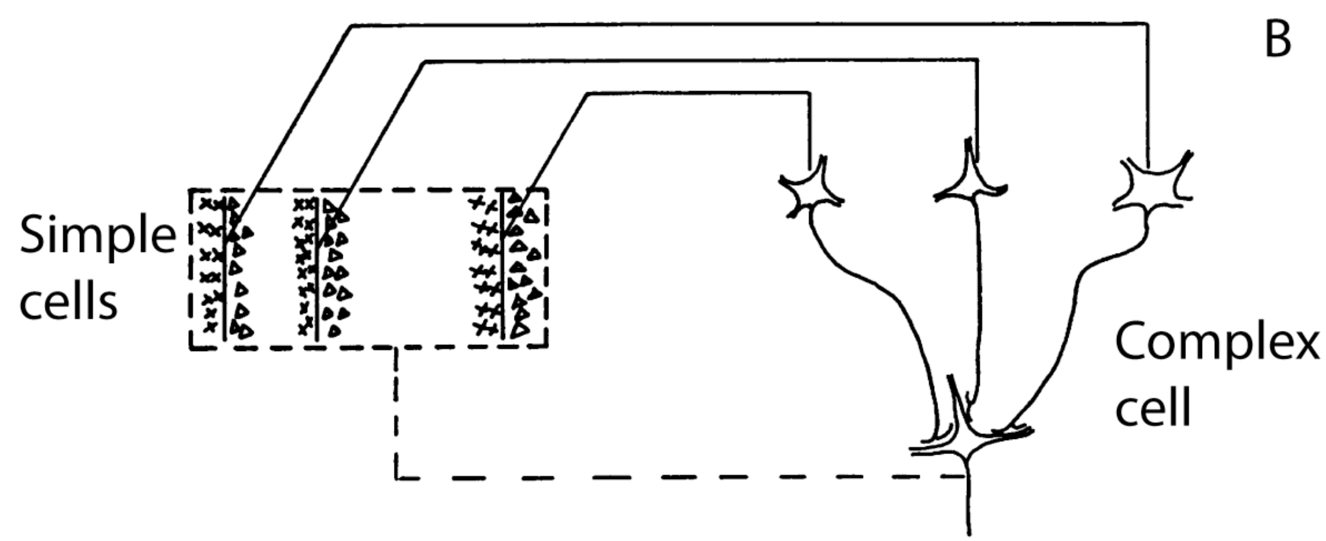
Simple cells Gabor linear models



$$\psi(x) = \theta(x)e^{i\xi x}$$

Complex Cells

- Non-linear
- Large receptive fields
- Some forms of invariance



«What» Pathway towards V4:

- More specialized invariance
- «Grand mother cells»

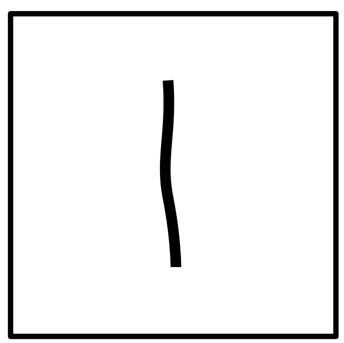
No Regularity

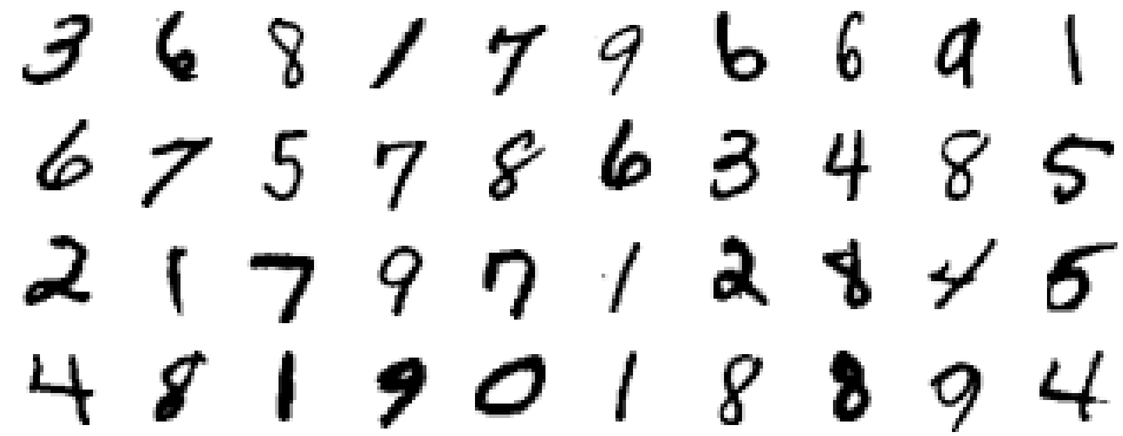
- Image classes do not define regular manifold structures.

No Regularity

- Image classes do not define regular manifold structures.

- Example of hand-written digit images

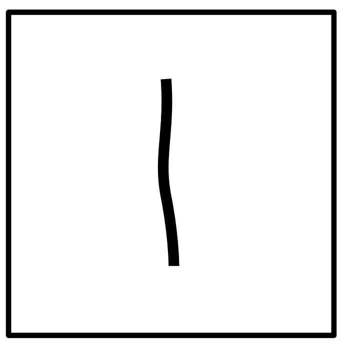


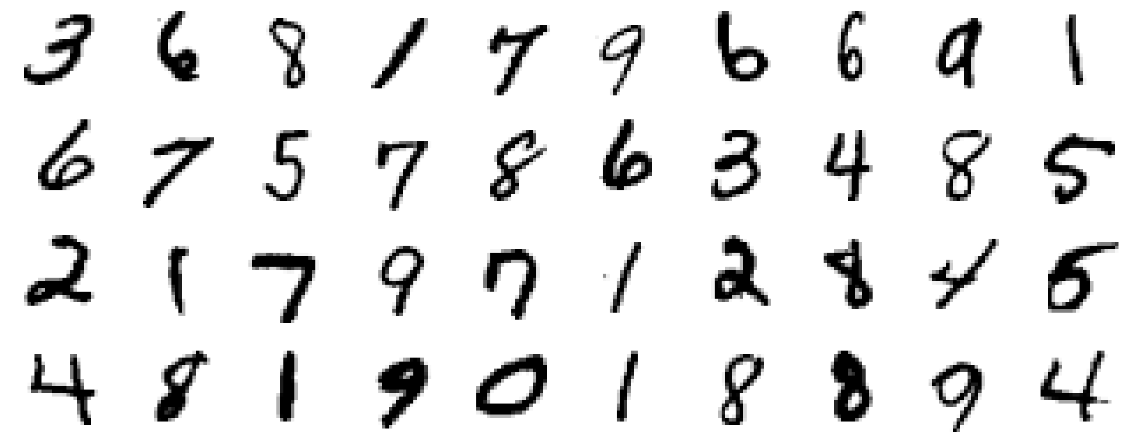


No Regularity

- Image classes do not define regular manifold structures.

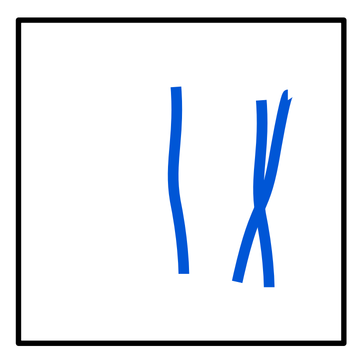
- Example of hand-written digit images

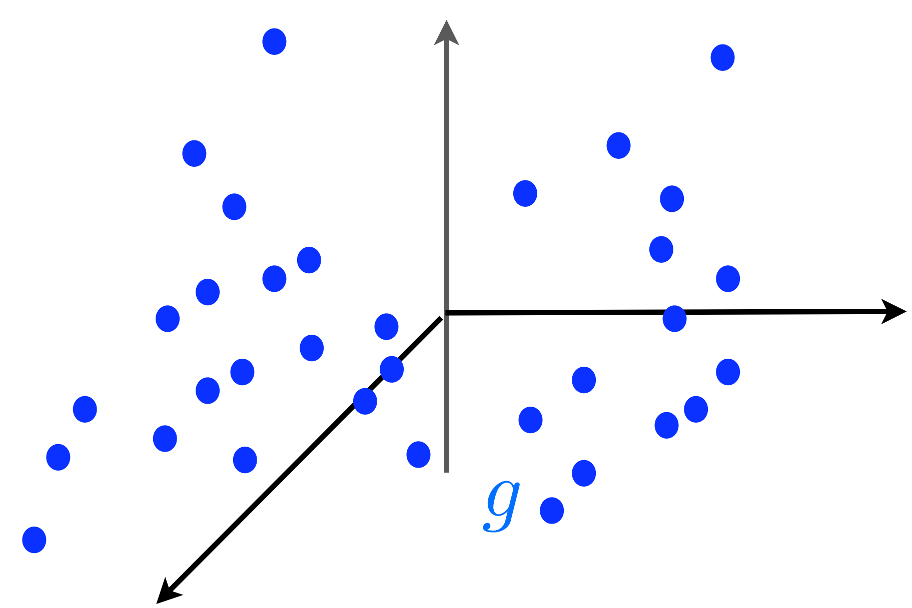




- A digit is a low-dimensional but irregular class because of:

  - Translations
  - Deformations

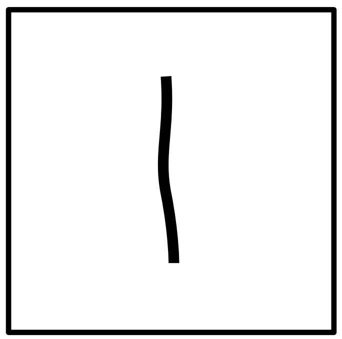


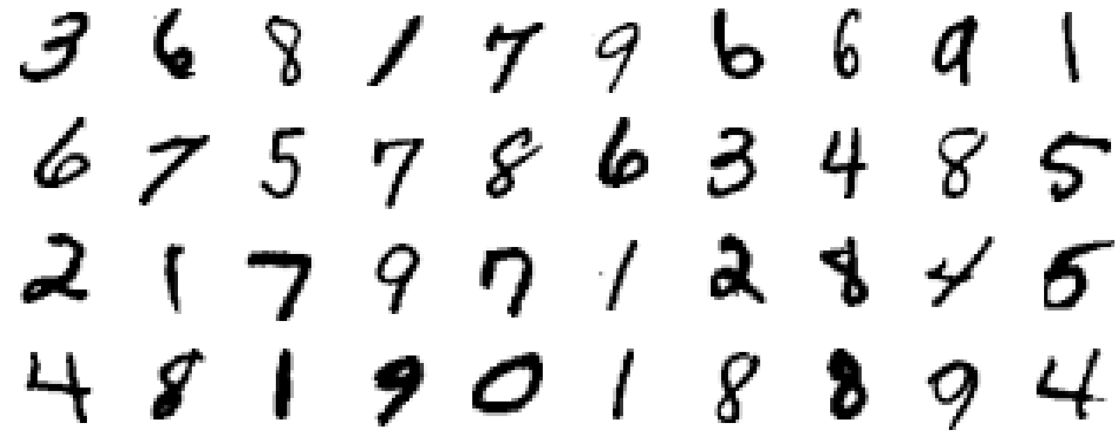


No Regularity

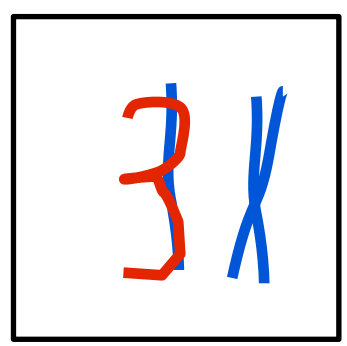
- Image classes do not define regular manifold structures.

- Example of hand-written digit images





- A digit is a low-dimensional but irregular class because of:
  - Translations
  - Deformations



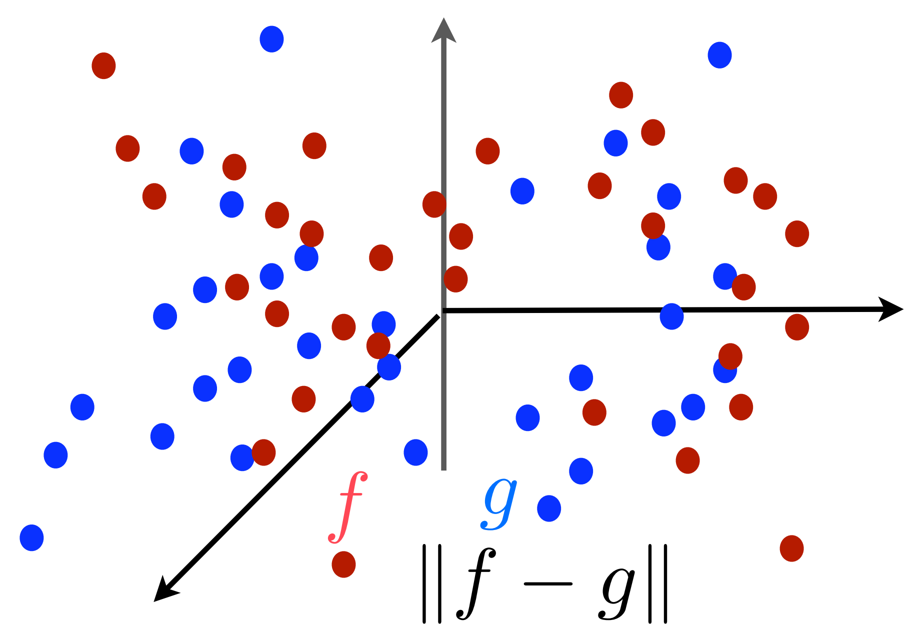


Image Classes are High Dimensional

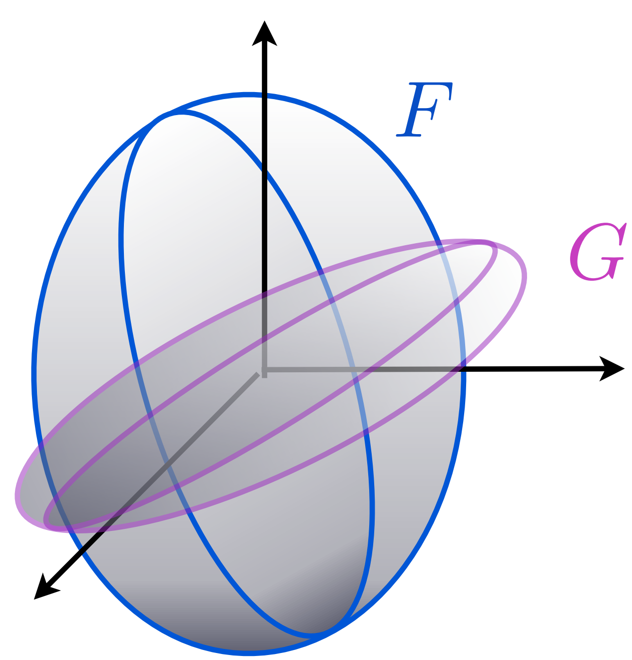
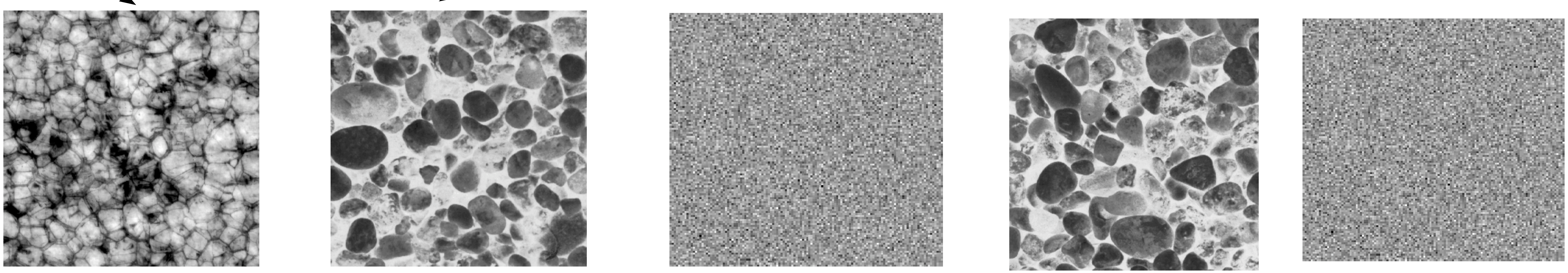
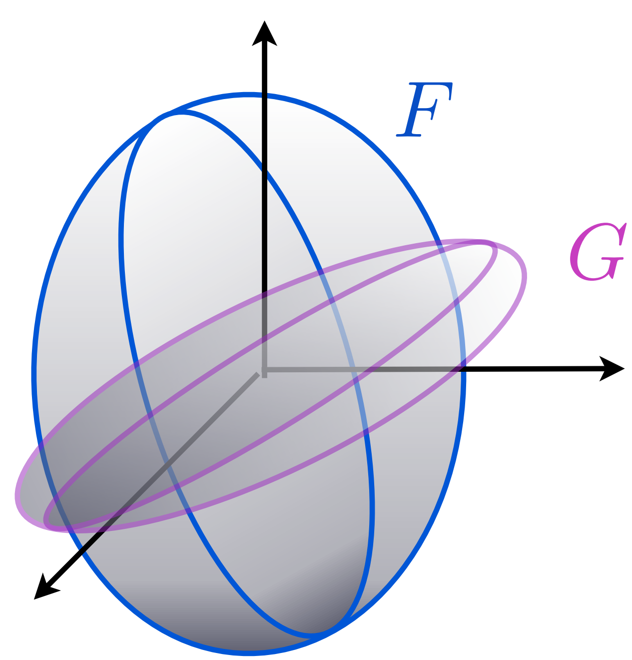


Image Classes are High Dimensional

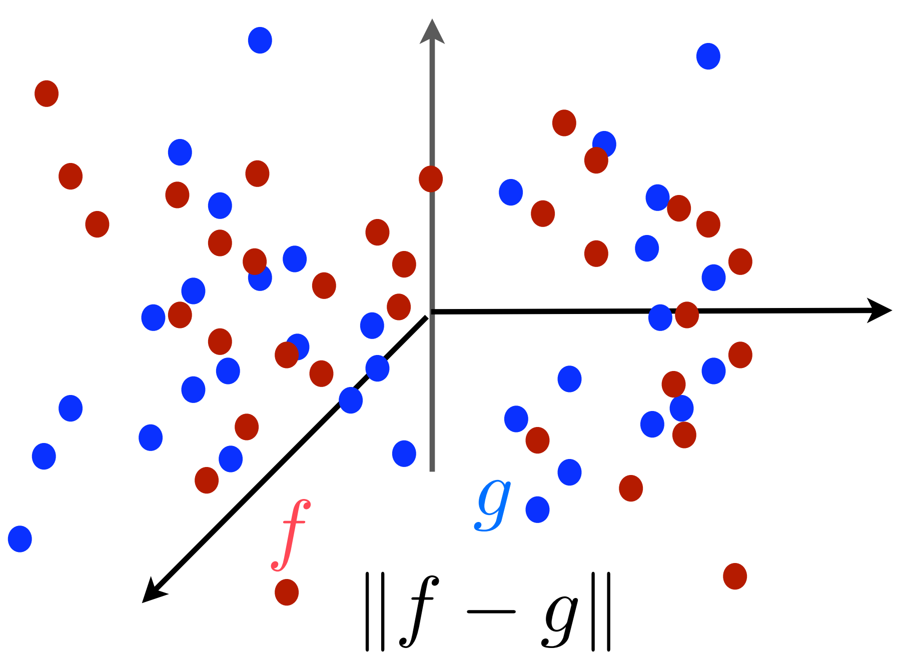
- Textures define high-dimensional image classes.
  - Realizations of stationary processes F but typically not Gaussian and not Markovian.

same power spectrum



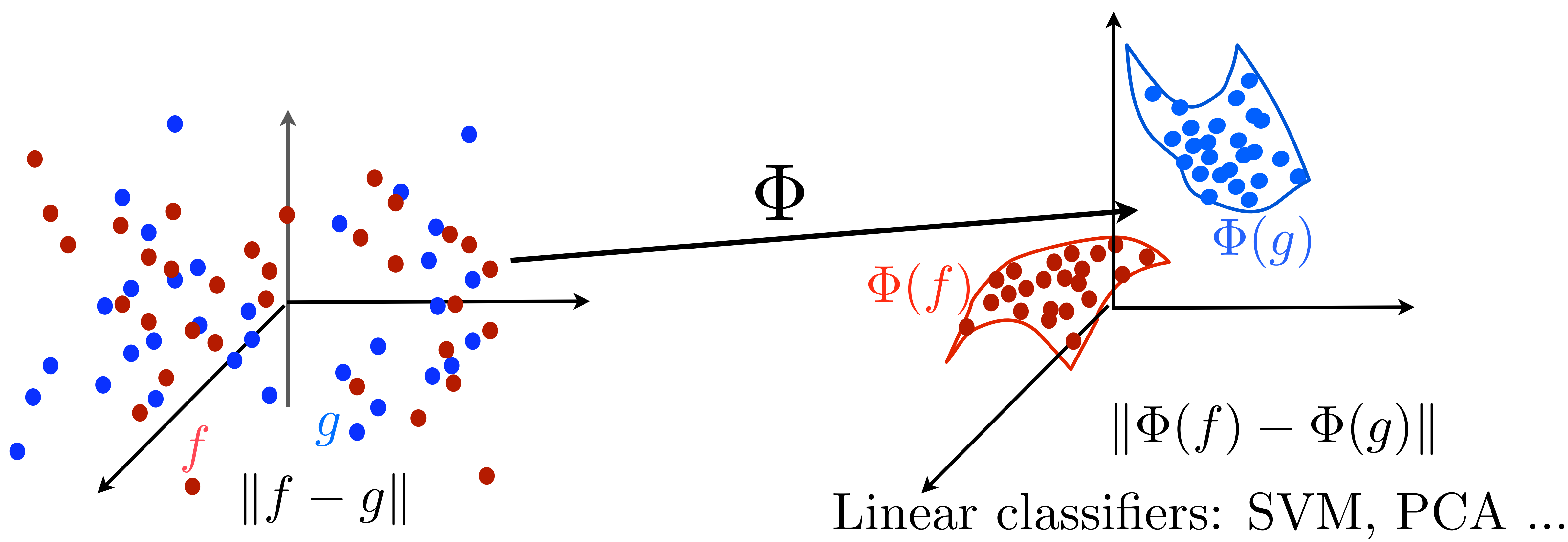


Representation for Classification



Representation for Classification

- Need to find a representation Φ which maps signals to lower-dimensional, regular manifolds by:
  - Reducing intra-class variability (***invariants***)
  - Creating a Lipschitz continuous manifold structure (***stable***)
  - Maintaining discriminability (***informative)***



Perceptual Distance

- Invariance to translations and scaling: variability reduction.

Perceptual Distance

- Invariance to translations and scaling: variability reduction.
- Sensitive to elastic deformations: natural metric.

Perceptual Distance

- Invariance to translations and scaling: variability reduction.

- Sensitive to elastic deformations: natural metric.

- Deformation of $f(x)$ into $D_\tau f(x) = f(x - \tau(x))$

$$\tau(x) \approx \tau(x_0) + \nabla\tau(x_0)(x - x_0)$$

Metric: elastic deformation amplitude $\|\nabla\tau\|_\infty = \sup_x |\nabla\tau(x)|$

Distance from Representations

- Euclidean distance on a representation: $\|\Phi(f) - \Phi(g)\|$

- **Invariance** to groups of operators $\{D_\tau\}_\tau$ such as rigid translations $D_\tau f(x) = f(x-\tau)$:

$$\Phi(D_\tau f) = \Phi(f) \quad : \quad \textit{weak property.}$$

- **Stability:** Lipschitz continuity to deformations

$$D_\tau f(x) = f(x - \tau(x))$$

$$\|\Phi(f) - \Phi(D_\tau f)\| \leq C\, \|f\| \|\nabla \tau\|_\infty \ .$$

Linearizes small deformations.

Overview

- Averaging, Fourier and wavelets.
- Invariance through scattering: Convolution Networks

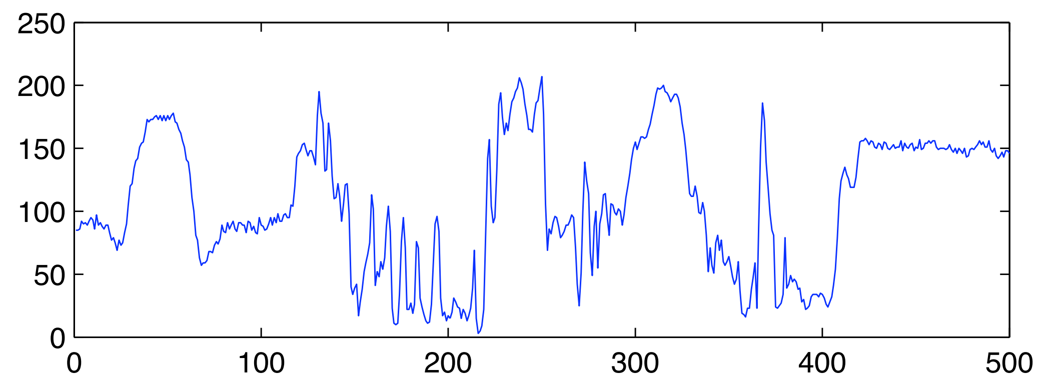
  $$\textit{Variability reduction with iterative contractions}$$

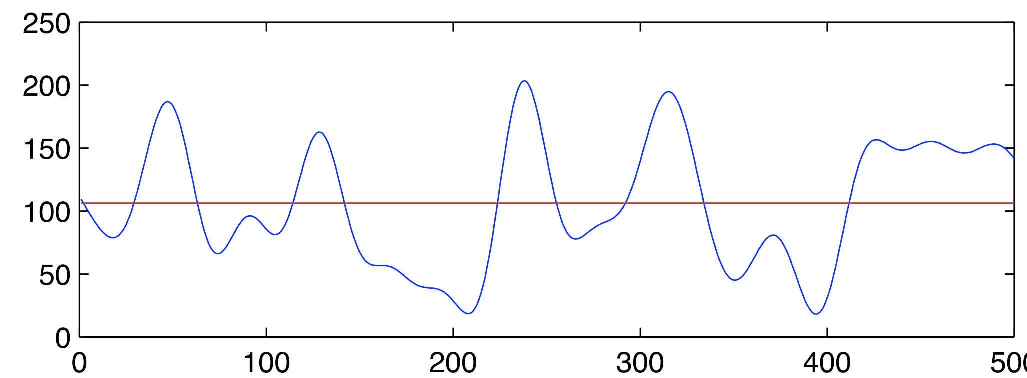
- Representation of stationary processes for textures
- Scattering PCA classification of patterns and textures
- General group invariance and learning

Invariance by Averaging

- Averaging kernel: $\phi_J(x) = 2^{-J}\,\phi(2^{-J}x)$.

$$f(x) \qquad\qquad f\star\phi_J(x) \underset{J\to\infty}{\longrightarrow} \int f(u)\,du$$





- $f\star\phi_J$ is invariant to translations small relatively to 2^J

- $f\star\phi_J$ looses too much information for discriminability.

Deformation Instability of Fourier

- Fourier modulus is invariant to translations

If $D_\tau f(x) = f(x-\tau)$ then $\widehat{D_\tau f}(\omega) = e^{-i\tau\omega}\hat{f}(\omega)$

so $|\widehat{D_\tau f}(\omega)| = |\hat{f}(\omega)|$: $\Phi(f) = |\hat{f}|$.

- For deformations $D_\tau f(x) = f(x-\tau(x))$

$|\hat{f}(\omega)|$ is unstable at high frequencies ξ:

$$\| \, |\widehat{D_\tau f}| - |\hat{f}| \, \| \sim \|f\| \, \|\nabla\tau \cdot \xi\|_\infty$$

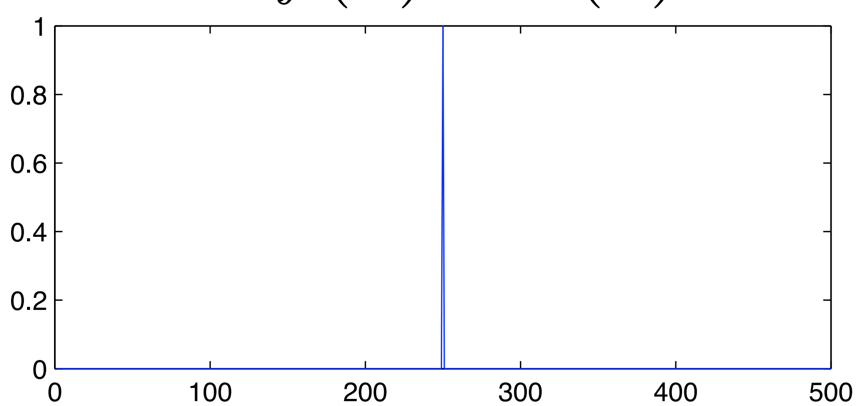
with $\|f\|^2 = \int |f(x)|^2 \, dx$

Loss of Discriminability

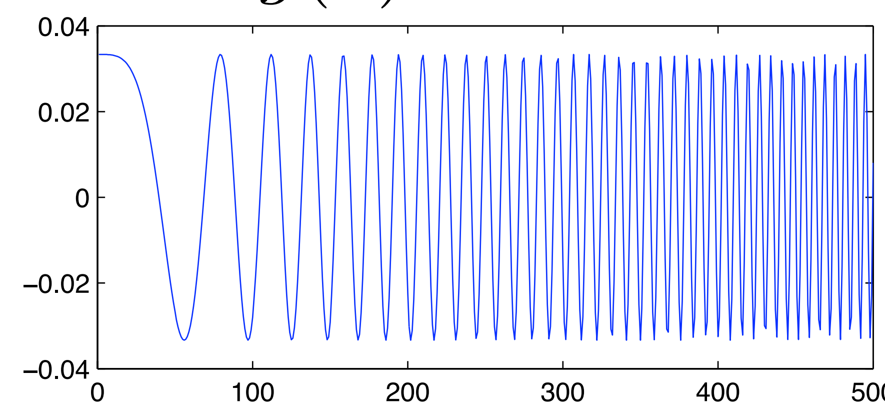
- The loss of the Fourier phase eliminates too much information.

  $\delta(x)$ and e^{ix^2} have same Fourier modulus (constant).

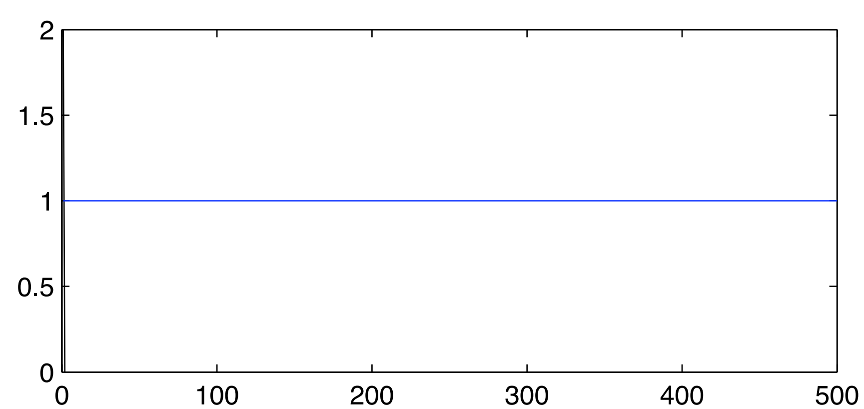
$$f(x) = \delta(x)$$



$$g(x) = \lambda\, e^{ix^2}$$

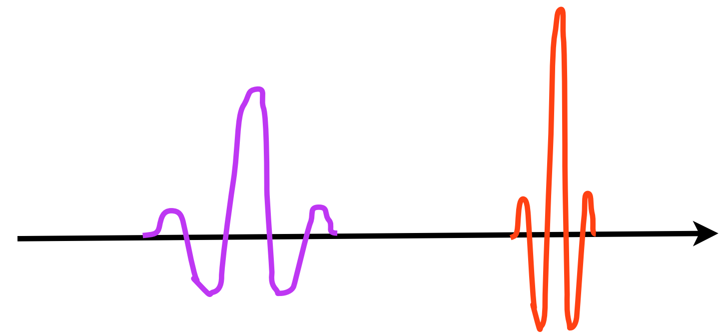


$$|\hat{f}(\omega)| = |\hat{g}(\omega)|$$

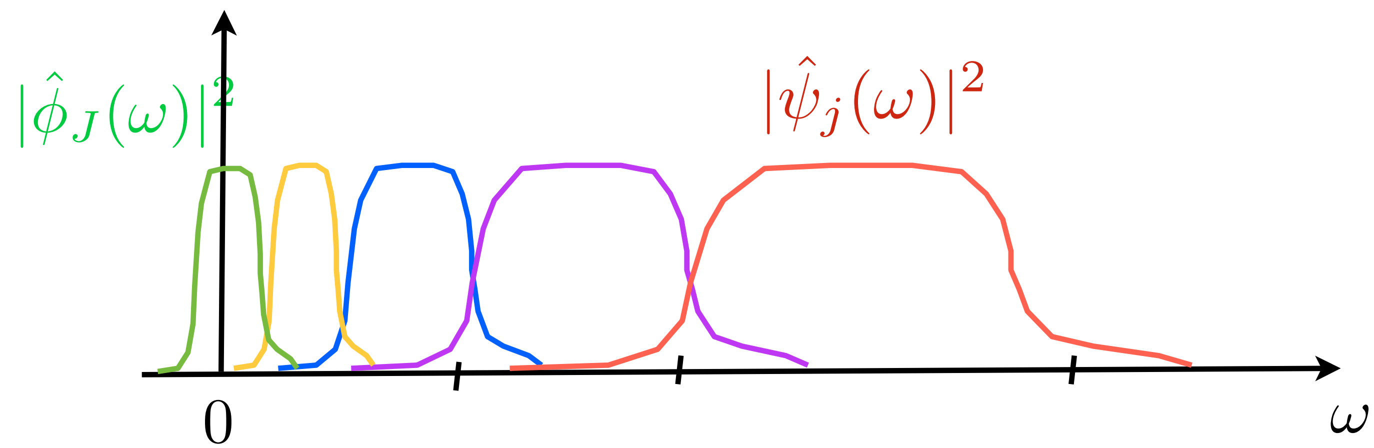


Wavelets

- In 1D, dilated wavelets:

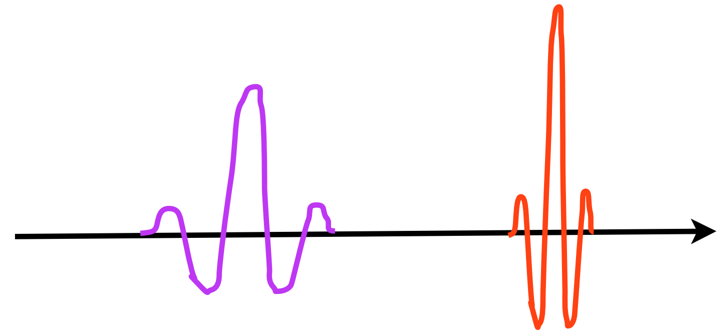


$$\psi_j(x) = 2^{-j}\,\psi(2^{-j}x)$$

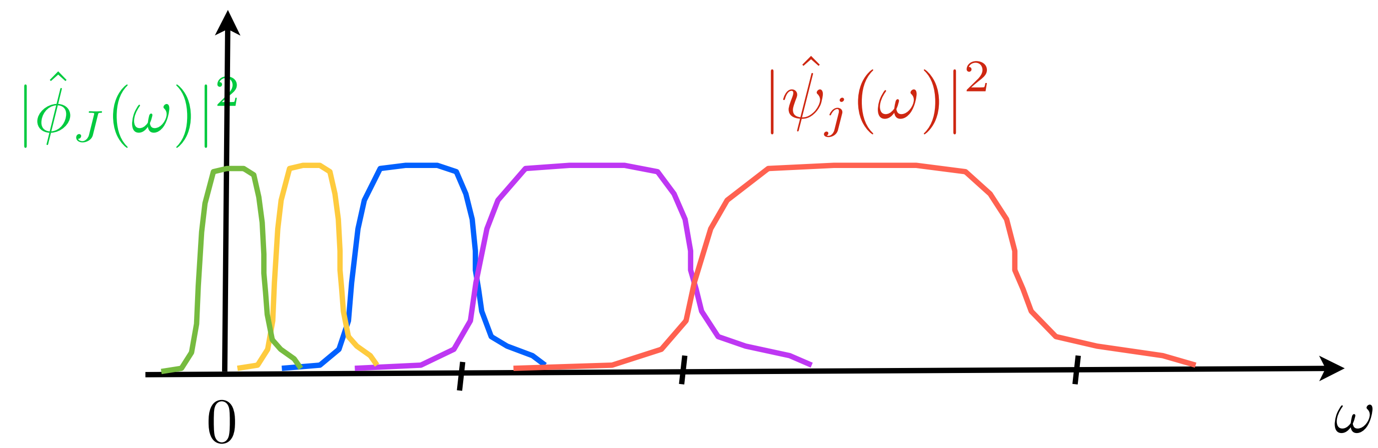


Wavelets

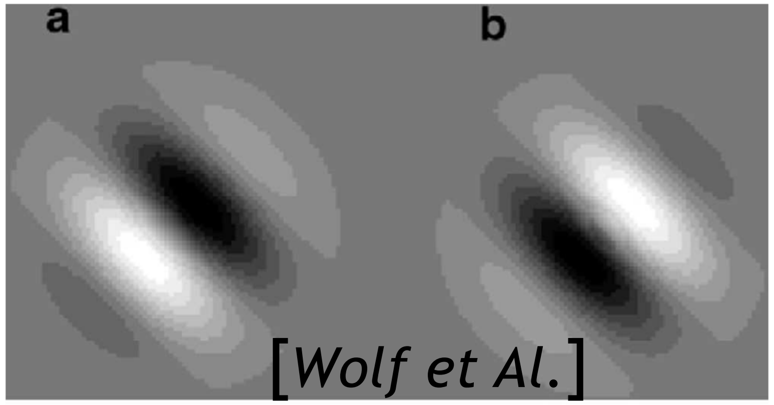
- In 1D, dilated wavelets:



$$\psi_j(x) = 2^{-j}\,\psi(2^{-j}x)$$

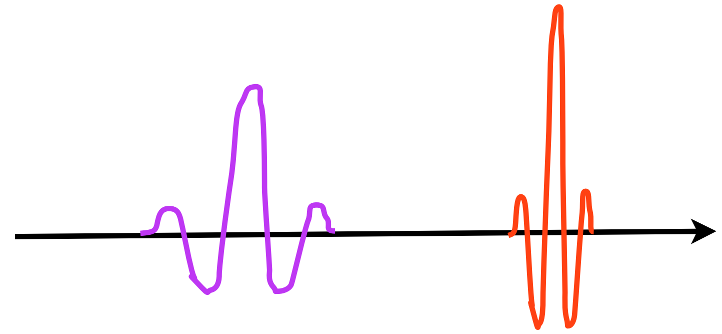


- In 2D, dilated and rotated wavelets: $\psi(x) = \theta(x)\, e^{i\xi x}$

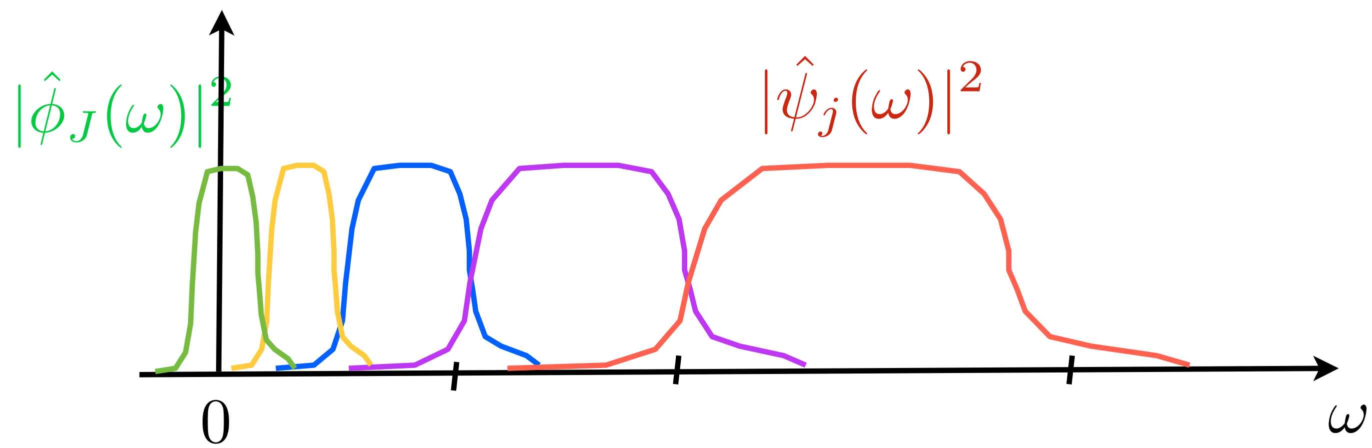


Wavelets

- In 1D, dilated wavelets:

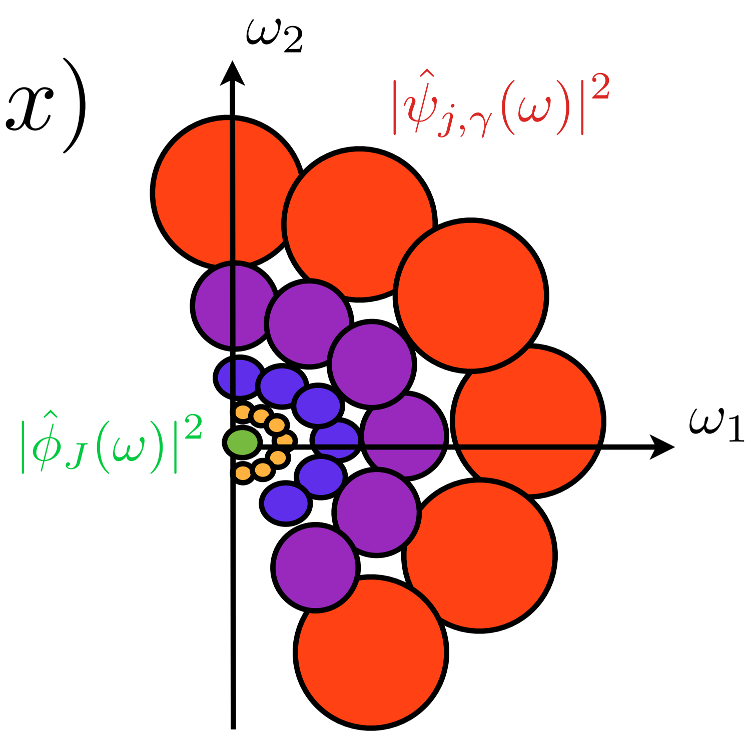


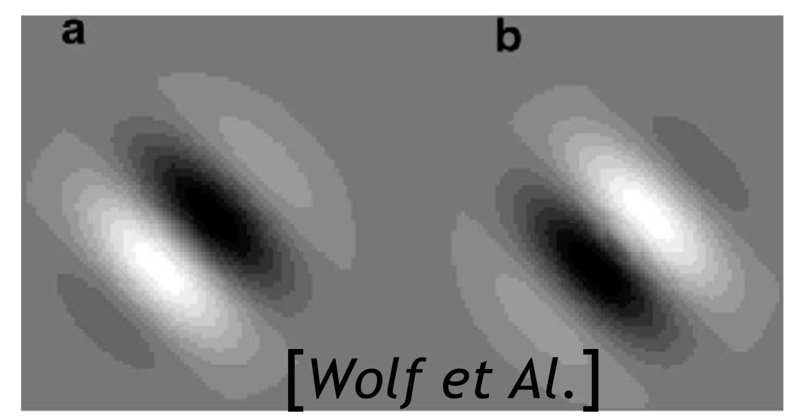
$$\psi_j(x) = 2^{-j}\,\psi(2^{-j}x)$$



- In 2D, dilated and rotated wavelets: $\psi(x) = \theta(x)\, e^{i\xi x}$

$$\psi_{j,\gamma}(x) = 2^{-2j}\,\psi(2^{-j}R_\gamma x)$$





Wavelet Transforms

- Wavelet transform of f at a scale 2^J:

$$W_J f(x) = \left(\begin{array}{c} f \star \phi_J(x) \\ f \star \psi_{j,\gamma}(x) \end{array} \right)_{j<J,\gamma\in\Gamma}$$

- Unitary:

$$\|W_J f\|^2 = \|f \star \phi_J\|^2 + \sum_{j<J,\gamma\in\Gamma} \|f \star \psi_{j,\gamma}\|^2 = \|f\|^2$$

Image and Audio Descriptors

- How to build invariant descriptors from wavelet coefficients ?
- If f is translated then $f \star \psi_{j,\gamma}$ is translated

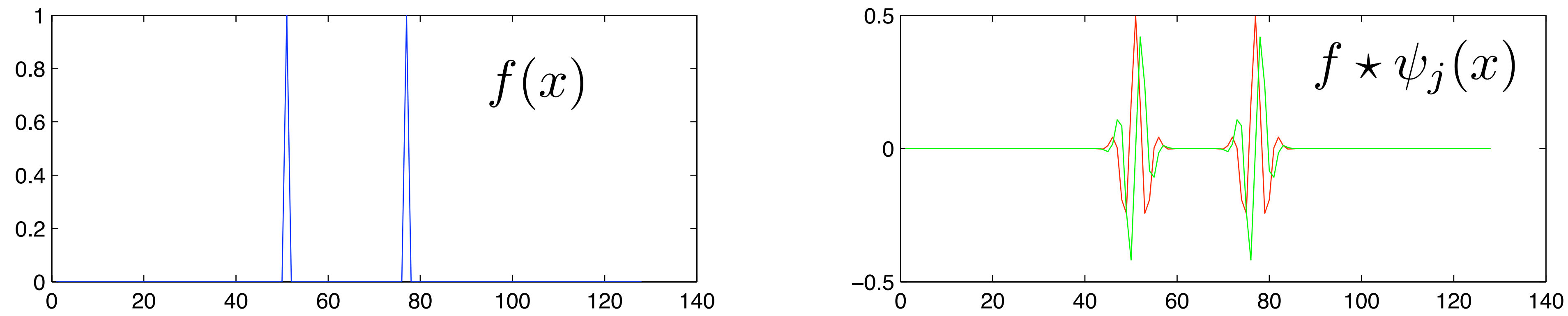
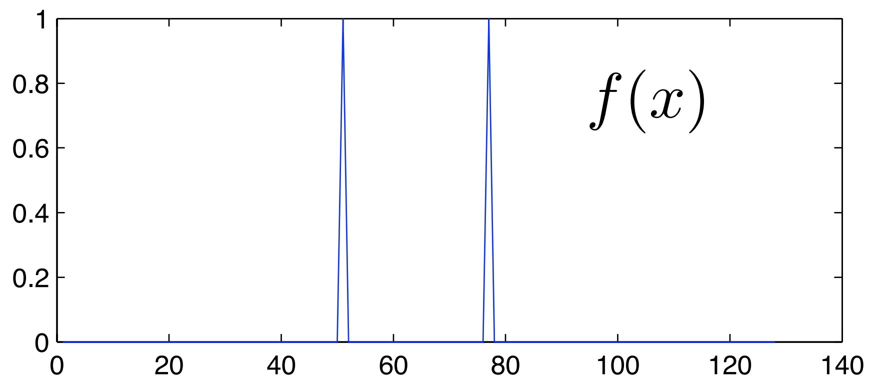
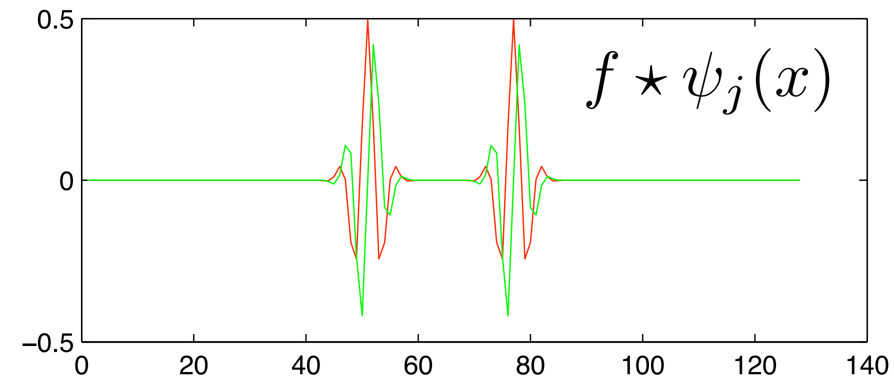


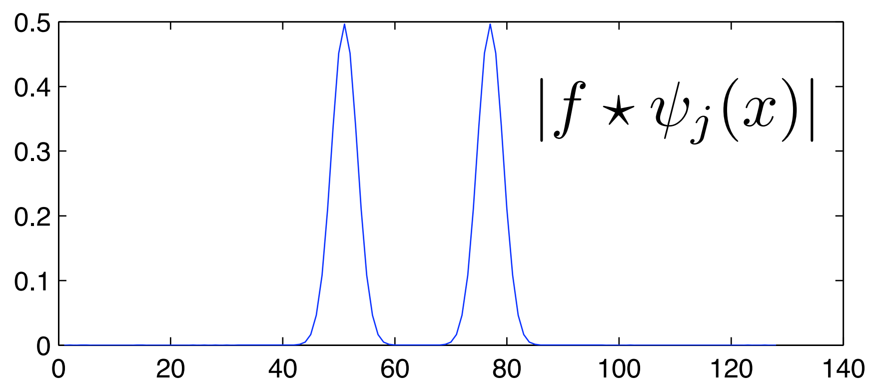
Image and Audio Descriptors

- How to build invariant descriptors from wavelet coefficients ?
- If f is translated then $f \star \psi_{j,\gamma}$ is translated





- $|f \star \psi_{j,\gamma}| \star \phi_J$ (SIFT, MFSC) locally translations if $\tau \ll 2^J$.



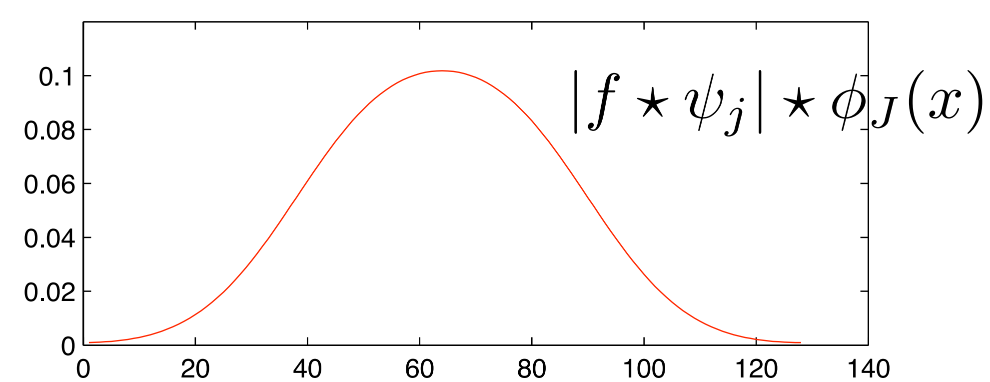
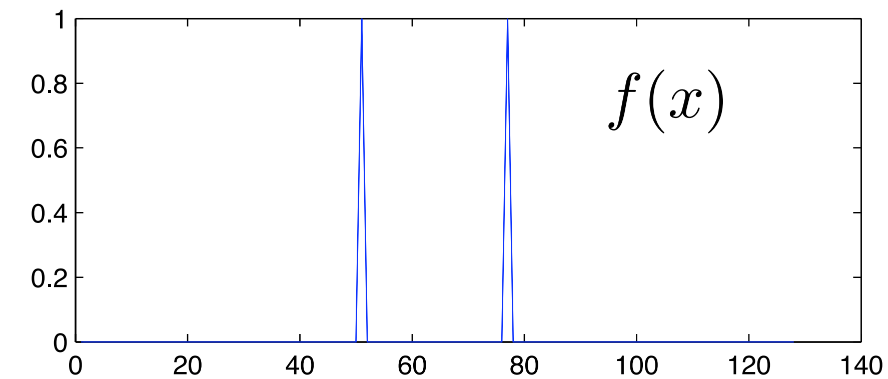
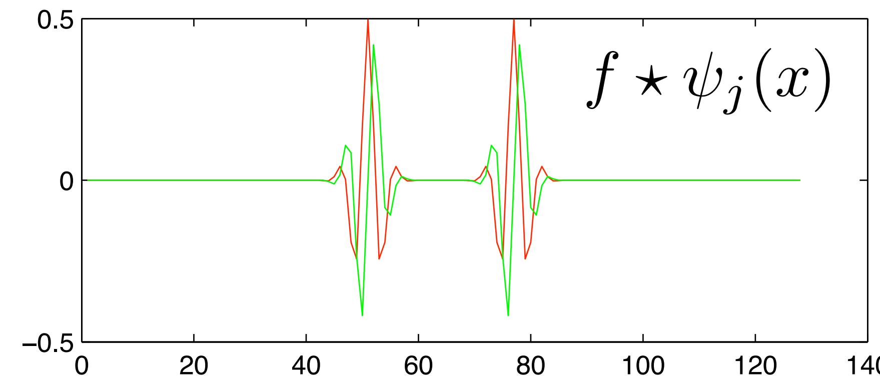


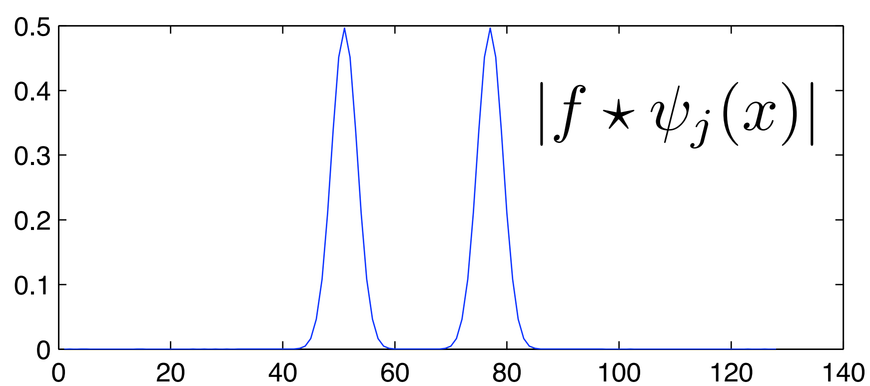
Image and Audio Descriptors

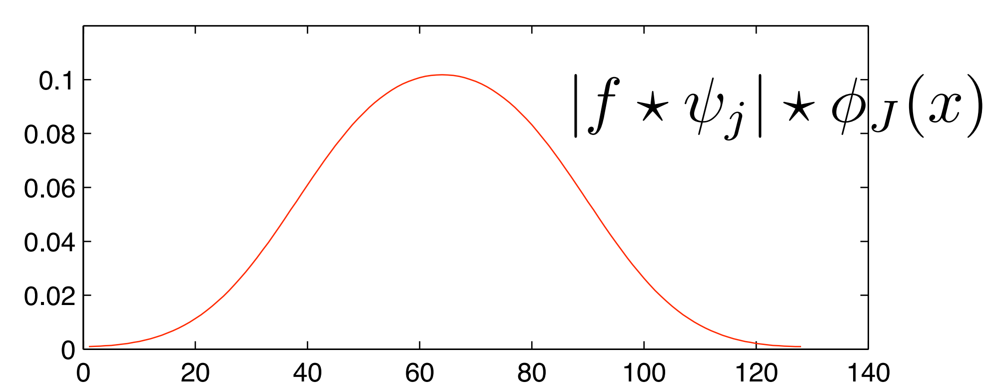
- How to build invariant descriptors from wavelet coefficients ?
- If f is translated then $f \star \psi_{j,\gamma}$ is translated





- $|f \star \psi_{j,\gamma}| \star \phi_J$ (SIFT, MFSC) locally translations if $\tau \ll 2^J$.





- **Problem:** Important loss of information by averaging.
- Can we recover information that remains locally invariant ?

Scattering Operators

$$|f \star \psi_{j_1,\gamma_1}| \star \phi_J$$

Scattering Operators

$$W_J(|f * \psi_{j_1,\gamma_1}|) = \begin{pmatrix} |f \star \psi_{j_1,\gamma_1}| \star \phi_J \\ \\ |f \star \psi_{j_1,\gamma_1}| \star \psi_{j_2,\gamma_2} \end{pmatrix}_{\substack{j_2 < J \\ \gamma_2 \in \Gamma}} \longrightarrow |f \star \psi_{j_1,\gamma_1}|$$

Scattering Operators

$$W_J(|f * \psi_{j_1,\gamma_1}|) = \begin{pmatrix} |f \star \psi_{j_1,\gamma_1}| \star \phi_J \\ \\ \\ |f \star \psi_{j_1,\gamma_1}| \star \psi_{j_2,\gamma_2} \end{pmatrix}_{\substack{j_2 < J \\ \gamma_2 \in \Gamma}}$$

$$\Big\downarrow \; \text{Translation invariance}$$

$$||f \star \psi_{j_1,\gamma_1}| \star \psi_{j_2,\gamma_2}| \star \phi_J \ .$$

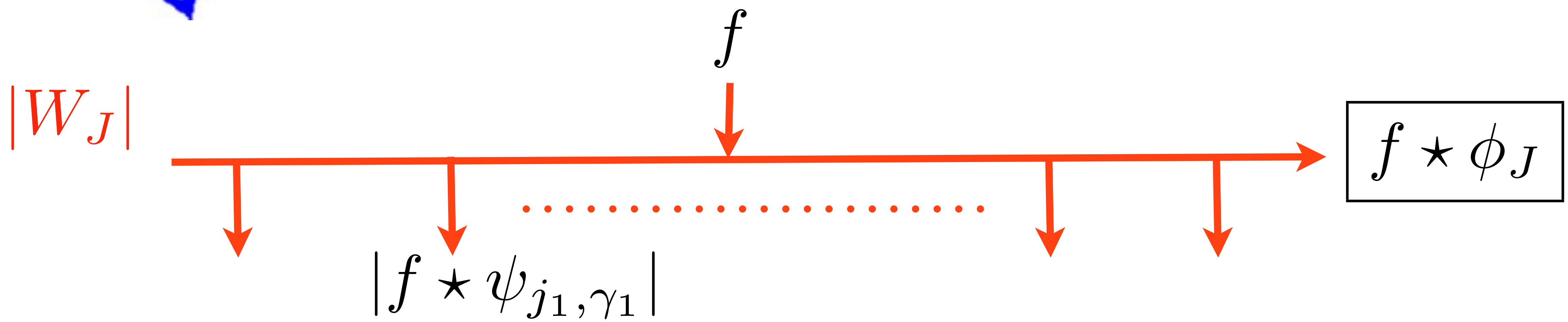
Co-occurrence at scales 2^{j_1}, 2^{j_2} and directions γ_1, γ_2.

Scattering: Convolution Network

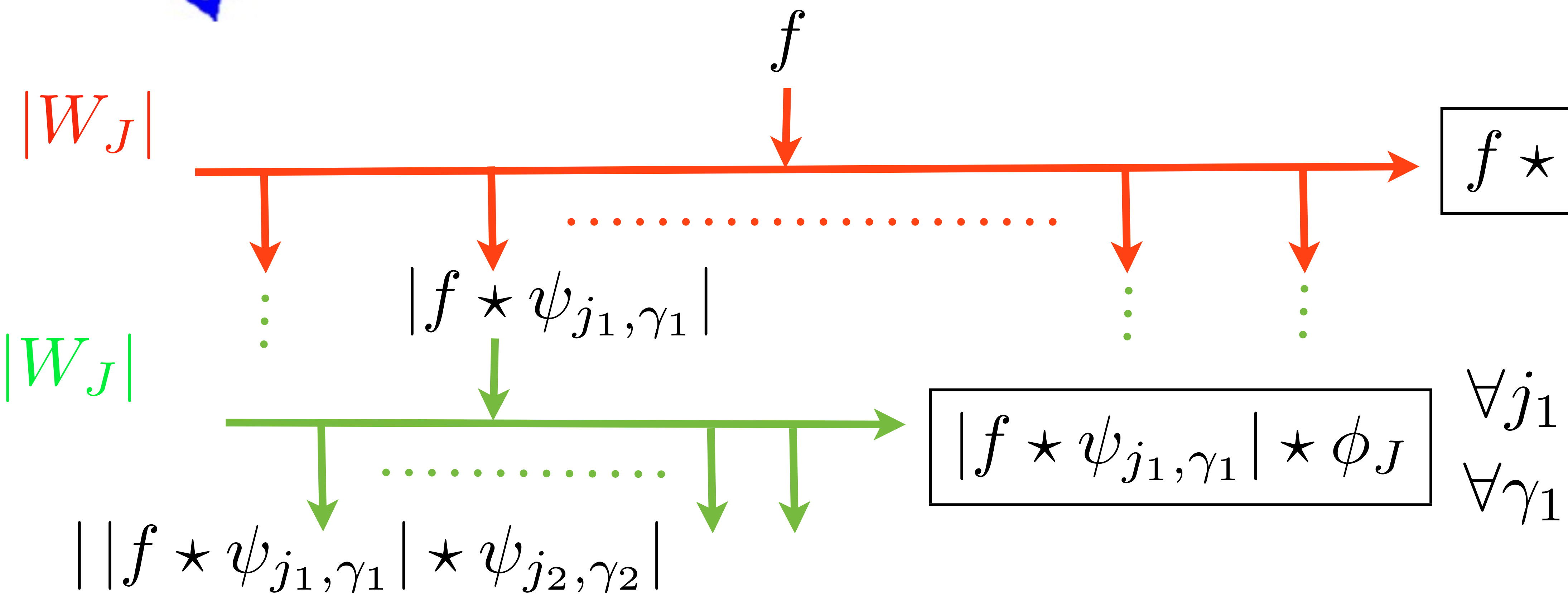
Scattering: Convolution Network

f

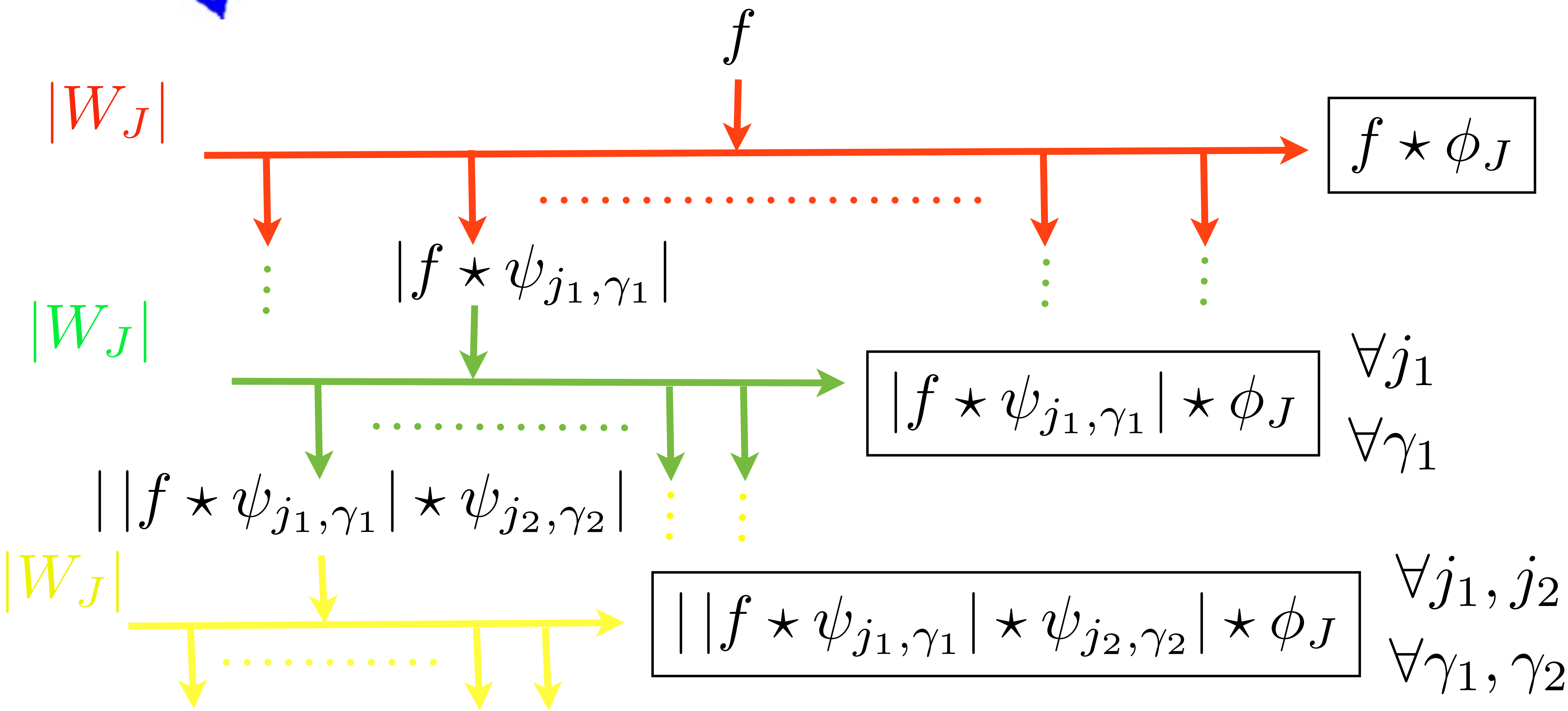
Scattering: Convolution Network



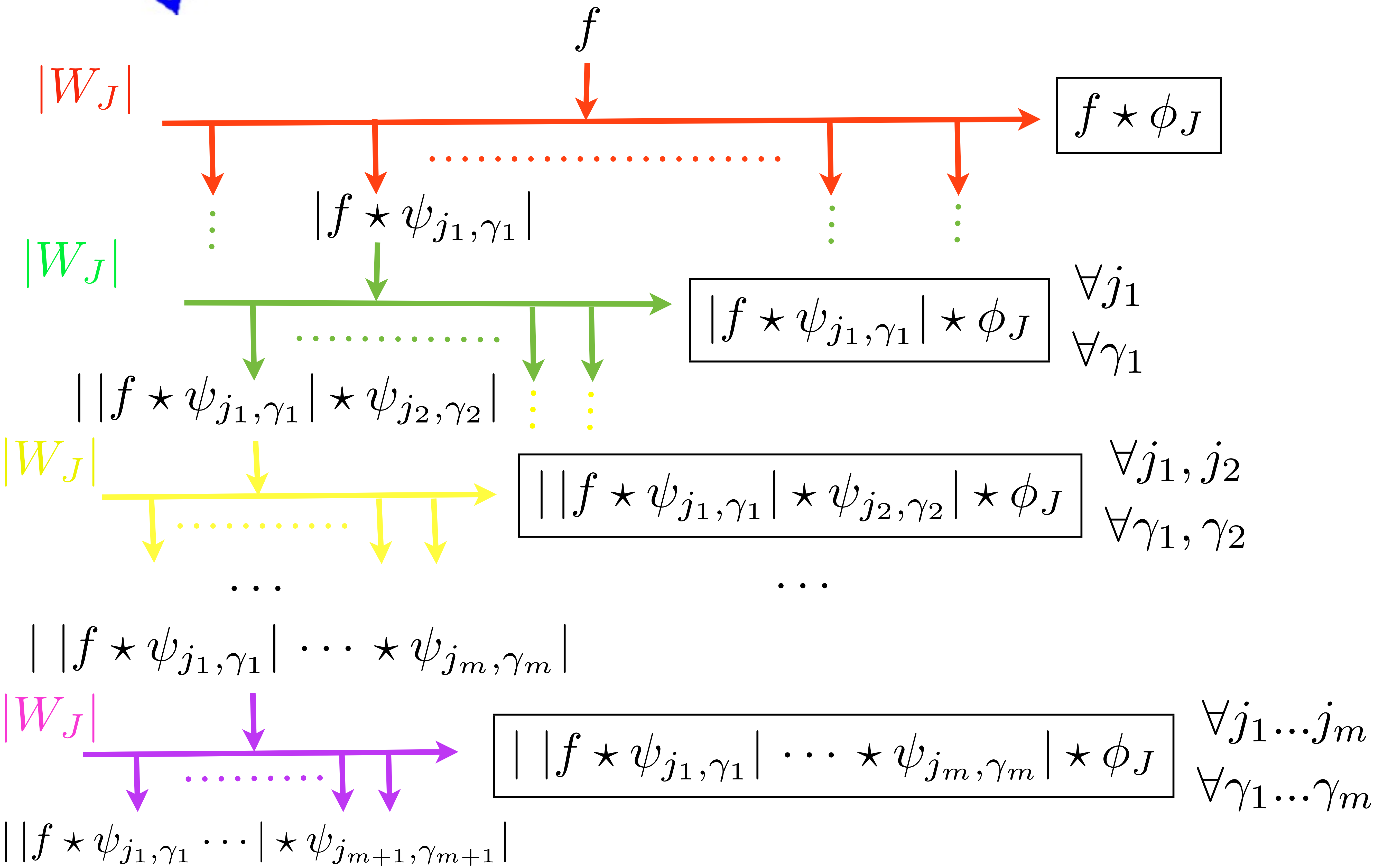
Scattering: Convolution Network



Scattering: Convolution Network



Scattering: Convolution Network



Cascade of contractive operators.

Scattering Representation

$$S_J f(x) = \begin{pmatrix} f \star \phi_J(x) \\ |f \star \psi_{j_1,\gamma_1}| \star \phi_J(x) \\ ||f \star \psi_{j_1,\gamma_1}| \star \psi_{j_2,\gamma_2}| \star \phi_J(x) \\ \cdots \\ |\,|f \star \psi_{j_1,\gamma_1}| \cdots \star \psi_{j_m,\gamma_m}| \star \phi_J(x) \end{pmatrix}_{\substack{\forall j_1 \ldots j_m \\ \forall \gamma_1 \ldots \gamma_m}}$$

Scattering norm:

$$\|S_J f\|^2 = \sum_{m=0}^{+\infty} \sum_{\substack{j_1 \cdots j_m \\ \gamma_1 \cdots \gamma_m}} \|\,|\,|f \star \psi_{j_1,\gamma_1}| \cdots \star \psi_{j_m,\gamma_m}| \star \phi_J\|^2$$

Contractive because cascade of contractive operators $|W_J|$:

$$\|S_J f - S_J g\| \leq \|f - g\|.$$

Scattering Energy Conservation

Theorem: For appropriate complex wavelets

$$\lim_{m\to\infty} \sum_{\substack{(j_1 \dots j_m)\in \mathbf{Z}^m \\ (\gamma_1 \dots \gamma_m)\in\Gamma^m}} \| \, | \, |f \star \psi_{j_1,\gamma_1}| \cdots | \star \psi_{j_m,\gamma_m}| \, \|^2 = 0$$

so a scattering is unitary:

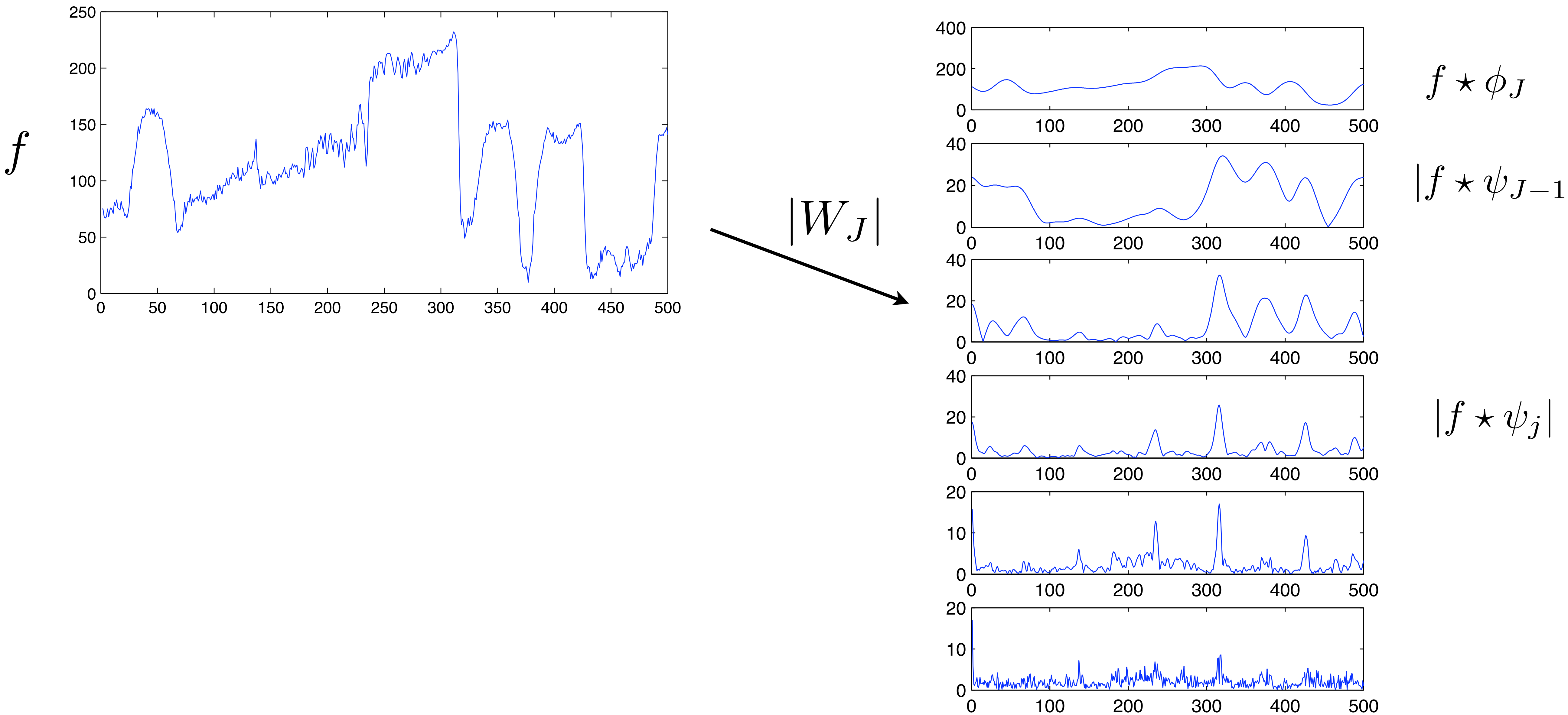
$$\|S_J f\|^2 = \|f\|^2 \ .$$

Completness and Reconstruction

Theorem *(with Waldspurger)*: For appropriate wavelets

$$|W_J|\, f = \begin{pmatrix} f \star \phi_J \\ |f \star \psi_j| \end{pmatrix}_{j<J}$$

is invertible over band-limited signals.

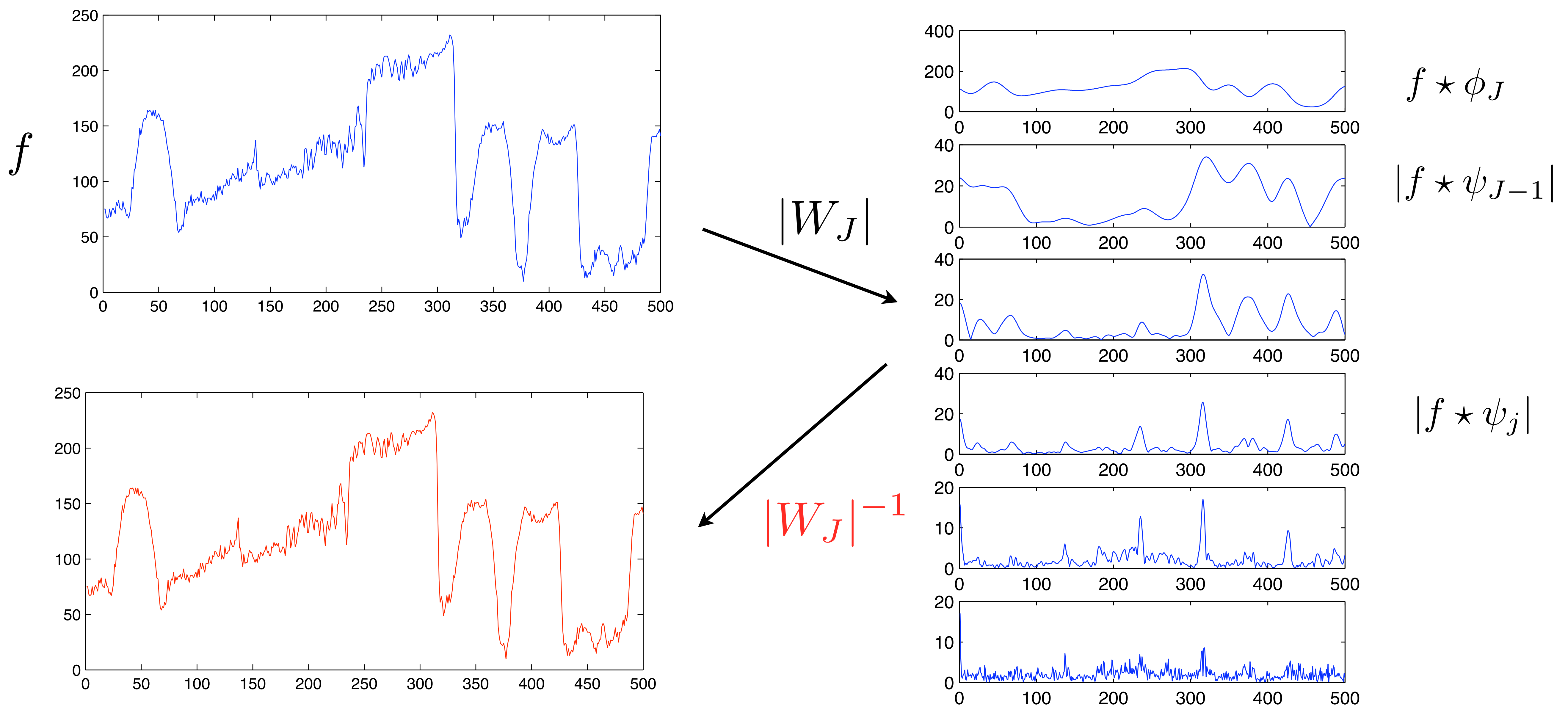


Completness and Reconstruction

Theorem *(with Waldspurger):* For appropriate wavelets

$$|W_J|\, f = \begin{pmatrix} f \star \phi_J \\ |f \star \psi_j| \end{pmatrix}_{j<J}$$

is invertible over band-limited signals.



Computational Complexity

- Scattering coefficients $S_J f(x)$ are averaged by ϕ_J.

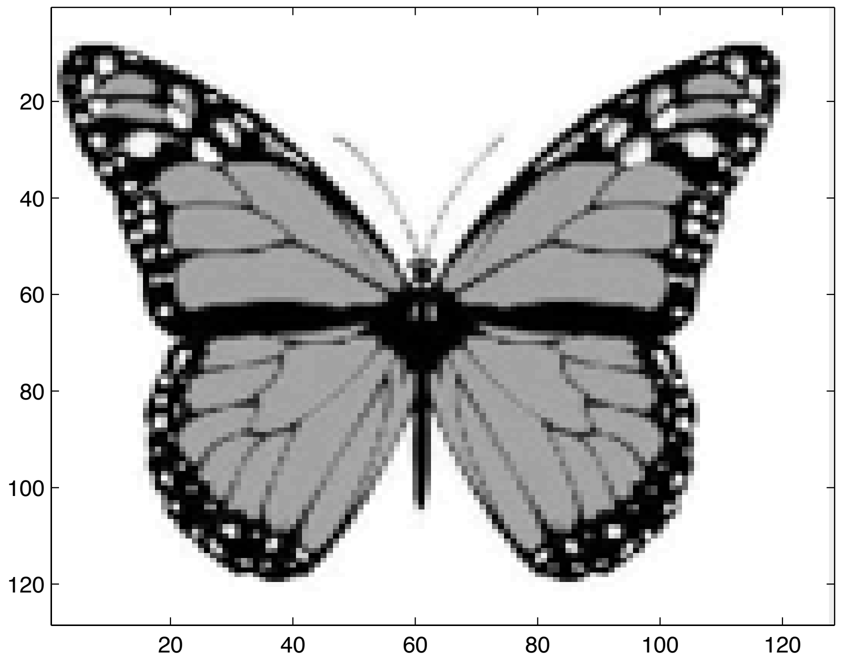
- If $f(n)$ is of size N

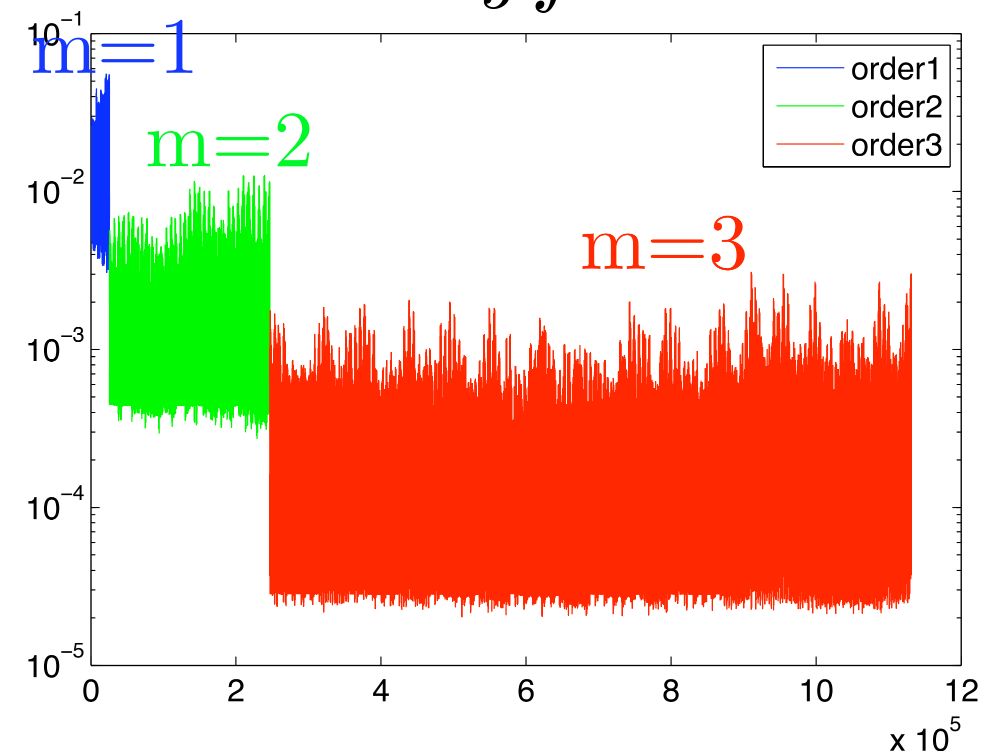
  Compute only $S_J f(2^J n)$: $2^{-2J} N$ scattering vectors.

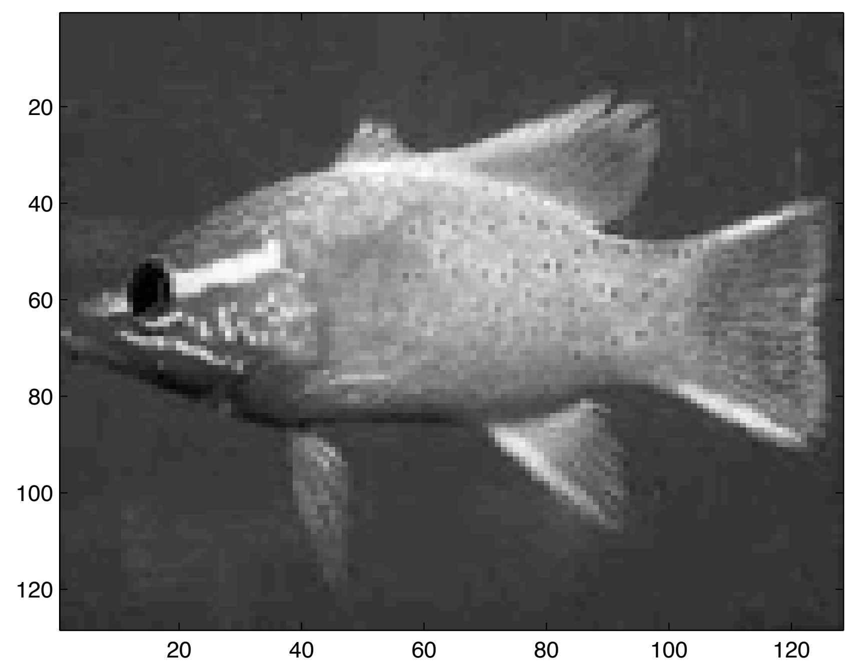
  $O(N)$ coefficients computed with $O(N \log N)$ operations.

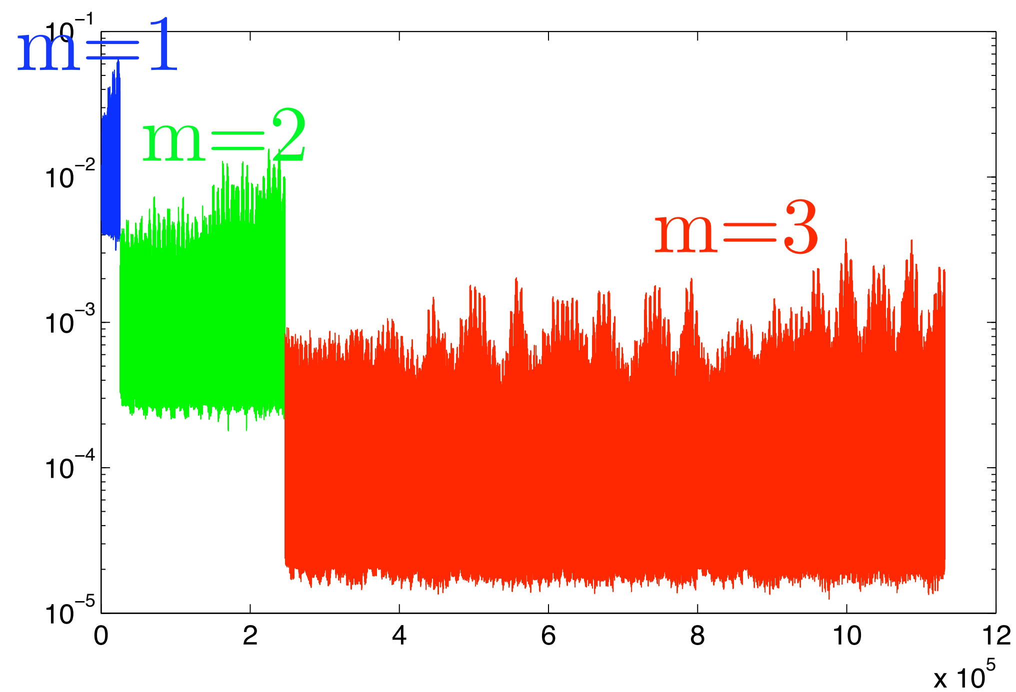
Scattering Examples

f $\quad\quad\quad$ $S_J f$









Translation Invariance

- When 2^J increases coefficients converge:

$$\lim_{J\to\infty} 2^{2J} \,||f \star \psi_{j_1,\gamma_1}|... \star \psi_{j_m,\gamma_m}| \star \phi_J(x) = \int ||f \star \psi_{j_1,\gamma_1}|... \star \psi_{j_m,\gamma_m}(u)|\, du.$$

Theorem: $\lim_{J\to\infty} \|S_J f - S_J g\|$ converges and

if $D_\tau f(x) = f(x-\tau)$ is a translation then

$$\lim_{J\to\infty} \|S_J f - S_J(D_\tau f)\| = 0\ .$$

Continuity to Deformations

Theorem If $D_\tau f(x) = f(x - \tau(x))$ with $\|\nabla \tau\|_\infty < 1$

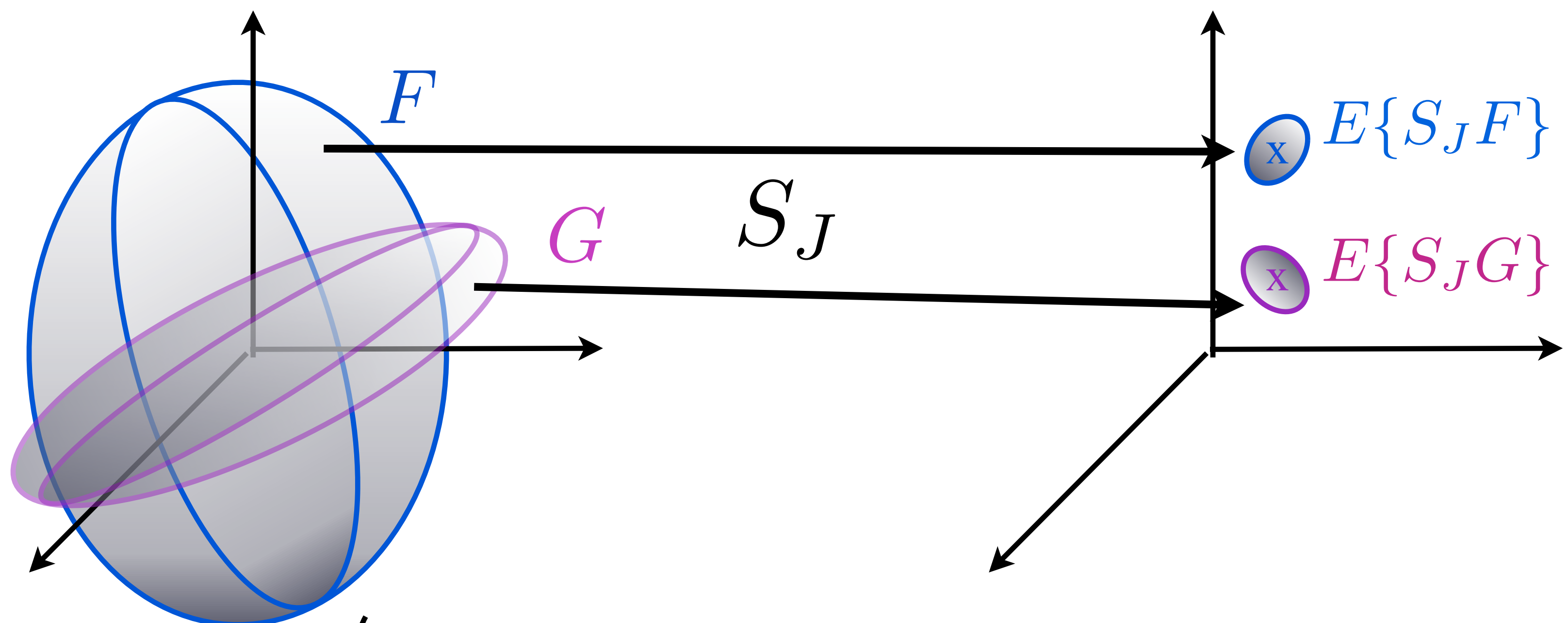
then for $J > \log \frac{\|\tau\|_\infty}{\|\nabla\tau\|_\infty}$

$$\|S_J f - S_J(D_\tau f)\| \leq C\, m\, \|f\|\, \log\left(\frac{\|\tau\|_\infty}{\|\nabla\tau\|_\infty}\right)\, \|\nabla\tau\|_\infty$$

Scattering Stationary Processes

Conjecture: for a wide class of "ergodic" stationary processes

$$\lim_{J \to \infty} \|S_J F - E\{S_J\, F\}\| = 0 \ \ : \ \text{with probability 1.}$$

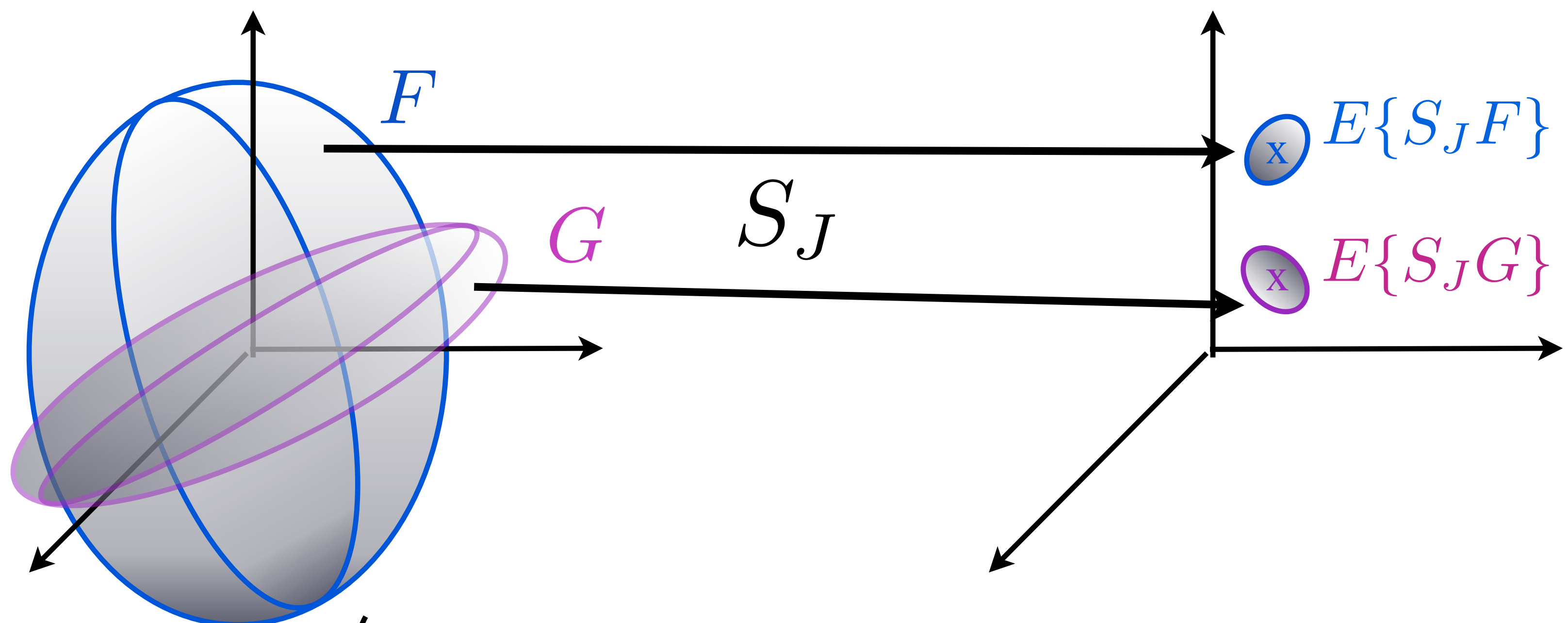


$$E\{S_J F(x)\} = \begin{pmatrix} E\{F\} \\ E\{|F \star \psi_{j_1,\gamma_1}|\} \\ \cdots \\ E\{|\,|F \star \psi_{j_1,\gamma_1}| \cdots \star \psi_{j_m,\gamma_m}|\} \end{pmatrix}_{\substack{\forall j_1 \ldots j_m \\ \forall \gamma_1 \ldots \gamma_m}}$$

Scattering Stationary Processes

Conjecture: for a wide class of "ergodic" stationary processes

$$\lim_{J\to\infty} \|S_J F - E\{S_J\, F\}\| = 0 \ \ : \ \text{with probability 1.}$$



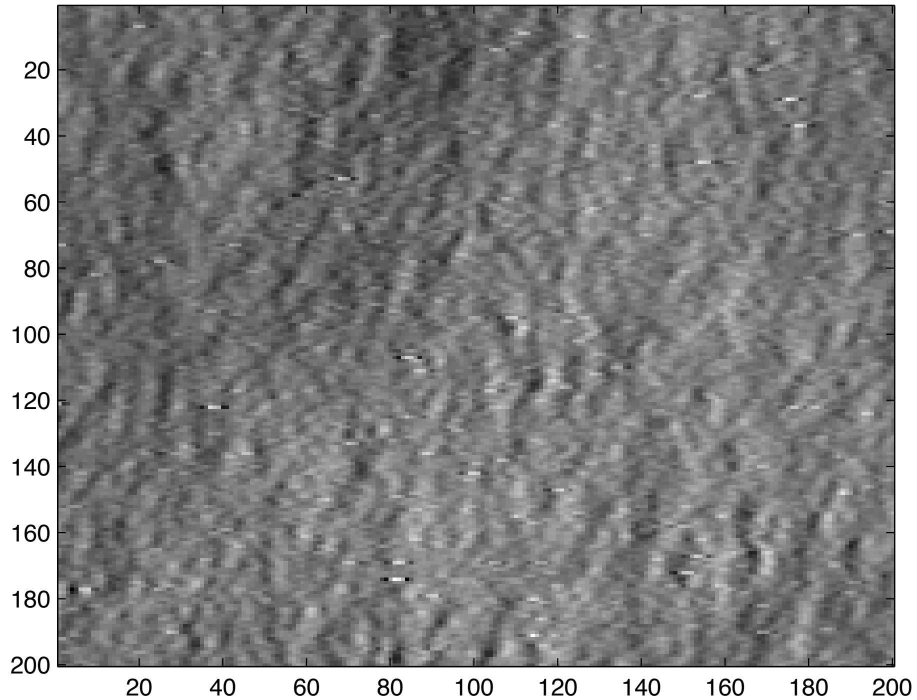
$$E\{S_J F(x)\} = \begin{pmatrix} E\{F\} \\ E\{|F\star\psi_{j_1,\gamma_1}|\} \\ \cdots \\ E\{|\,|F\star\psi_{j_1,\gamma_1}|\cdots\star\psi_{j_m,\gamma_m}|\} \end{pmatrix}_{\substack{\forall j_1\ldots j_m \\ \forall \gamma_1\ldots\gamma_m}}$$

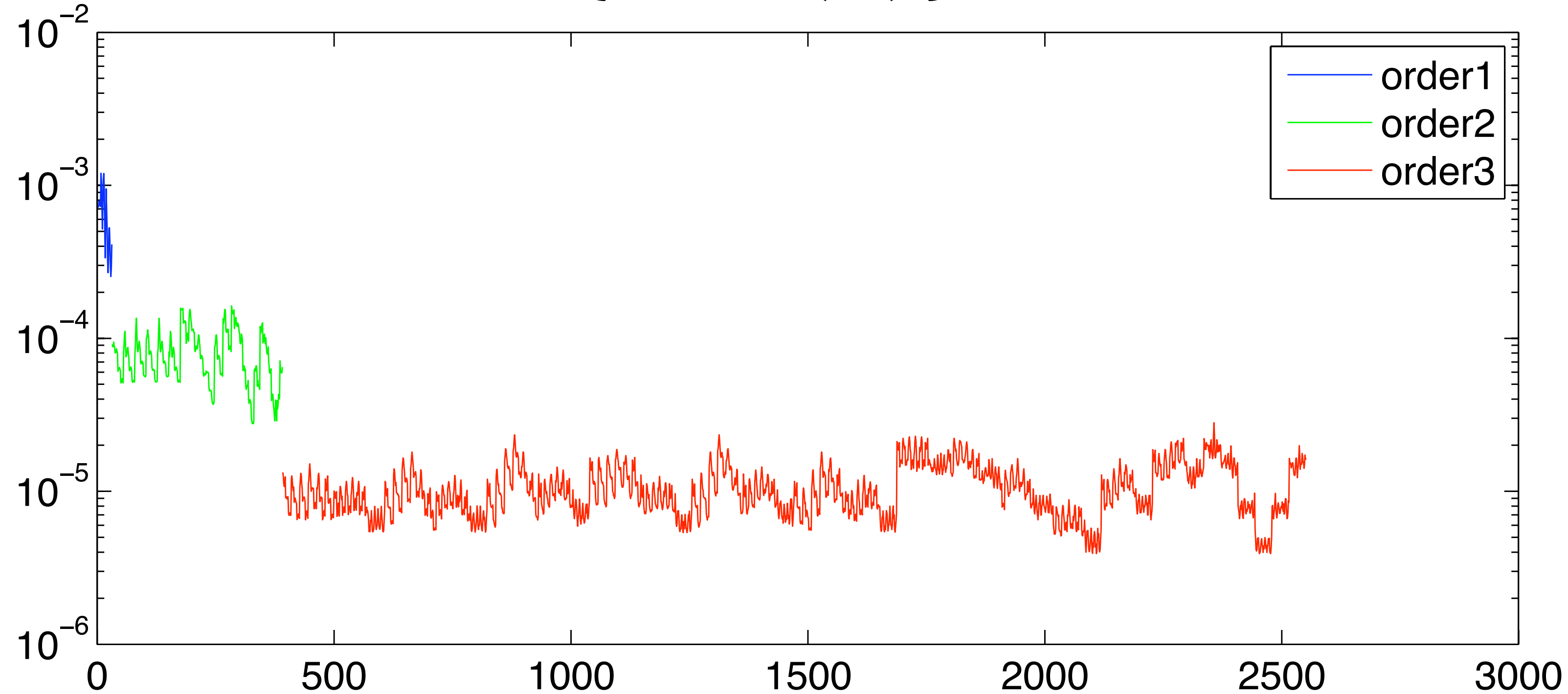
Theorem : $E\{|S_J F(x)|^2\} = E\{|F(x)|^2\}$.

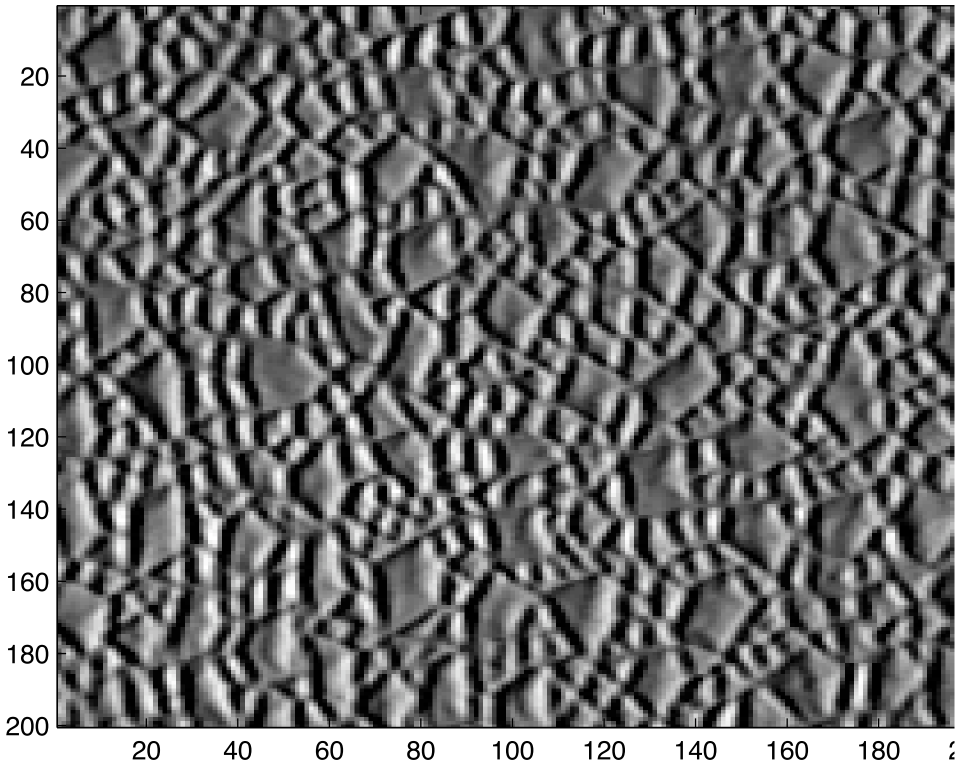
Scattering of Stationary Processes

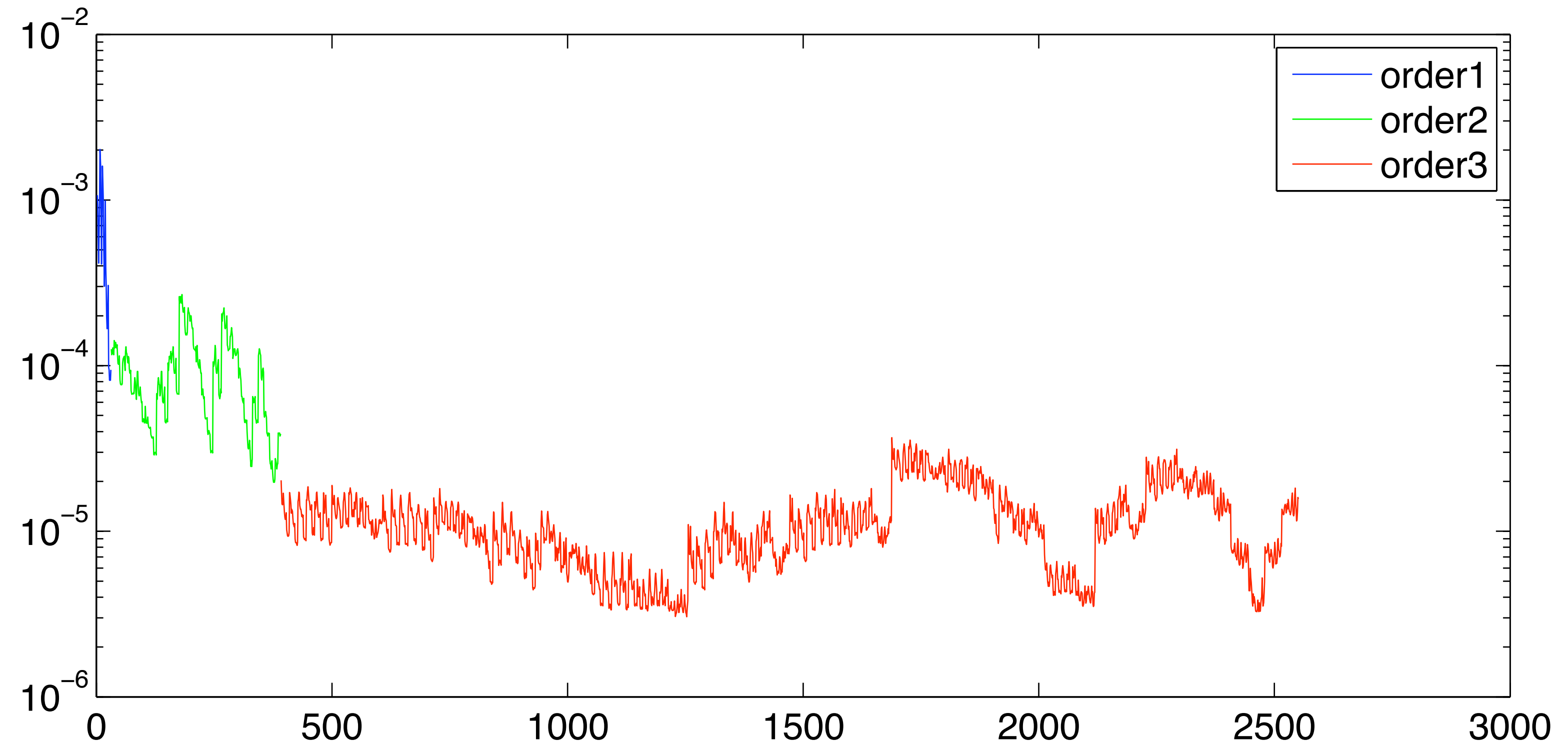
F

$E\{S_J F(x)\}$





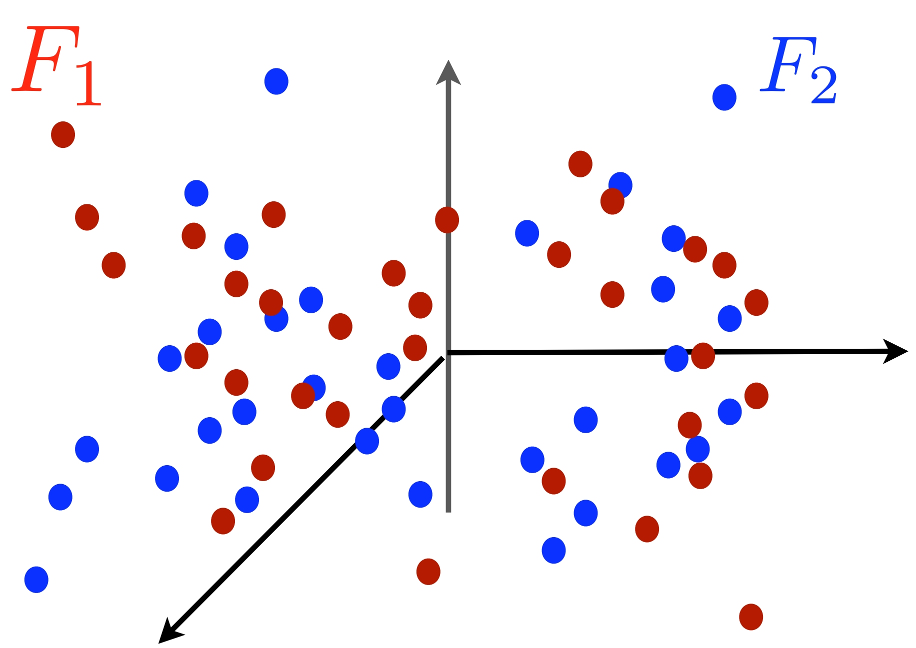


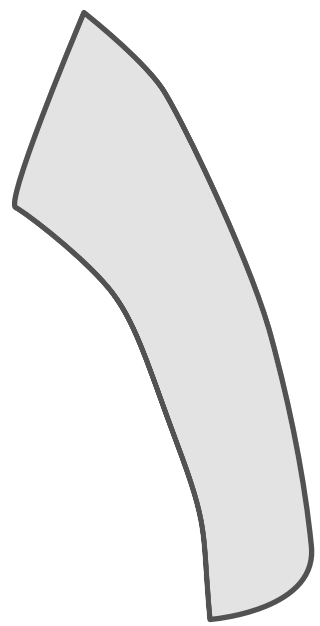


Classification : *Joan Bruna*

- K classes corresponding to K (non stationary) processes $\{F_k\}_{k \leq K}$

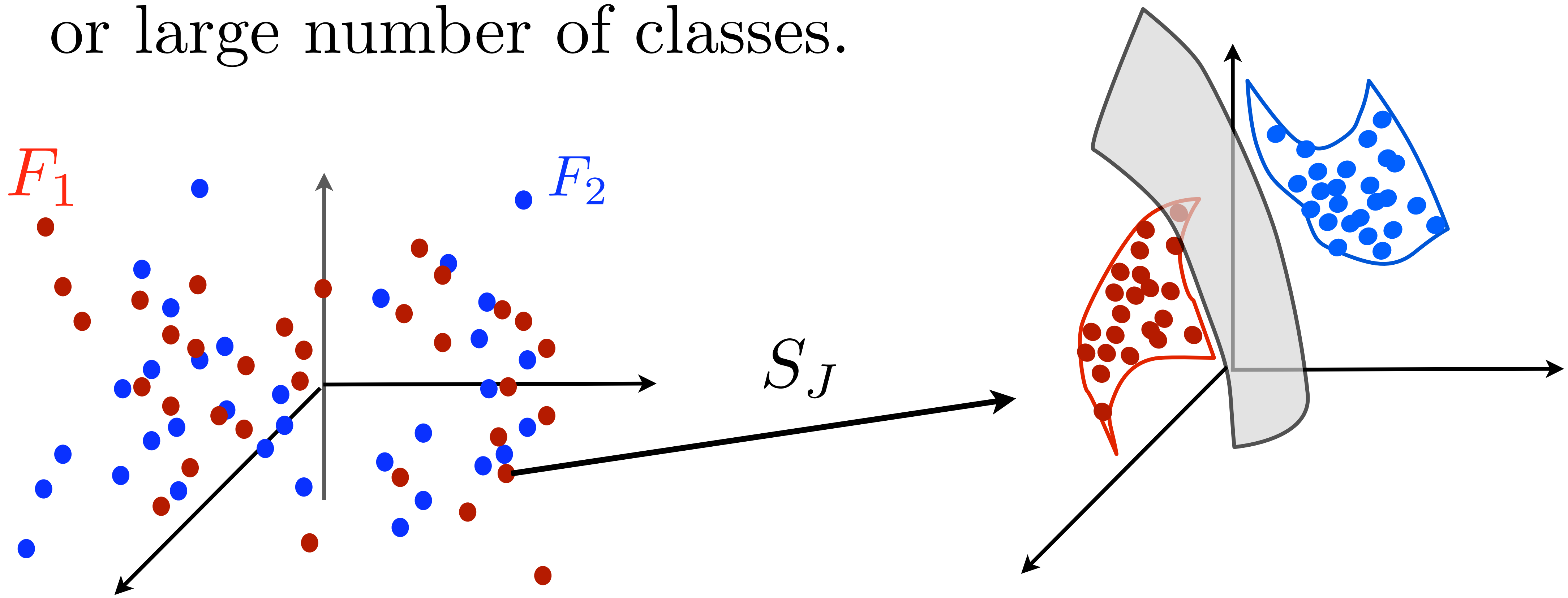
- Two possible strategies: discriminant or generative classifiers.
  - Discriminant (e.g. SVM) is asymptotically optimal.
  - Generative can be much better on small training sets or large number of classes.





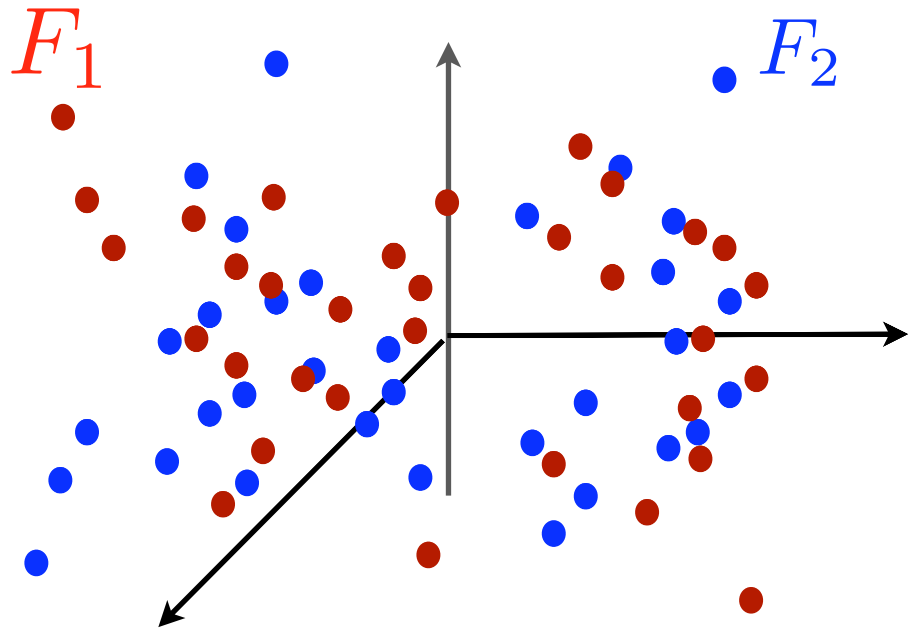
Classification : *Joan Bruna*

- K classes corresponding to K (non stationary) processes $\{F_k\}_{k \leq K}$
- Scattering transformation.
- Two possible strategies: discriminant or generative classifiers.
  - Discriminant (e.g. SVM) is asymptotically optimal.
  - Generative can be much better on small training sets or large number of classes.

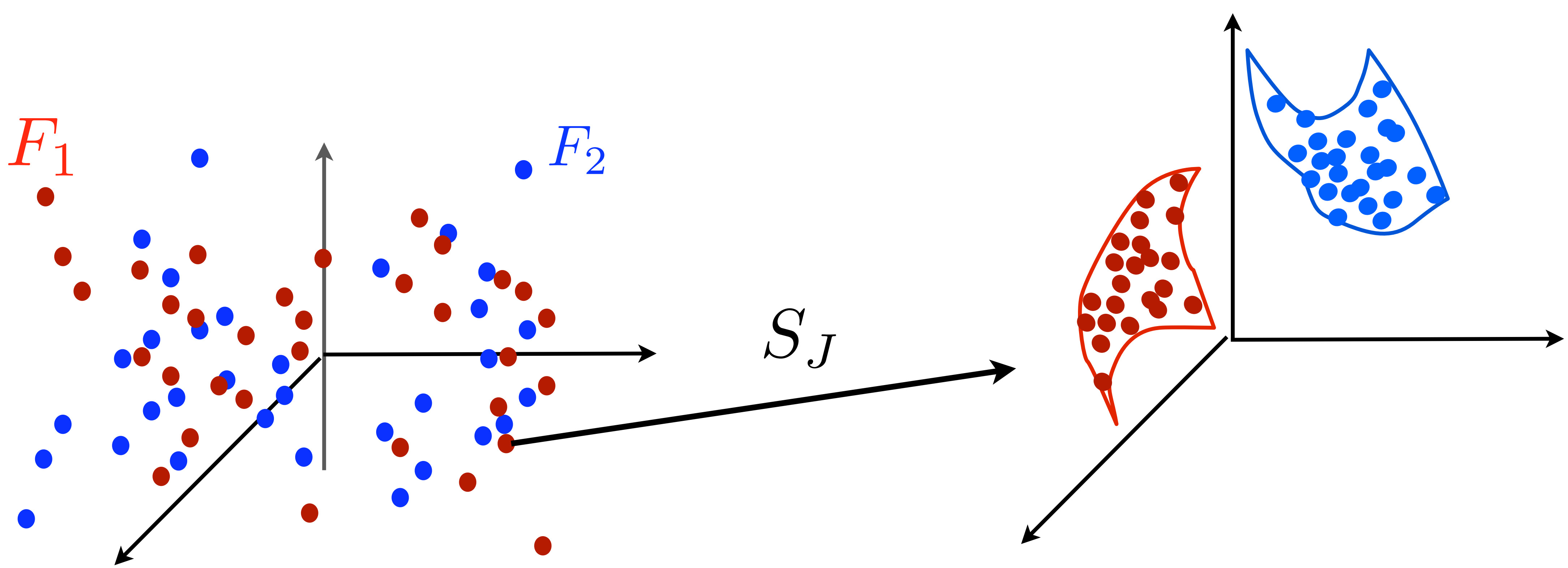


Generative: Affine Space Selection

$\{F_k\}_{k \leq K}$



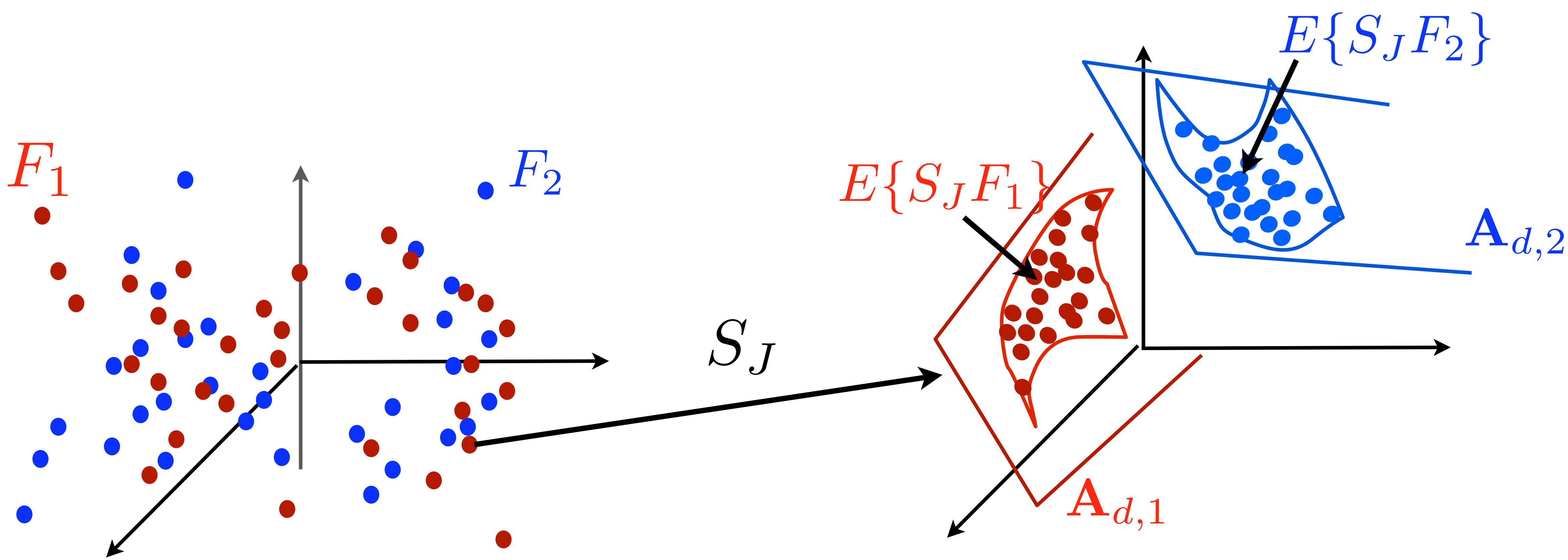
$$\{F_k\}_{k \le K}$$



Generative: Affine Space Selection

- Each class is represented by the centroid $E\{S_J F_k\}$ and $\{F_k\}_{k \leq K}$
  a space $\mathbf{V}_{d,k}$ of principal variance directions (PCA).

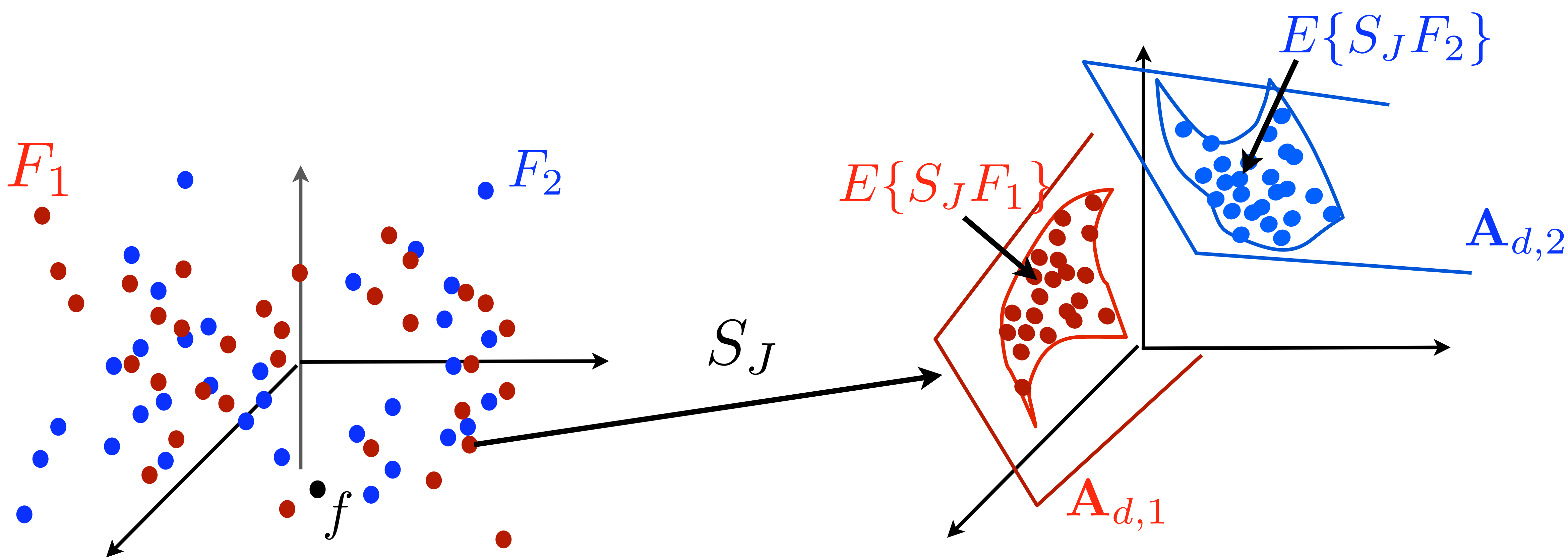
Affine space model $\mathbf{A}_{d,k} = E\{S_J F_k\} + \mathbf{V}_{d,k}$.



Generative: Affine Space Selection

- Each class is represented by the centroid $E\{S_J F_k\}$ and a space $\mathbf{V}_{d,k}$ of principal variance directions (PCA). $\{F_k\}_{k \leq K}$

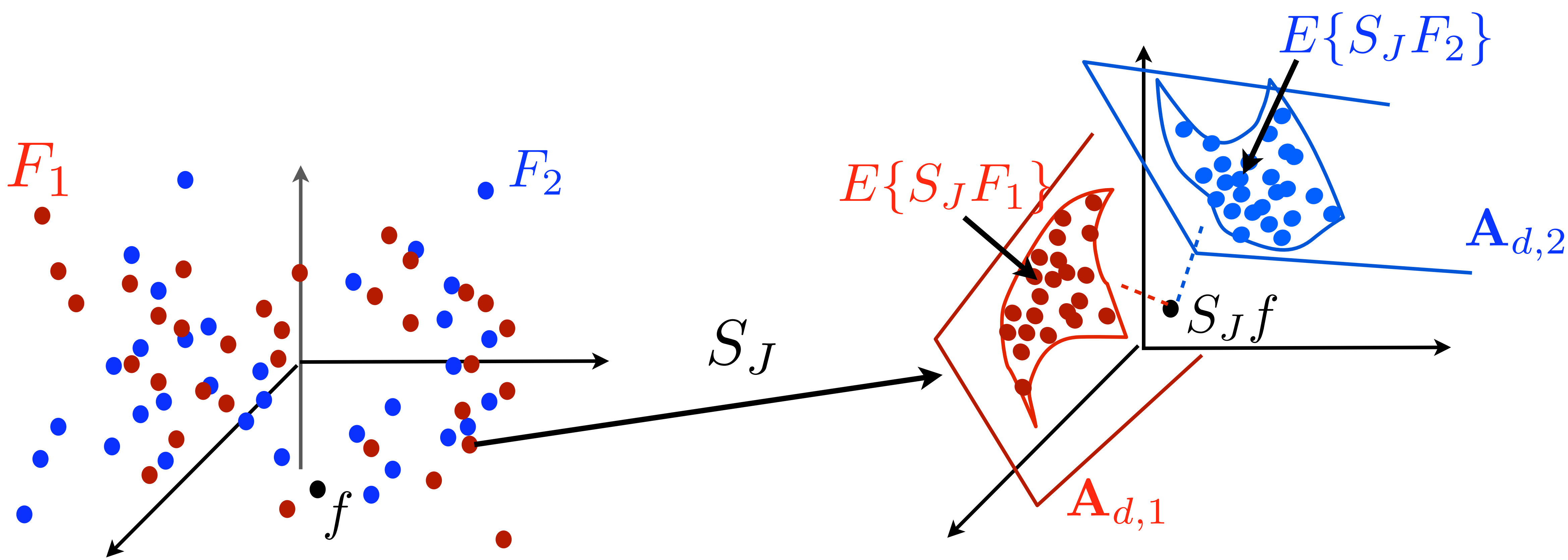
Affine space model $\mathbf{A}_{d,k} = E\{S_J F_k\} + \mathbf{V}_{d,k}$.



Generative: Affine Space Selection

- Each class is represented by the centroid $E\{S_J F_k\}$ $\{F_k\}_{k \leq K}$ and a space $\mathbf{V}_{d,k}$ of principal variance directions (PCA).

Affine space model $\mathbf{A}_{d,k} = E\{S_J F_k\} + \mathbf{V}_{d,k}$.



Scattering PCA Model Selection

- **PCA** calculation of the d dimensional spaces $\mathbf{V}_{k,d}$ of maximum variability of $S_J F_k - E\{S_J F_k\}$ from training samples of F_k

- **Classification** by best scattering affine model selection:

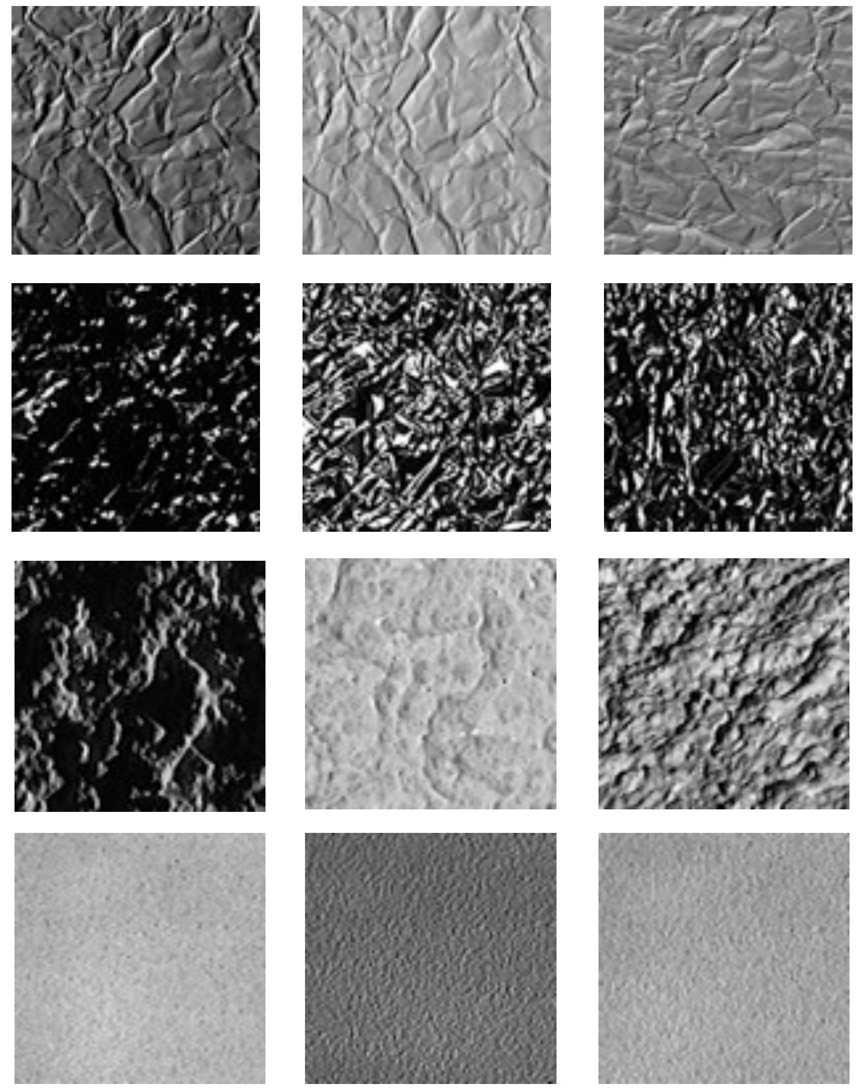
$$k(f) \quad = \quad \arg \min_{1 \leq k \leq K} \|S_J f - P_{\mathbf{A}_{k,d}}(S_J f)\| \ .$$

- **Cross-validation:**
  - d : dimension of the variability reduction.
  - J : maximum scattering scale.

Classification of Textures

CUREt database

61 classes



Rotations and illumination variations.

Scattering $J = \log_2 N$

Training per class	PCA $m = 2$	SVM $m = 2$	Mark. Rand.
23	**0.9%**	3.3%	22.43%
46	**0.09%**	1.1%	2.46%

Non-Gaussian Process Characterization

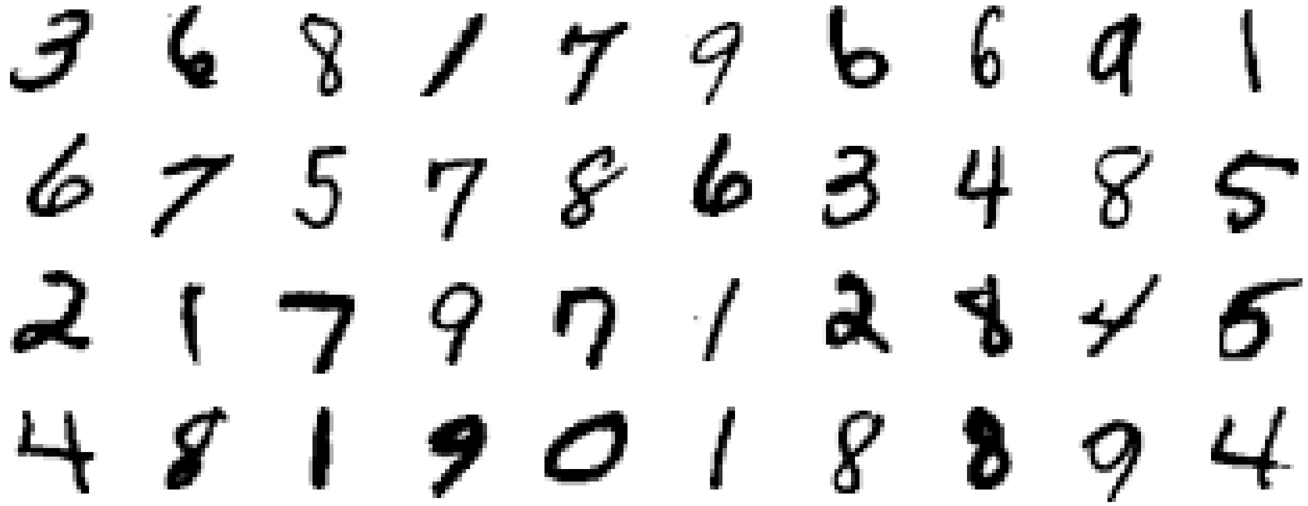
- Usual approaches use high order moments: large variance estimators. Not enough training samples.

- Non-gaussian process models with first and second order moments of scattering coefficients: co-occurrence information (Bela Julesz conjecture).

- Effective for audio classification: characterizes attacks, beating...

- What are the properties of these stochastic models ?

Digit Classification: MNIST



Scattering with $J = 3$

Training Size	Conv. Net.	PCA $m = 2$	Space dim. d
300	7.18	**6.05**	24
5000	1.52	**1.22**	40
40000	**0.65**	0.78	180

Combined Scattering

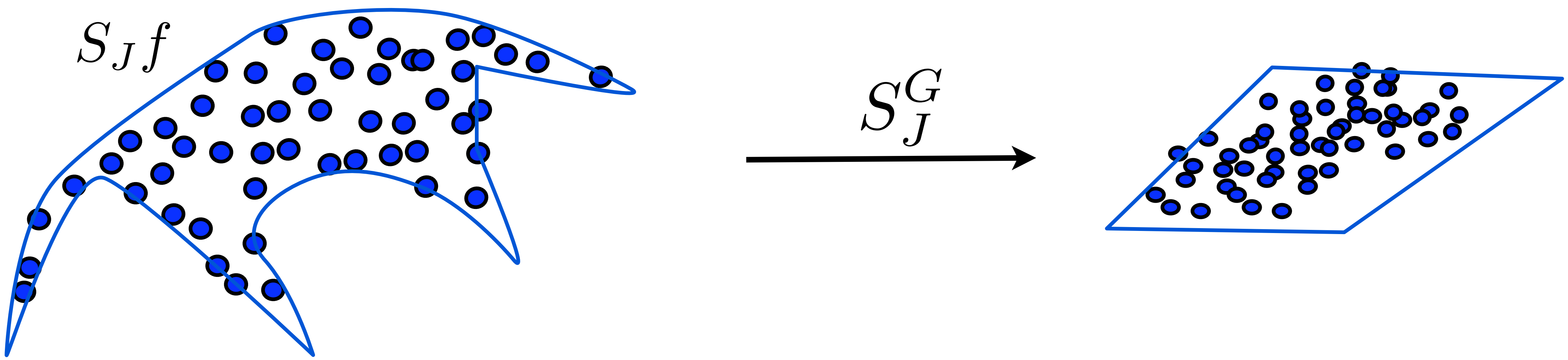
- Translation group scattering: not sufficient for complex classes

Combined Scattering

- Translation group scattering: not sufficient for complex classes

  - Intra-class variability need to be further reduced:

$$f \longrightarrow \boxed{S_J^{\text{Trans}}} \longrightarrow \boxed{S_{J'}^{G}} \longrightarrow \ldots$$

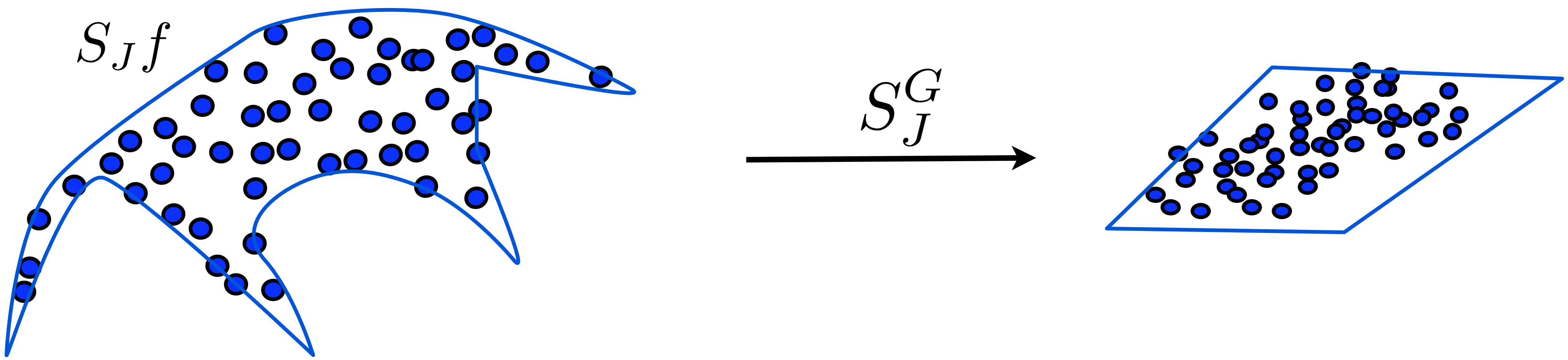


Combined Scattering

- Translation group scattering: not sufficient for complex classes

- Intra-class variability need to be further reduced:

$$f \longrightarrow \boxed{S_J^{\mathrm{Trans}}} \longrightarrow \boxed{S_{J'}^{G}} \longrightarrow \ldots$$



- Scattering $S_{J'}^{G}$ over a compact Lie group G with iterated wavelet tranforms over G cascaded with modulus operators.

Curvature reduction with iterated contractions.

Conclusion

- High dimensional signal classification strategy by reducing intra-class variability with iterated contractions.

- A multiscale scattering is invariant, Lipschitz continuous to deformations and informative. How to do it otherwise ?

- Important for image and audio perception: neurophysiology.

- Papers/softwares: www.cmap.polytechnique.fr/scattering