Stat 212b: Topics in Deep Learning
Lecture 4

Joan Bruna
UC Berkeley

DeepMind makes Nature Cover

DeepMind designed an algorithm that beat a professional GO player for the first time, using MCTS and two CNNs trained with supervised learning and reinforcement learning.

[http://www.nature.com/news/google-ai-algorithm-masters-ancient-game-of-go-1.19234]

- Thus $\Phi(x) = |Vx|$ satisfies

$$\boxed{\forall\ x, t\ ,\ \ \Phi(\varphi_t(x)) = \Phi(x)\ .}$$

- Indeed,

$$A = V^* \mathrm{diag}(\lambda_1, \ldots, \lambda_n) V \implies e^{itA} = V^* \mathrm{diag}(e^{it\lambda_1}, \ldots, e^{it\lambda_n}) V\ .$$

$$\begin{aligned} V\varphi_t x &= V e^{itA} x = V V^* \mathrm{diag}(e^{it\lambda_1}, \ldots, e^{it\lambda_n}) V x \\ &= \mathrm{diag}(e^{it\lambda_1}, \ldots, e^{it\lambda_n}) V x \end{aligned}$$

$$\text{thus } \Phi(\varphi_t x) = |V\varphi_t x| = |Vx|\ .$$

Review: Limits of Group Diagonalisation

- A shallow (1 layer) network is thus sufficient to achieve invariance to commutative group transformations:

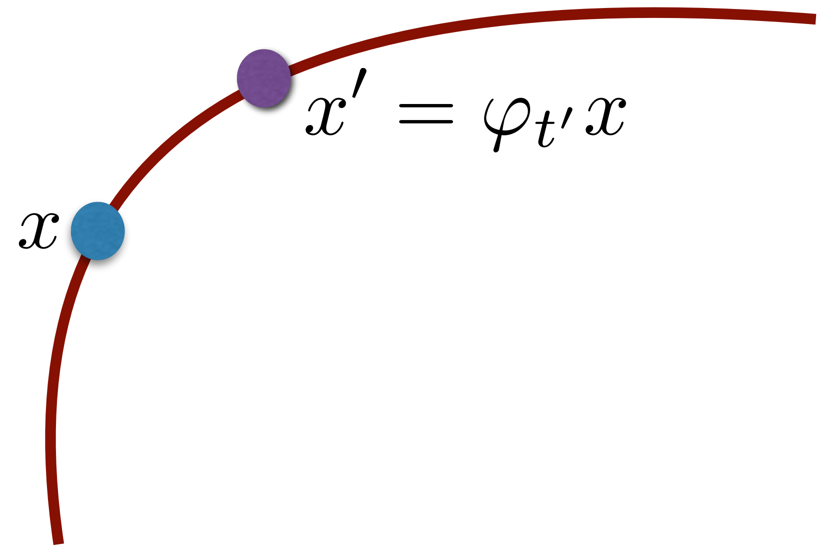
$$x \longrightarrow \boxed{V} \longrightarrow \boxed{|z|} \longrightarrow \Phi(x)$$

- However, this architecture has a number of shortcomings.
    - Not applicable to non-commutative, discrete symmetry groups
    - Not discriminative in general
    - Not stable

Objectives

- Wavelets
- Point-Wise non-linearities
- Scattering Representations for the Translation Group
- Properties

Local invariants and convolution

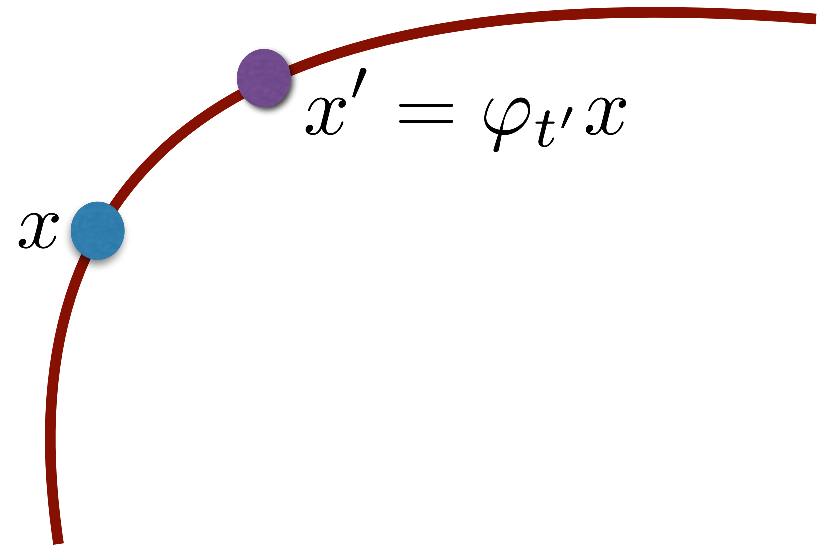


- Local translation invariance:

$$\|\Phi(x) - \Phi(\varphi_v x)\| \leq C 2^{-J} \|v\| \text{ , or}$$

$$\forall\ v, \|x\| = 1\ ,\ \frac{\|\Phi(x) - \Phi(\varphi_v x)\|}{\|v\|} \leq C 2^{-J}\ .$$

Local invariants and convolution



- Local translation invariance:

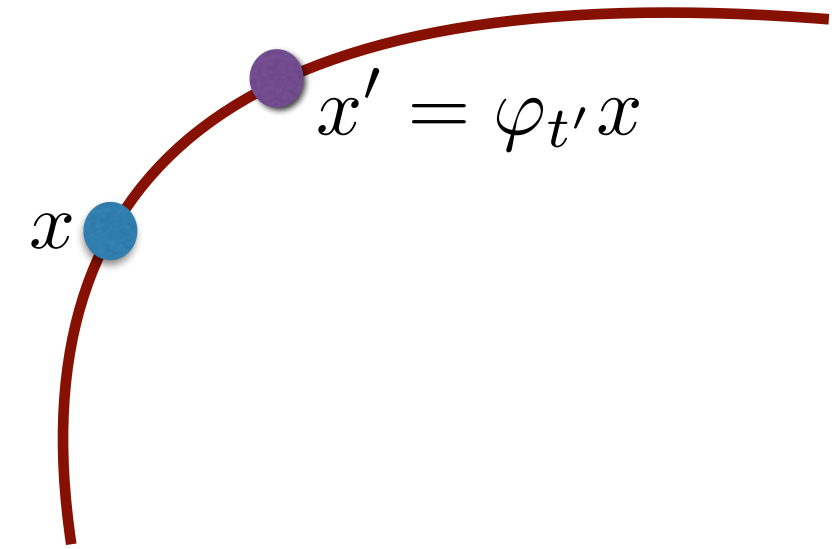
$$\|\Phi(x) - \Phi(\varphi_v x)\| \leq C 2^{-J} \|v\| \text{ , or}$$

$$\forall\ v, \|x\| = 1\ ,\ \frac{\|\Phi(x) - \Phi(\varphi_v x)\|}{\|v\|} \leq C 2^{-J}\ .$$

- So, we want to *smooth* along the orbits.
- Local averaging within the translation orbit:

$$\Phi(x) = 2^{-dJ} \int_v \phi(2^{-J} v) \varphi_v x dv\ ,\ \left(\int \phi(v) dv = 1, \phi \geq 0 \right)\ .$$

Local invariants and convolution



- Local averaging within the translation orbit:

$$\Phi(x) = 2^{-dJ} \int_v \phi(2^{-J} v) \varphi_v x dv \ , \quad \left(\int \phi(v) dv = 1, \phi \geq 0 \right) \ .$$

- In coordinates, it becomes

$$\Phi(x)(u) = \int \phi_J(v) x(u-v) dv = x * \phi_J(u) \ , \text{with}$$

$$\phi_J(v) = 2^{-Jd} \phi(2^{-J} v)$$

Local average and stability

Proposition: The local averaging $\Phi(x) = x * \phi_J$ satisfies
$\forall \ \|x\| = 1 \in L^2 \ , \tau \ , \ \|\Phi(x) - \Phi(\varphi_\tau x)\| \leq C \|\tau\| \ .$

Local average and stability

Proposition: The local averaging $\Phi(x) = x * \phi_J$ satisfies $\forall \, \|x\| = 1 \in L^2 \,,\, \tau \,,\; \|\Phi(x) - \Phi(\varphi_\tau x)\| \leq C \|\tau\| \;.$

- Not surprising, since this operator removes the problematic high-frequencies.
- Are there other linear operators with the same property?

Average and uniqueness

- The only linear, translation-invariant operator is the average:

$$\forall\ v\ ,\ \ \Phi(x) = \Phi(\varphi_v x) \Longrightarrow \Phi(x) = \frac{1}{|G|}\int \Phi(\varphi_v x) dv$$

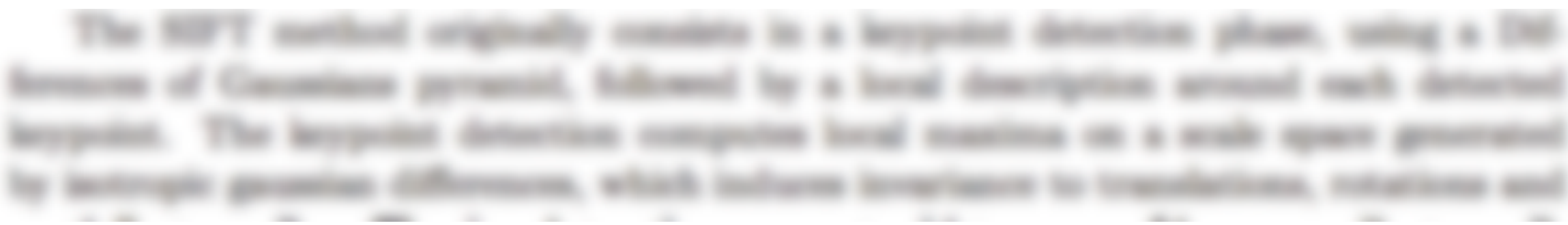
$$\Longrightarrow \Phi(x) = \Phi\left(\frac{1}{|G|}\int \varphi_v x dv\right) = \Phi\left(\frac{1}{|G|}\int x(u) du\right)\ .$$

- And a similar argument can be used locally.

From averages to Wavelets

- Low-pass information is insufficient:

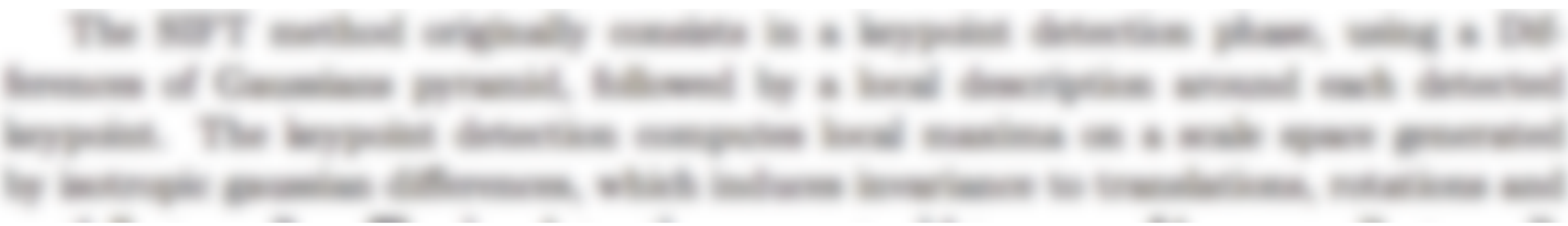
The SIFT method originally consists in a keypoint detection phase, using a Differences of Gaussians pyramid, followed by a local description around each detected keypoint. The keypoint detection computes local maxima on a scale space generated by isotropic gaussian differences, which induces invariance to translations, rotations and



From averages to Wavelets

- Low-pass information is insufficient:

> The SIFT method originally consists in a keypoint detection phase, using a Differences of Gaussians pyramid, followed by a local description around each detected keypoint. The keypoint detection computes local maxima on a scale space generated by isotropic gaussian differences, which induces invariance to translations, rotations and

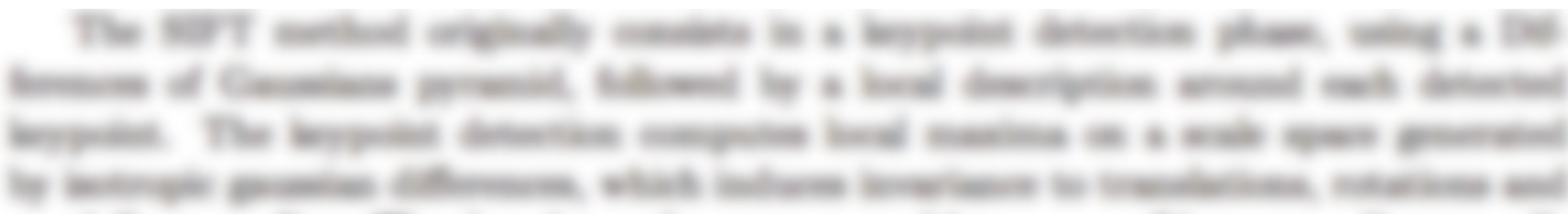


- Thus, we must capture high-frequency.
- These new measurements must involve a non-linearity.

From averages to Wavelets

- Low-pass information is insufficient:

> The SIFT method originally consists in a keypoint detection phase, using a Differences of Gaussians pyramid, followed by a local description around each detected keypoint. The keypoint detection computes local maxima on a scale space generated by isotropic gaussian differences, which induces invariance to translations, rotations and

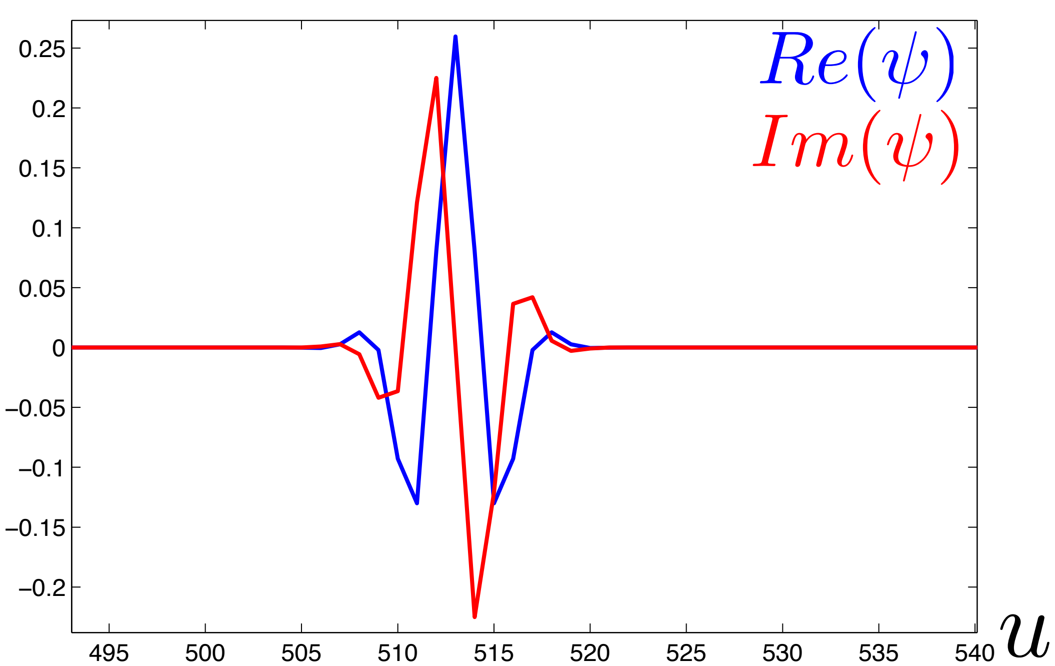


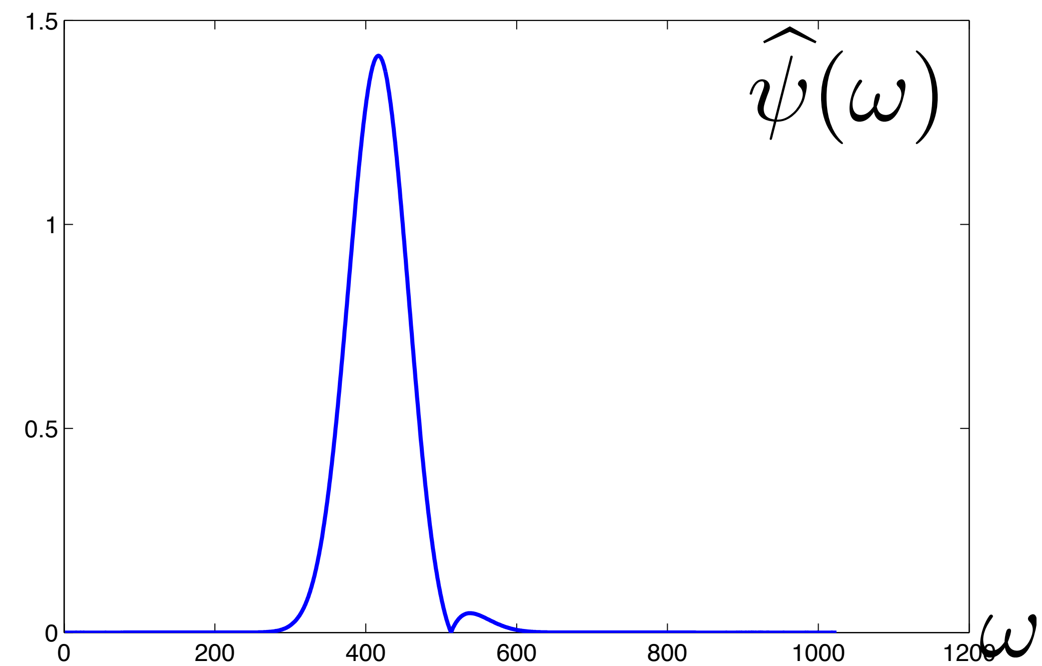
- Thus, we must capture high-frequency.
- These new measurements must involve a non-linearity.
- We want them to preserve stability to deformations.
- And we want them to preserve inter-class variability.

Wavelets

- ψ : bandpass (ie oscillating) signal, well localized in space and frequency.

Ex: Morlet wavelet



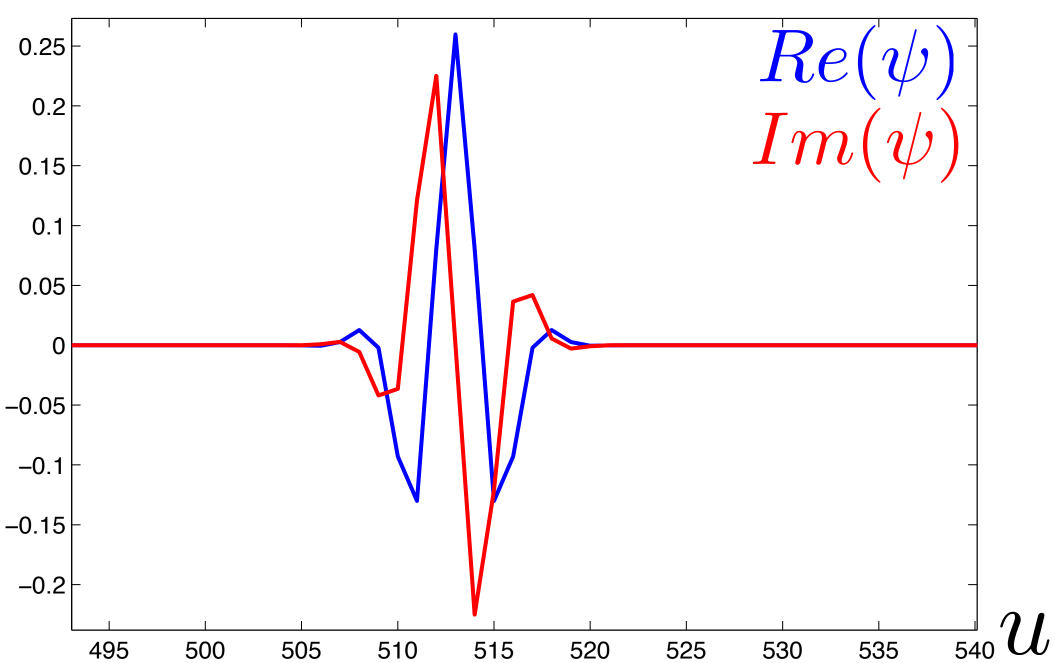


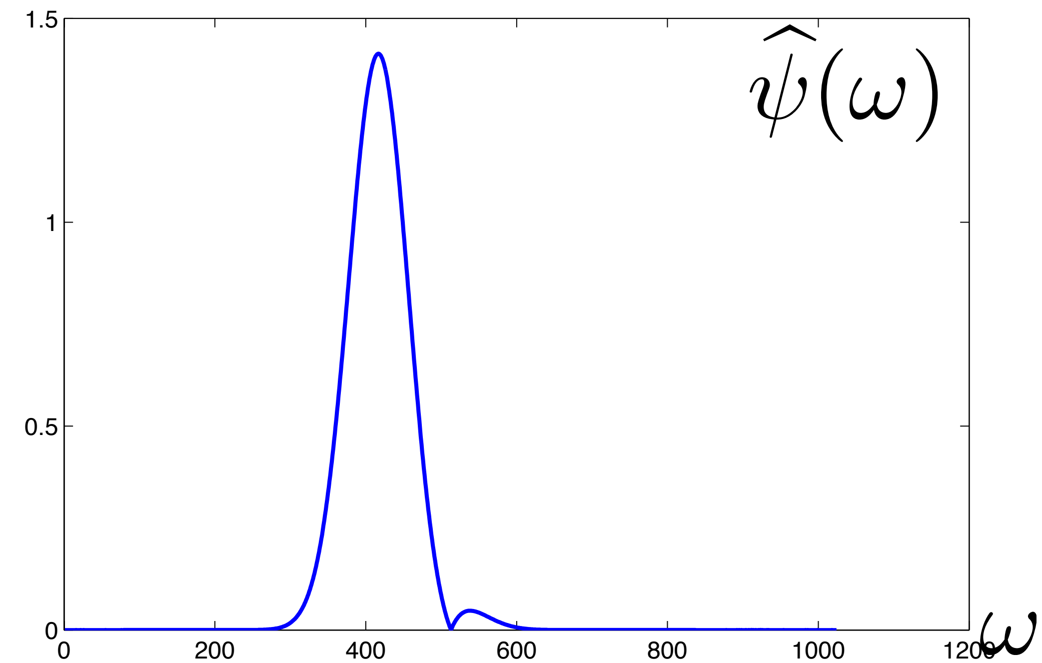
Wavelets

- ψ : bandpass (ie oscillating) signal, well localized in space and frequency.
- At least one vanishing moment: $\int \psi(u) du = 0$

(we say that ψ has k vanishing moments if $\int \psi(u) u^l du = 0$ for $l < k$)

Ex: Morlet wavelet





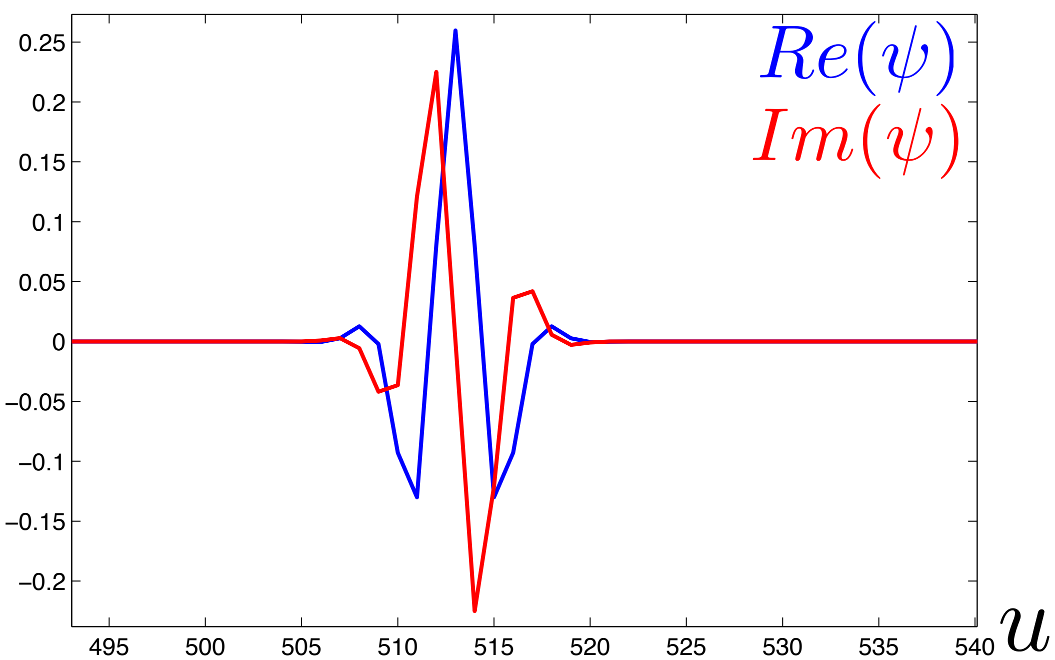
Wavelets

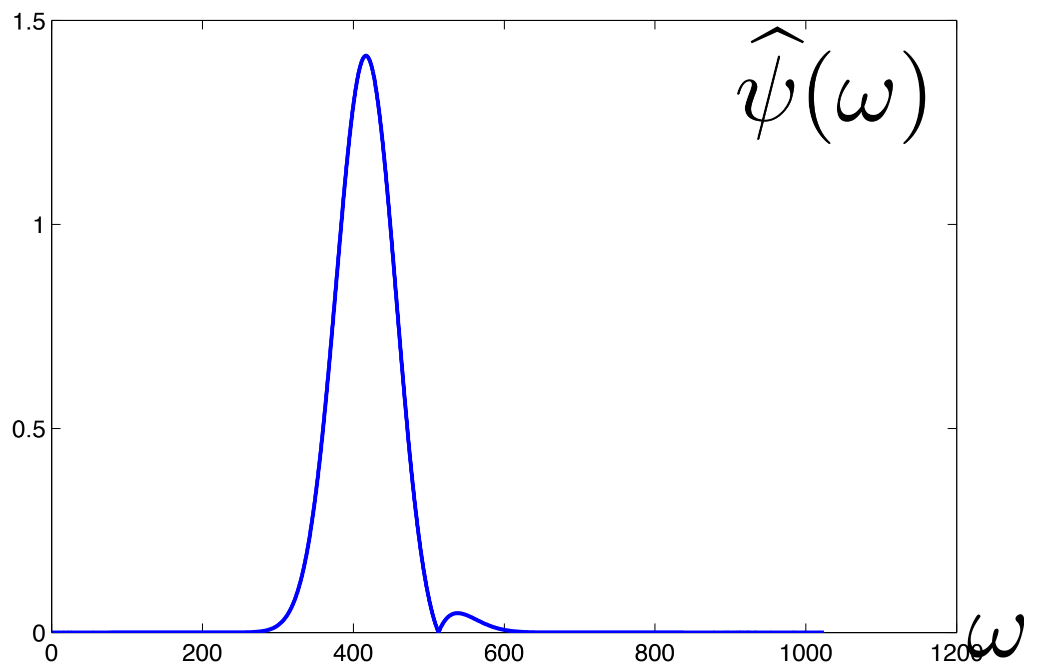
- ψ : bandpass (ie oscillating) signal, well localized in space and frequency.
- At least one vanishing moment: $\int \psi(u)du = 0$

(we say that ψ has k vanishing moments if $\int \psi(u)u^l du = 0$ for $l < k$)

- Can be real or complex. $\psi = \psi_r + i\psi_i$

Ex: Morlet wavelet



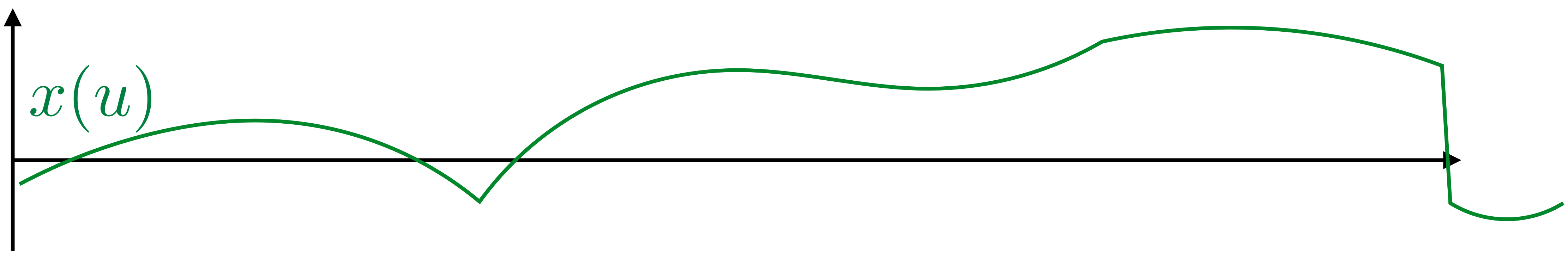


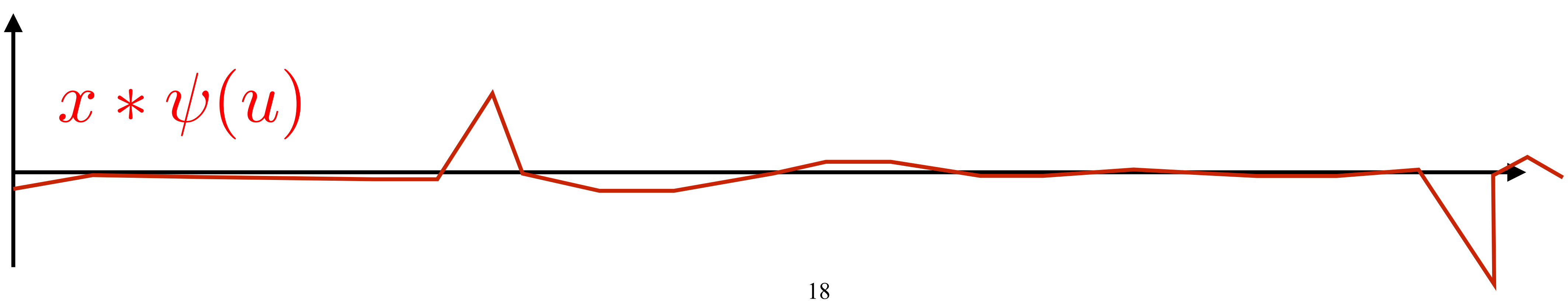
Wavelets

- ψ : bandpass (ie oscillating) signal, well localized in space and frequency.
- At least one vanishing moment: $\int \psi(u)du = 0$

(we say that ψ has k vanishing moments if $\int \psi(u)u^l du = 0$ for $l < k$)

- If $x(u)$ is piece-wise smooth, then $x * \psi(u)$ is mostly zero





Wavelets

- The local average $x * \phi$ is a blurry version of x, whereas
- $x * \psi$ carries the details lost by the blurring.

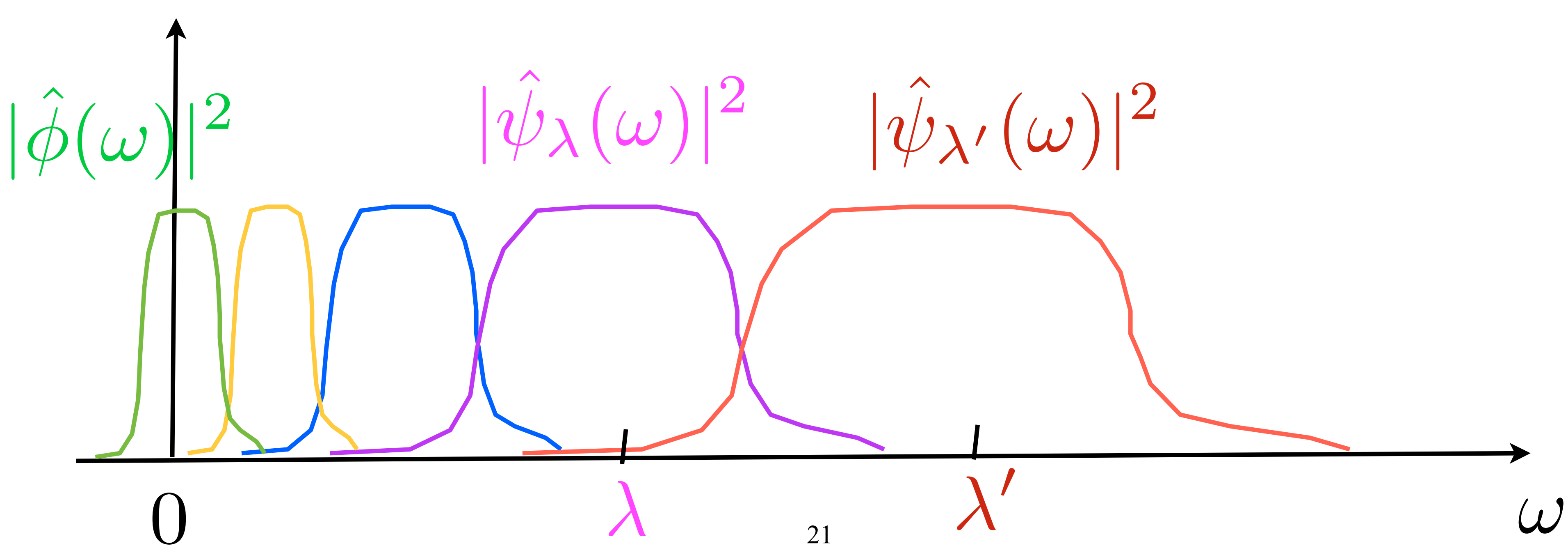
Wavelets

- The local average $x * \phi$ is a blurry version of x, whereas
- $x * \psi$ carries the details lost by the blurring.
- The details are relative to a given resolution. How to obtain a decomposition that captures details at *all* resolutions?

Wavelets

- The local average $x * \phi$ is a blurry version of x, whereas
- $x * \psi$ carries the details lost by the blurring.
- The details are relative to a given resolution. How to obtain a decomposition that captures details at *all* resolutions?

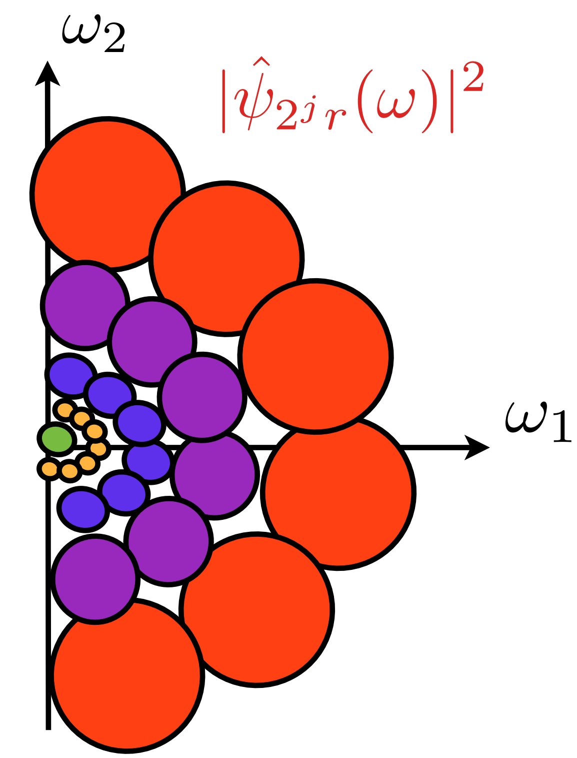
- Dilated wavelets: $\psi_j(u) = 2^{-j}\psi(2^{-j}u)\,,\ j \in \mathbb{Z}$



Littlewood-Paley Wavelet Filter Banks

- For images, dilated and rotated wavelets:

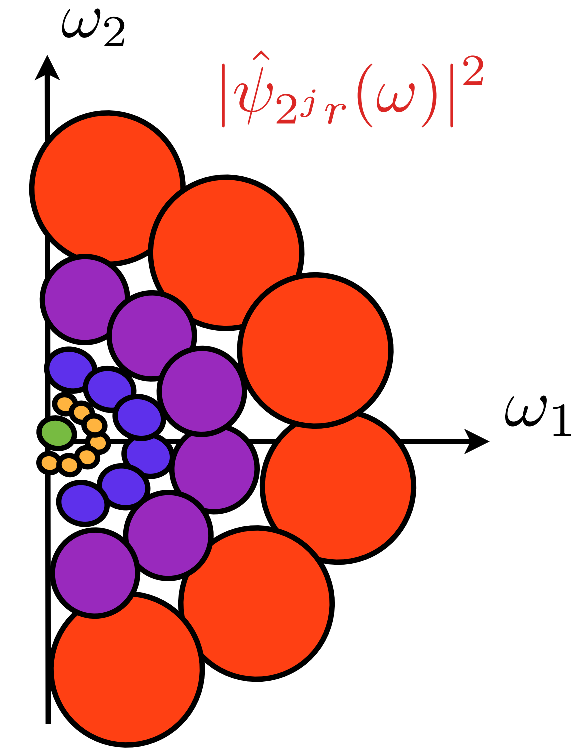
$$\psi_\lambda(u) = 2^{-j/2}\psi(2^{-j}ru) \ , \text{ with } \lambda = 2^j r$$



Littlewood-Paley Wavelet Filter Banks

- For images, dilated and rotated wavelets:

$$\psi_\lambda(u) = 2^{-j/2}\psi(2^{-j}ru)\ ,\ \text{with } \lambda = 2^j r$$



- Wavelet transform convolutional filter bank:

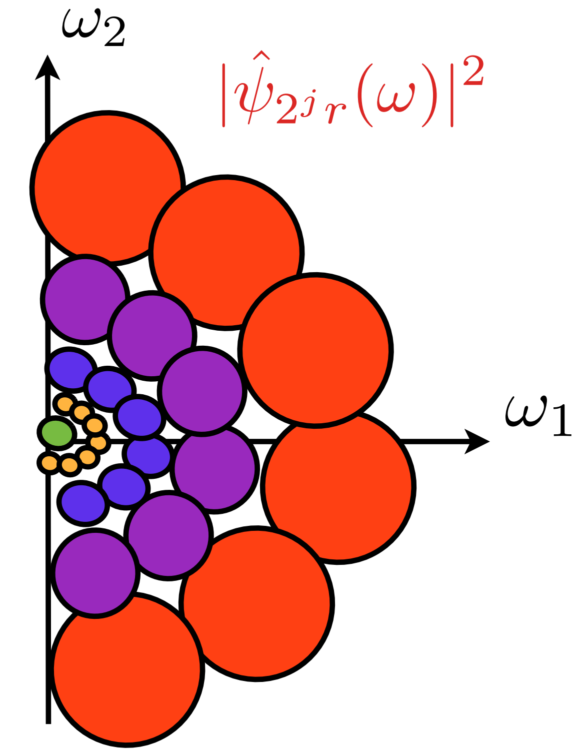
$$Wx = \{x \star \phi(u)\ ,\ x \star \psi_\lambda(u)\}_{\lambda \in \Lambda}$$

$$x \star \psi(u) = \int x(v)\psi(u-v)dv\ .$$

Littlewood-Paley Wavelet Filter Banks

- For images, dilated and rotated wavelets:

$$\psi_\lambda(u) = 2^{-j/2}\psi(2^{-j}ru)\ ,\ \text{with } \lambda = 2^j r$$



- Wavelet transform convolutional filter bank:

$$Wx = \{x \star \phi(u)\ ,\ x \star \psi_\lambda(u)\}_{\lambda\in\Lambda}$$

$$x \star \psi(u) = \int x(v)\psi(u-v)dv\ .$$

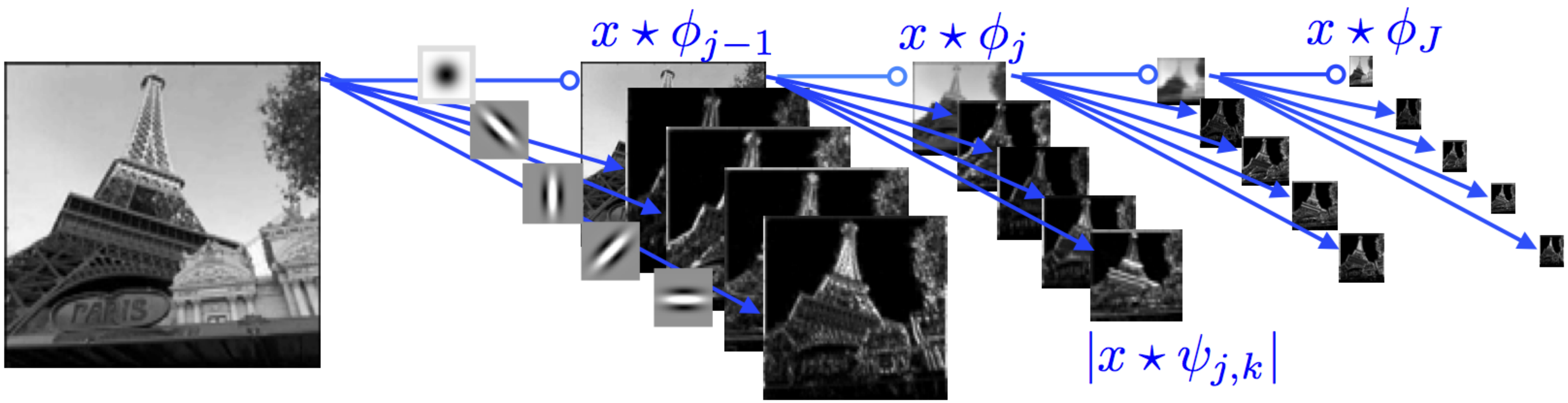
Theorem (Littlewood-Paley): If there exists $\delta > 0$ such that

$$\forall \omega > 0\ ,\ 1-\delta \leq |\hat{\phi}(\omega)|^2 + \frac{1}{2}\sum_\lambda |\hat{\psi}(\lambda^{-1}\omega)|^2 \leq 1\ ,$$

then $\forall\, x \in L^2$, $(1-\delta)\|x\|^2 \leq \|Wx\|^2 \leq \|x\|^2$.

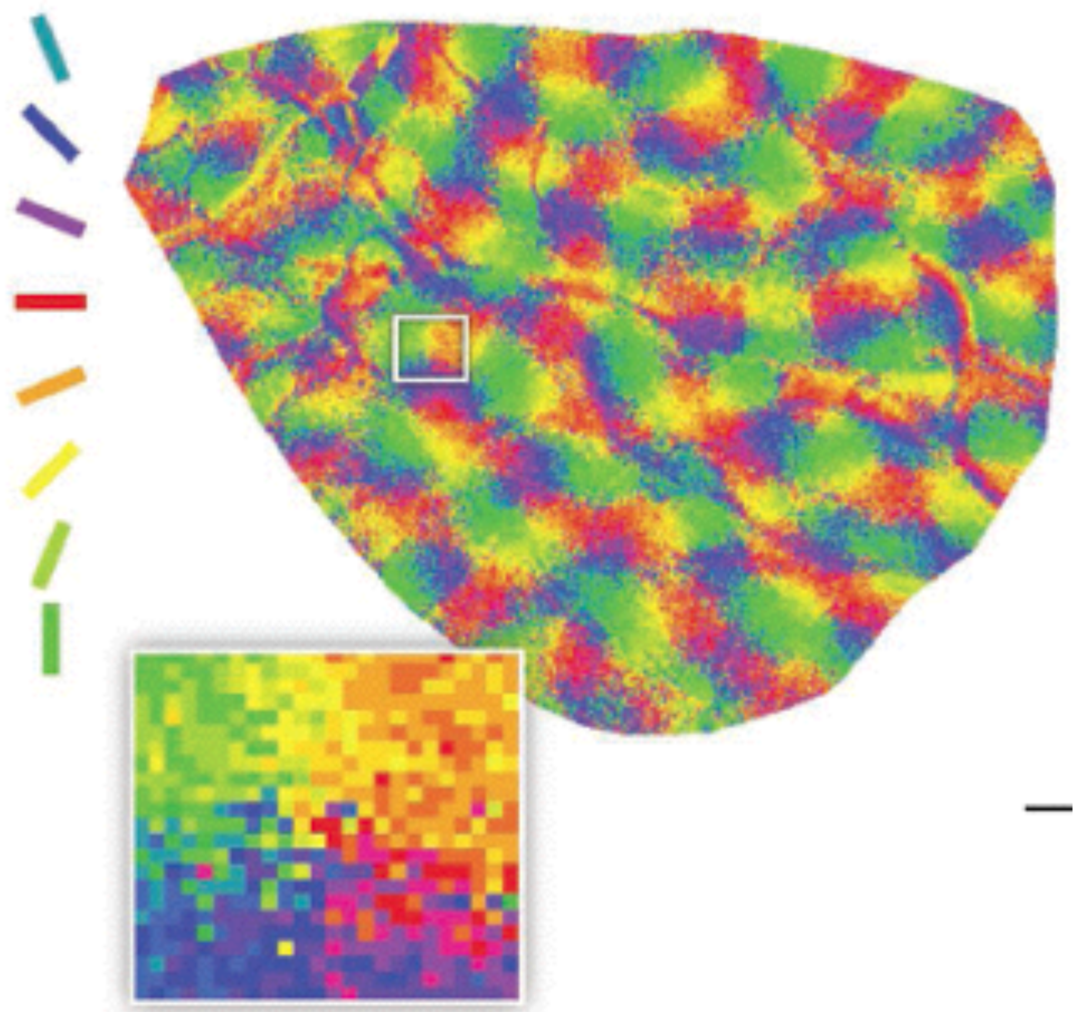
Wavelet Filter Banks

- We can compute a wavelets recursively in a fine-to-coarse transform:



Wavelets in Vision

- *V1* Model of Simple and Complex cells: First layer of processing is selective in orientation, scale and position.



  - cells are organized in *pinwheels*. (more on that later).

Why are wavelets a good model?

- We will see that they provide stability to deformations because they *commute* nicely with diffeomorphisms:

$$\|W\varphi_\tau x - \varphi_\tau W x\| \lesssim \|\tau\| \ .$$

Why are wavelets a good model?

- We will see that they provide stability to deformations because they *commute* nicely with diffeomorphisms:

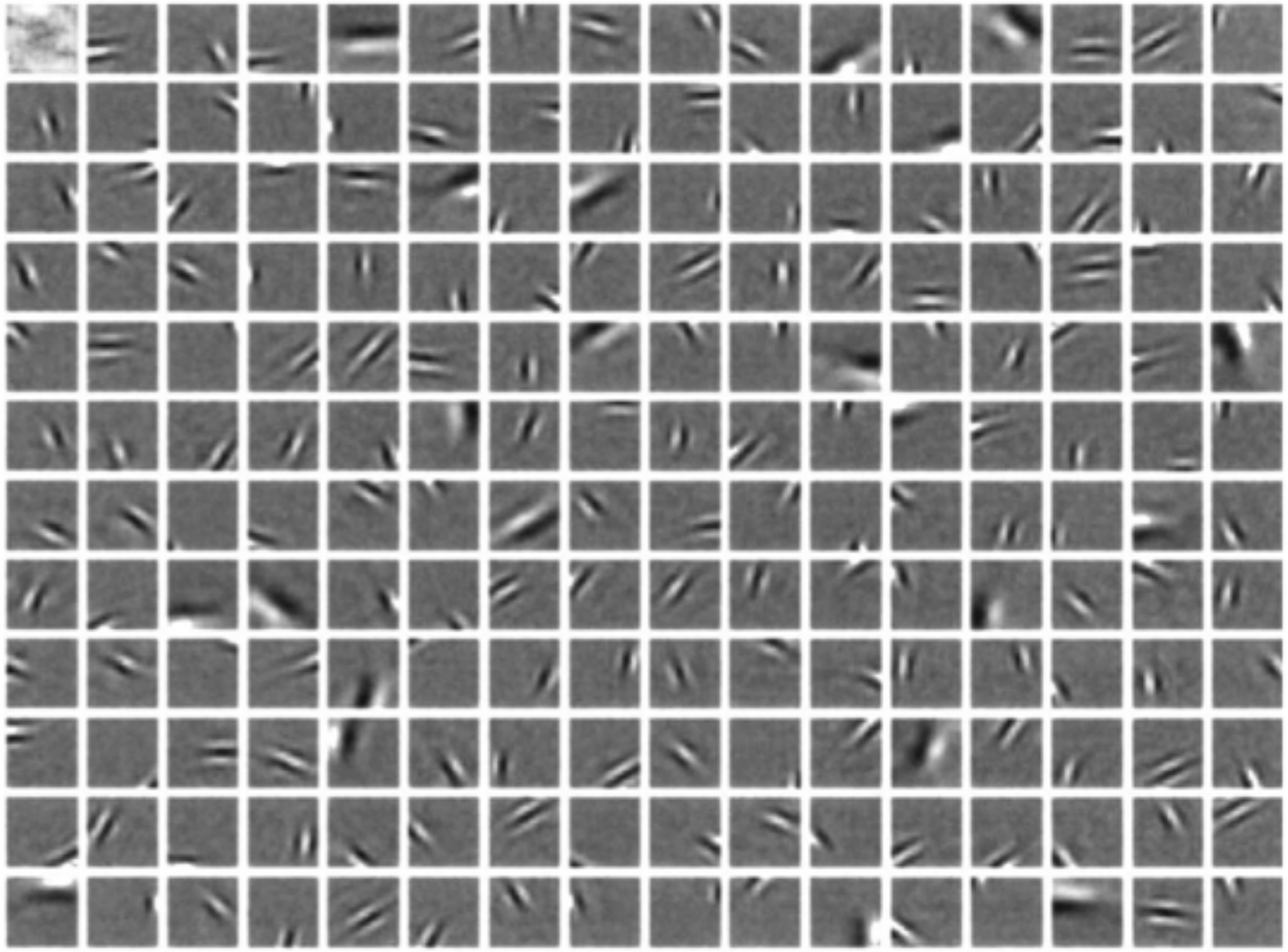
$$\|W\varphi_\tau x - \varphi_\tau W x\| \lesssim \|\tau\| \ .$$

- We will also see that the discriminability of $\Phi(x) = \rho(Wx)$ is controlled by the *sparsity* produced by W :

$$\{x * \psi_\lambda(u)\}_{\lambda,u} \text{ has few non-zero coefficients.}$$

Examples

- Olshausen and Field Sparse coding model trained on natural images:



$$\min_{W,z} \|X - Wz\|^2 + \lambda \|z\|_1$$

[Olshausen and Field,'96]

Examples

- Top performing shallow network unsupervised learning:

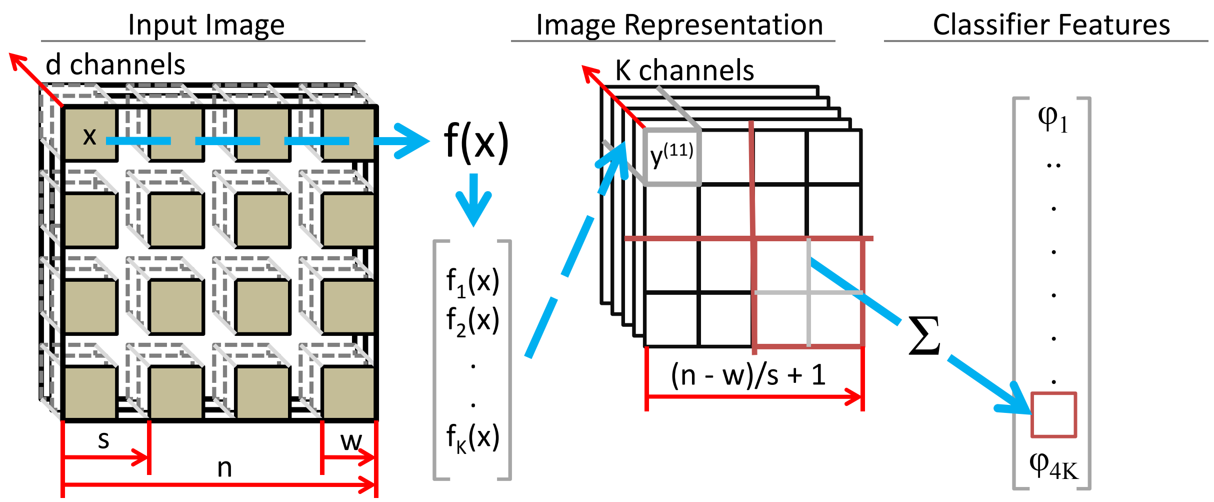
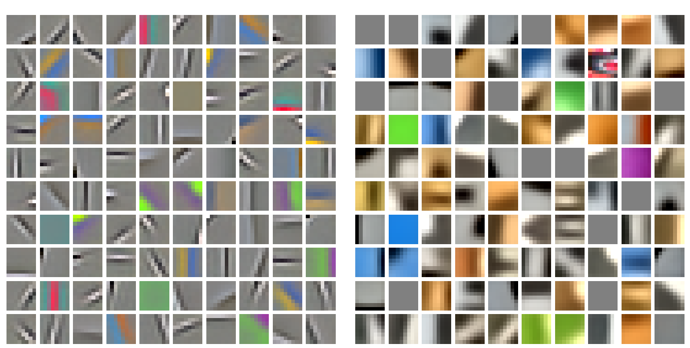
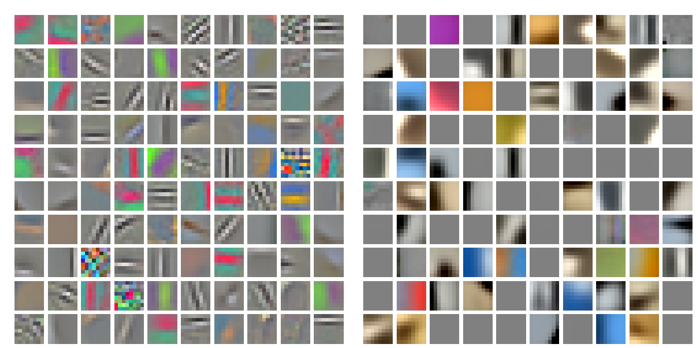


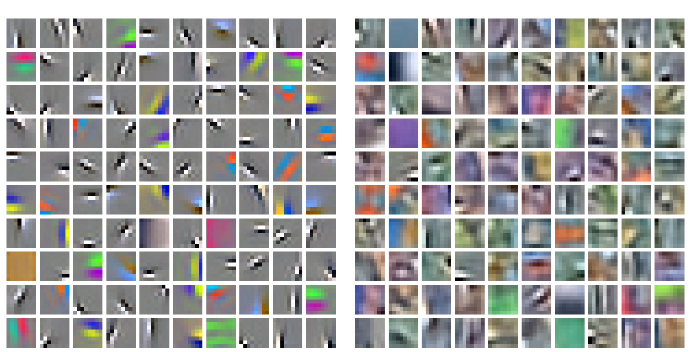
Figure 1: Illustration showing feature extraction using a w-by-w receptive field and stride s. We first extract w-by-w patches separated by s pixels each, then map them to K-dimensional feature vectors to form a new image representation. These vectors are then pooled over 4 quadrants of the image to form a feature vector for classification. (For clarity we have drawn the leftmost figure with a stride greater than w, but in practice the stride is almost always smaller than w.



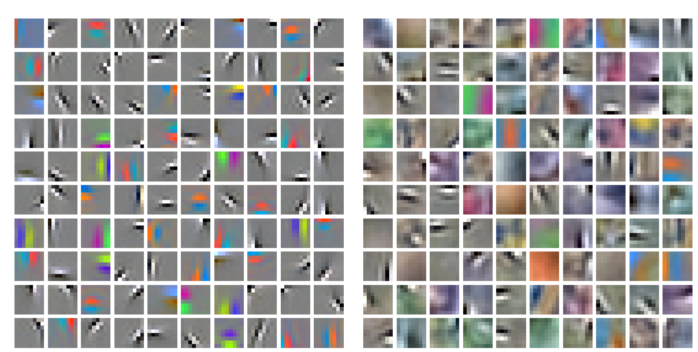
(a) K-means (with and without whitening)



(b) GMM (with and without whitening)



(c) Sparse Autoencoder (with and without whitening)



(d) Sparse RBM (with and without whitening)

[Coates, Le & Ng, '10]

Wavelets and Deformations

- We saw before that a blurring kernel is nearly invariant to deformations:

Proposition: The local averaging $\Phi(x) = x * \phi_J$ satisfies
$\forall \ \|x\| = 1 \in L^2 \ , \ \tau \ , \ \|\Phi(x) - \Phi(\varphi_\tau x)\| \leq C \|\tau\| \ .$

Wavelets and Deformations

- We saw before that a blurring kernel is nearly invariant to deformations:

  Proposition: The local averaging $\Phi(x) = x * \phi_J$ satisfies
  $\forall\ \|x\| = 1 \in L^2\ ,\ \tau\ ,\ \|\Phi(x) - \Phi(\varphi_\tau x)\| \leq C\|\tau\|\ .$

- What about the wavelet operator $\Phi(x) = \{x * \psi_\lambda\}_\lambda$?

Wavelets and Deformations

- We saw before that a blurring kernel is nearly invariant to deformations:

  Proposition: The local averaging $\Phi(x) = x * \phi_J$ satisfies $\forall \ \|x\| = 1 \in L^2 \ , \ \tau \ , \ \|\Phi(x) - \Phi(\varphi_\tau x)\| \leq C \|\tau\| \ .$

- What about the wavelet operator $\Phi(x) = \{x * \psi_\lambda\}_\lambda$?
  - We don't have local invariance, but we have a form of local covariance:

Proposition [Mallat]: For each $\delta > 0$ there exists $C > 0$ such that for all J and all $\tau \in C^2$ with $\|\nabla \tau\|_\infty \leq 1 - \delta$ we have

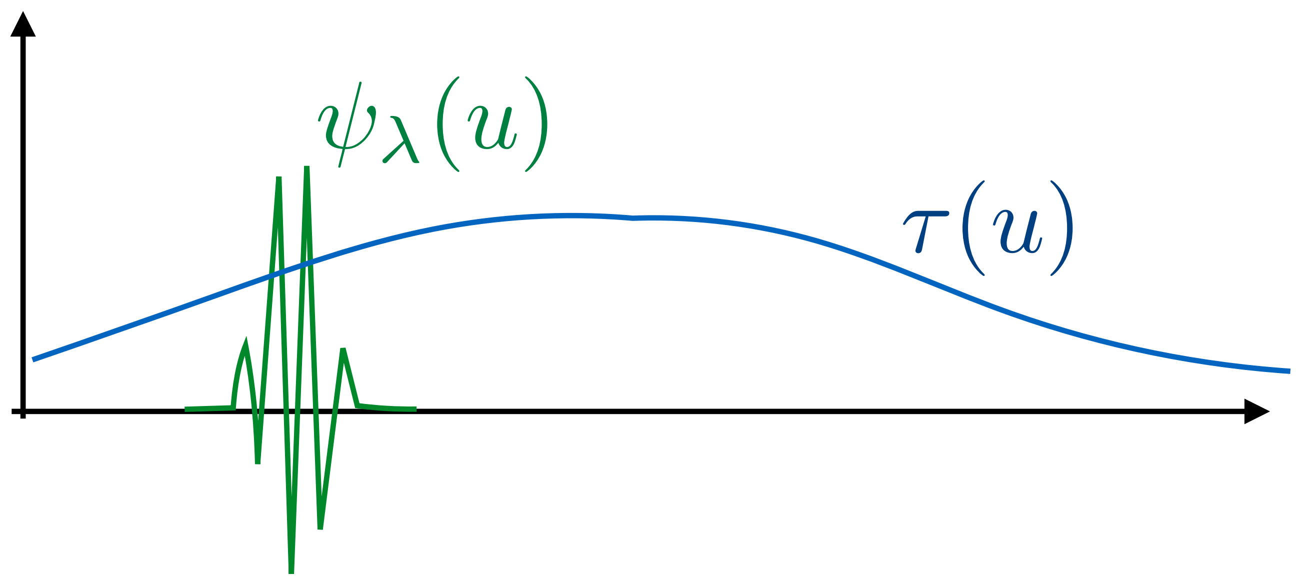
$$\|W_J \varphi_\tau - \varphi_\tau W_J\| \leq C (J \|\nabla \tau\|_\infty + \|H\tau\|_\infty) \ .$$

($H\tau$: Hessian of τ)

Wavelets and Deformations

- Qualitative idea behind this result:

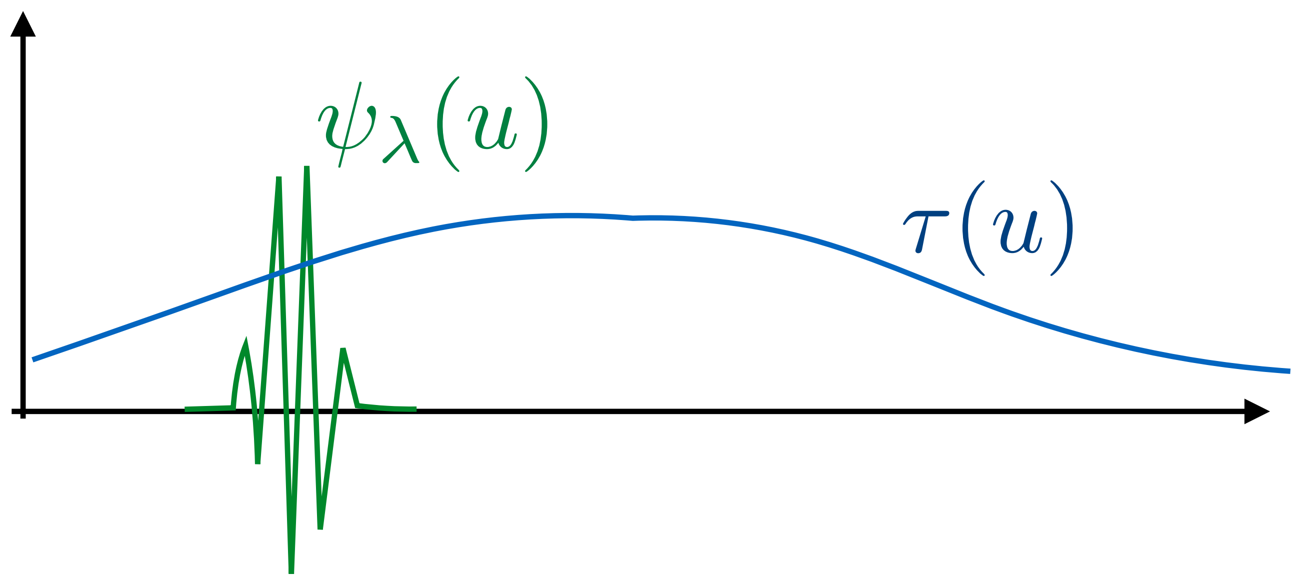
Each ψ_λ only "sees" the part of the deformation τ that intersects its support.



Wavelets and Deformations

- Qualitative idea behind this result:

Each ψ_λ only "sees" the part of the deformation τ that intersects its support.



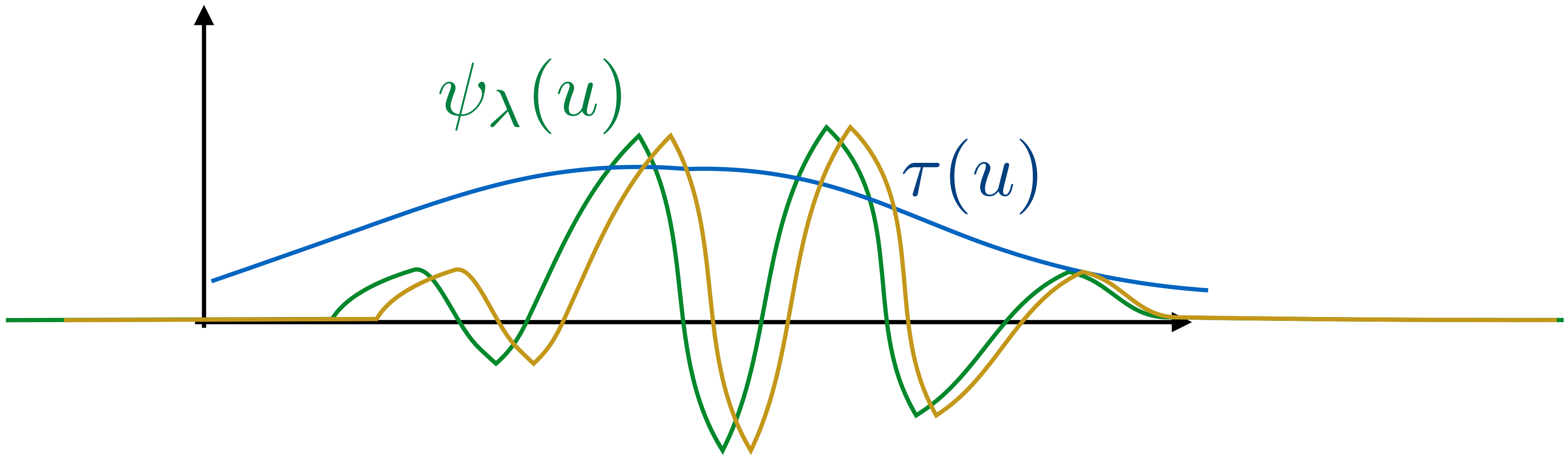
For small scales, ψ_λ has small support, and for u, v within that support, because τ is smooth, $|\tau(v) - \tau(u)| \sim 2^{-j} |\nabla \tau|_\infty$.

Thus $|(\varphi_\tau x) * \psi_\lambda(u) - x * \psi_\lambda(u - \tau(u))| \sim |\nabla \tau|_\infty$.

Wavelets and Deformations

- Qualitative idea behind this result:

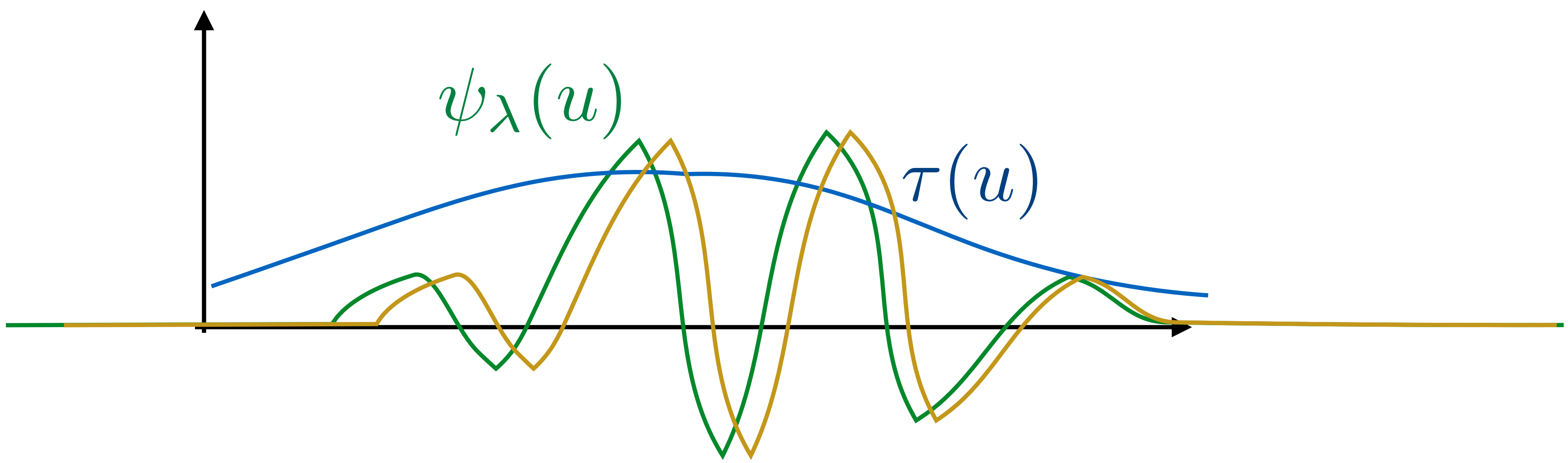
Each ψ_λ only "sees" the part of the deformation τ that intersects its support.



Wavelets and Deformations

- Qualitative idea behind this result:

Each ψ_λ only "sees" the part of the deformation τ that intersects its support.

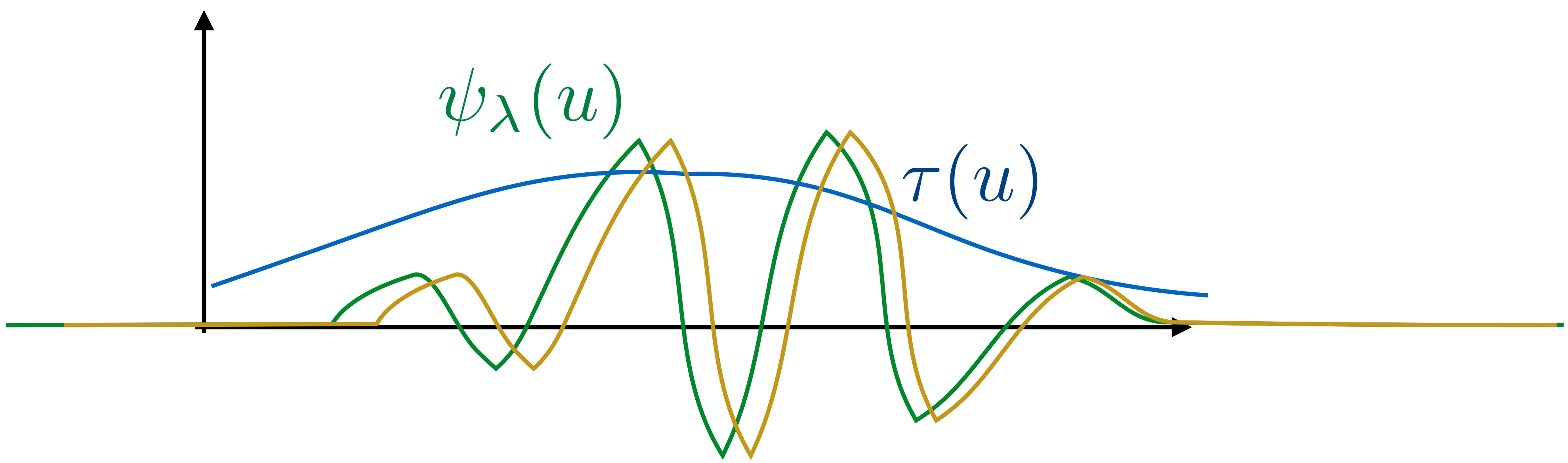


For large scales, ψ_λ is itself smooth, thus
$|\varphi_\tau(x * \psi_\lambda) - (\varphi_\tau x) * \psi_\lambda| \sim \|\nabla \tau\|_\infty$.

Wavelets and Deformations

- Qualitative idea behind this result:

Each ψ_λ only "sees" the part of the deformation τ that intersects its support.

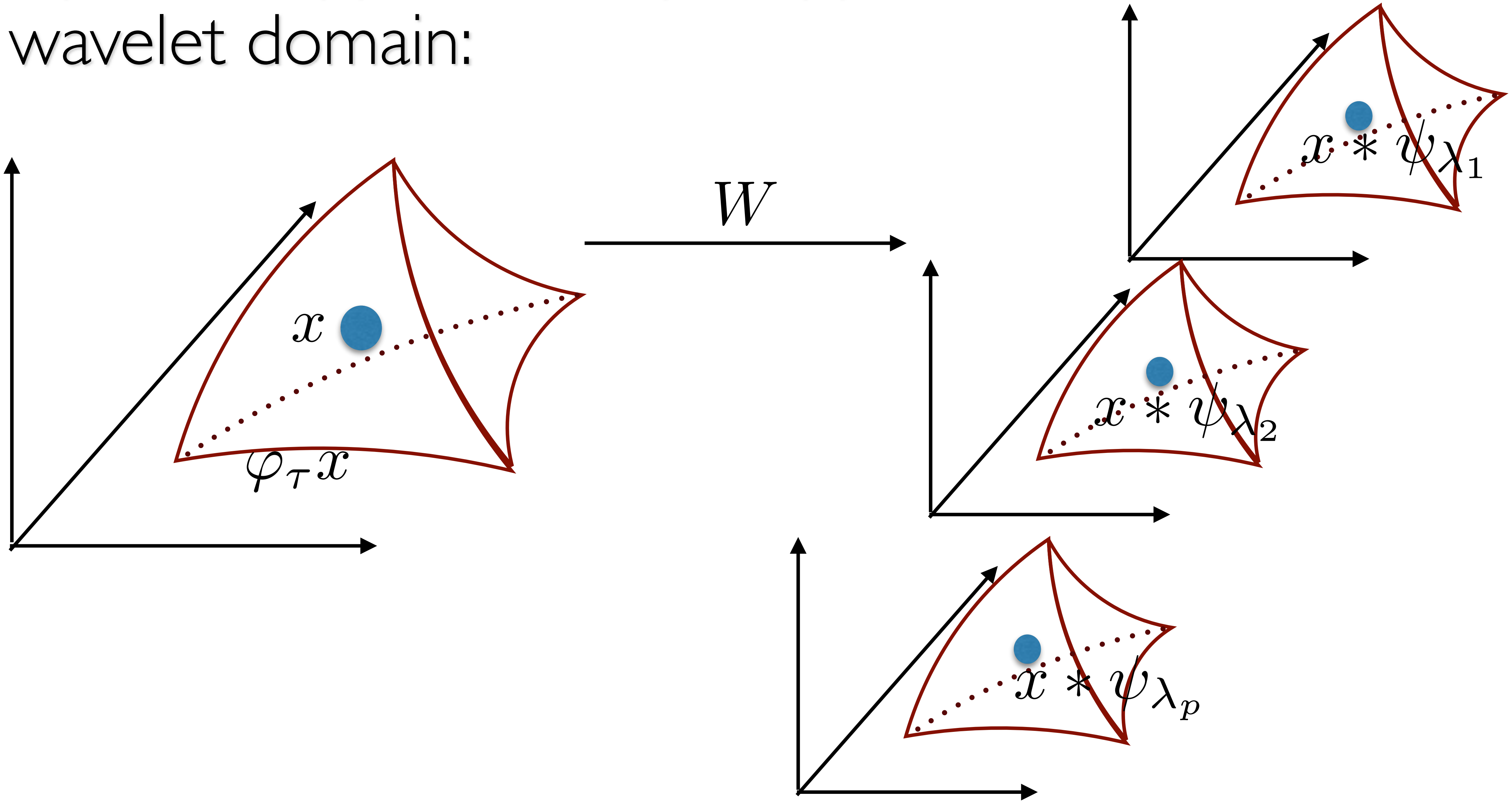


For large scales, ψ_λ is itself smooth, thus
$|\varphi_\tau(x * \psi_\lambda) - (\varphi_\tau x) * \psi_\lambda| \sim \|\nabla \tau\|_\infty$.

And, most importantly, wavelet separates scales
(so errors do not accumulate)

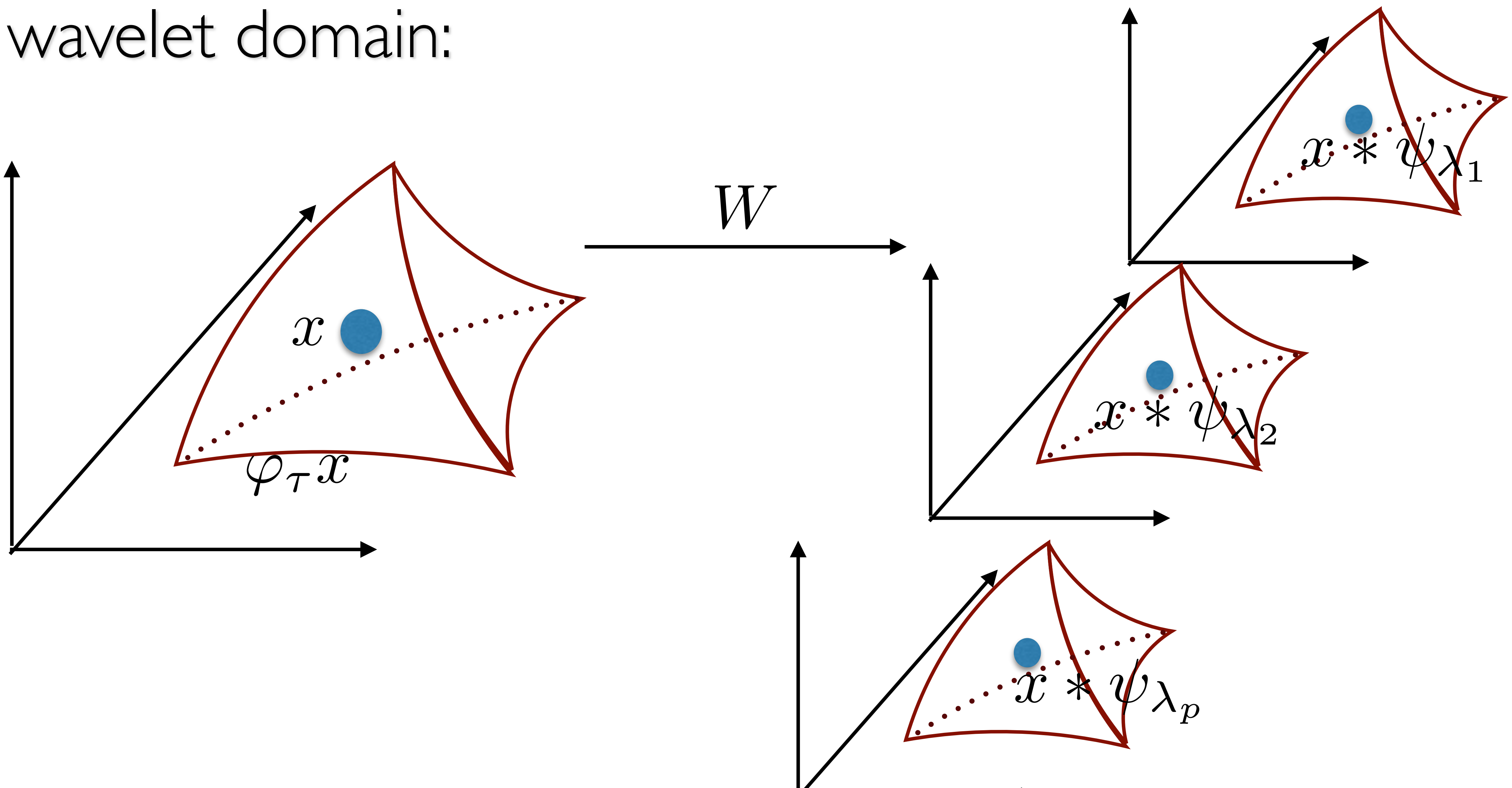
Wavelets and Non-linearities

- The commutation property says that deformations in the input are approximately mapped to deformations in the wavelet domain:



Wavelets and Non-linearities

- The commutation property says that deformations in the input are approximately mapped to deformations in the wavelet domain:



- We want to extract again stable measurements: *need non-linear operator.*

Characterization of stable non-linearities

- Preserve additive stability:

$$\|Mx - Mx'\| \leq \|x - x'\| \ . \qquad M \text{ non-expansive} \ .$$

Characterization of stable non-linearities

- Preserve additive stability:

$$\|Mx - Mx'\| \leq \|x - x'\| \;. \qquad M \text{ non-expansive} \quad .$$

- Preserve geometric stability: It is sufficient to commute with diffeomorphisms:

$$\left.\begin{array}{l} \Phi \text{ stable: } \|\Phi(\varphi_\tau x) - \Phi(x)\| \lesssim \|\tau\| \\ M \text{ commutes with } \varphi_\tau \ \forall \tau. \end{array}\right\} \Rightarrow$$

$M\Phi$ and ΦM stable:

$$\|\Phi M(\varphi_\tau x) - \Phi M(x)\| \lesssim \|\tau\|$$

$$\|M\Phi(\varphi_\tau x) - M\Phi(x)\| \lesssim \|\tau\|$$

Characterization of stable non-linearities

- Preserve additive stability:

$$\|Mx - Mx'\| \leq \|x - x'\| \ . \qquad M \text{ non-expansive } \ .$$

- Preserve geometric stability: It is sufficient to commute with diffeomorphisms.

Theorem: If M is non-expansive operator in L^2 such that $\varphi_\tau M = M \varphi_\tau$ for all τ, then M is point-wise:

$$Mx(u) = \rho(x(u)) \ .$$

Characterization of stable non-linearities

- Preserve additive stability:

$$\|Mx - Mx'\| \leq \|x - x'\| \,. \qquad M \text{ non-expansive} \quad .$$

- Preserve geometric stability: It is sufficient to commute with diffeomorphisms.

Theorem: If M is non-expansive operator in L^2 such that $\varphi_\tau M = M \varphi_\tau$ for all τ, then M is point-wise:

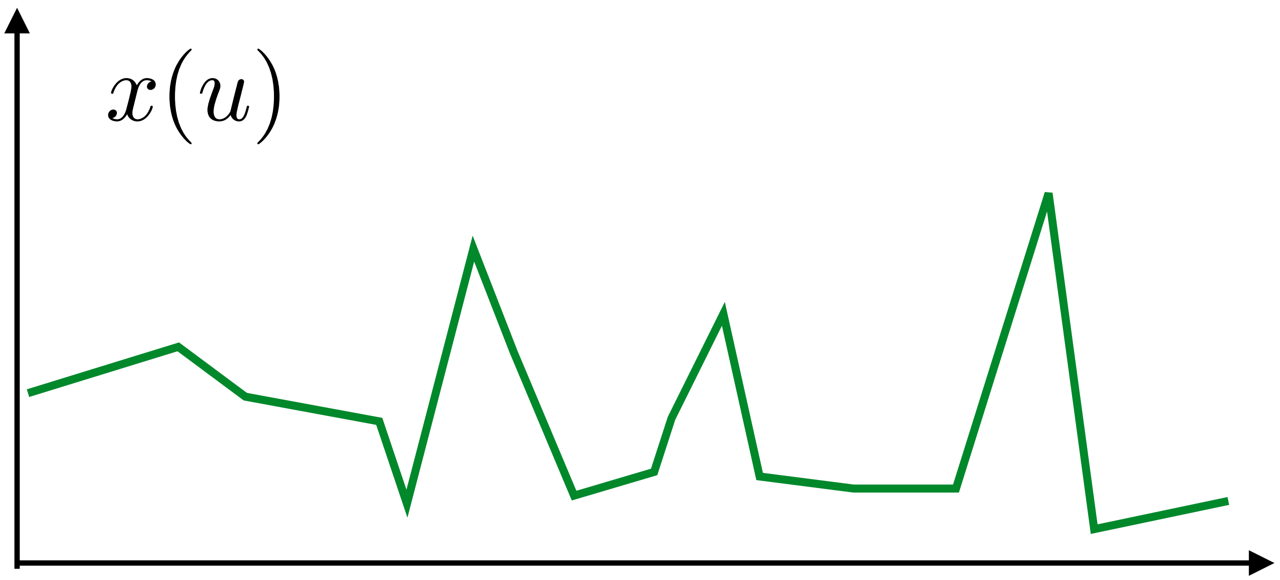
$$Mx(u) = \rho(x(u)) \,.$$

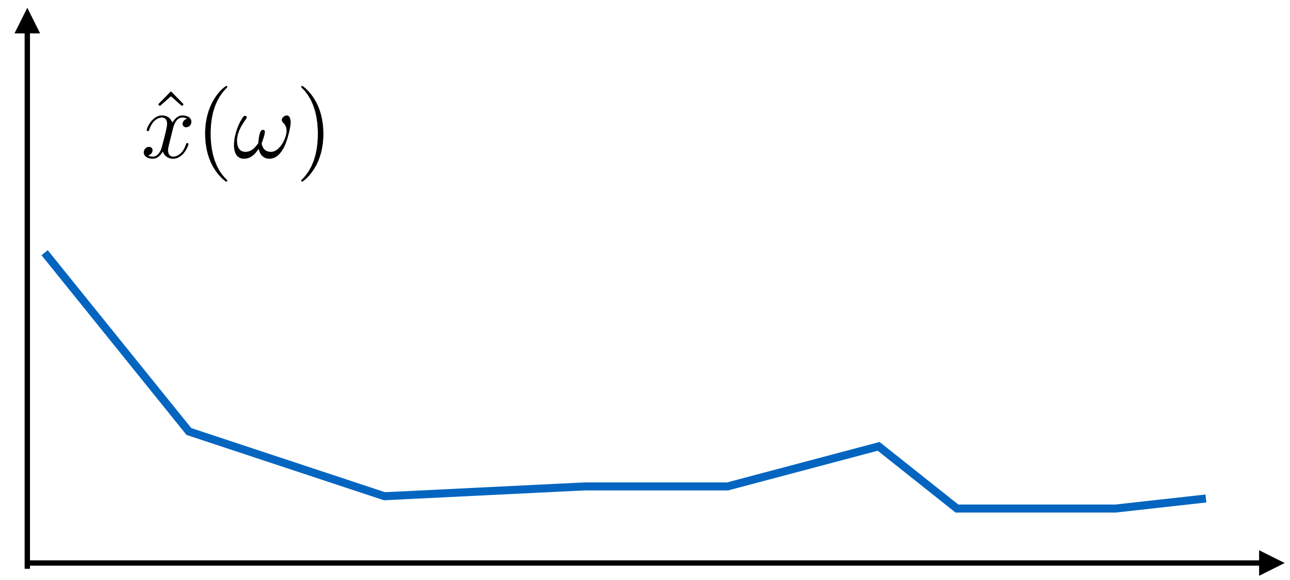
- Since we want to smooth orbits, we may choose a point-wise nonlinearity that reduces oscillations:

$$\rho(z) = |z| \text{ or } \rho(z) = \max(0, z)$$

Understanding the effect of nonlinearities

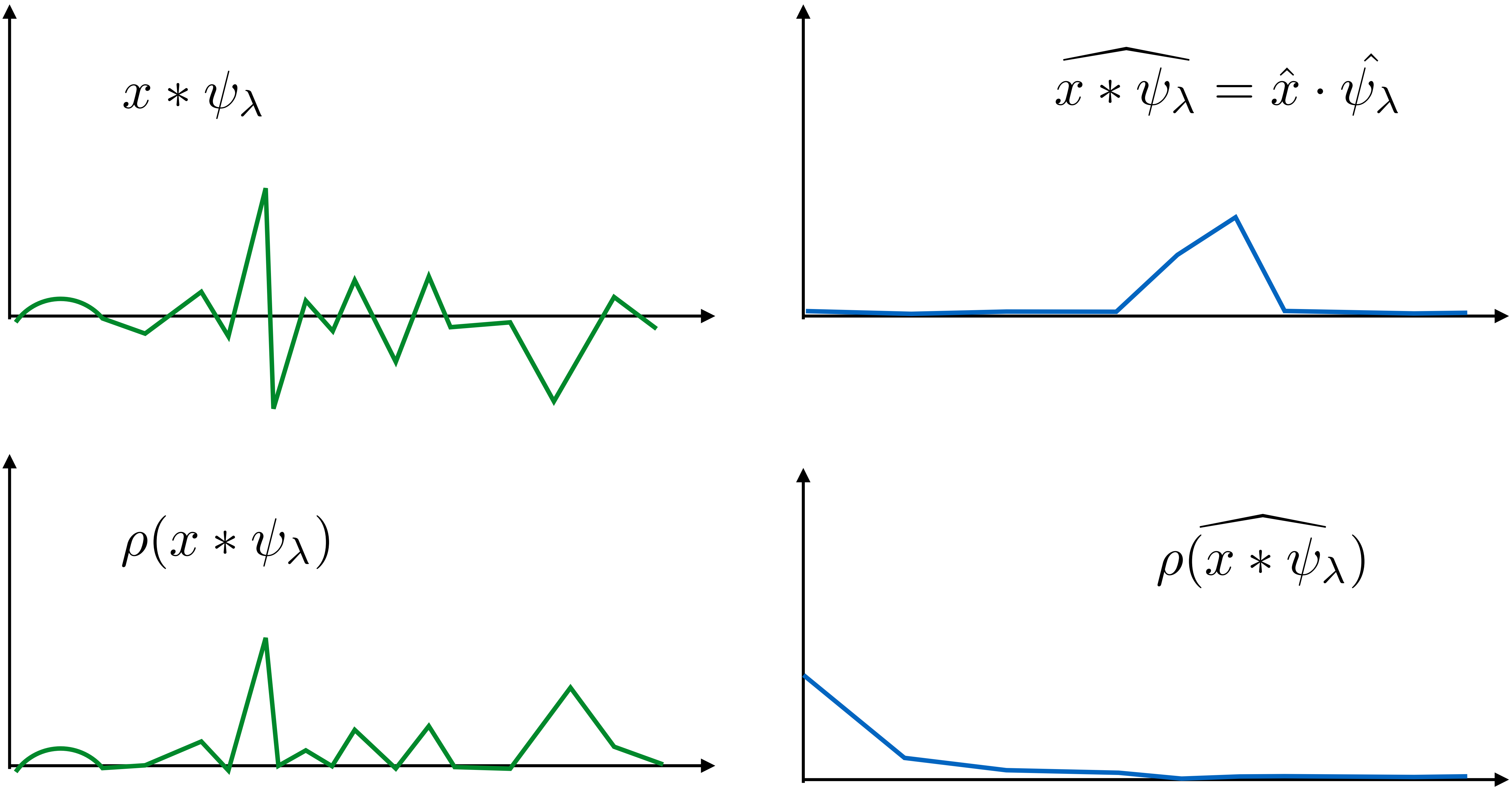
- Rectifiers thus perform a *non-linear demodulation*:





Understanding the effect of nonlinearities

- Rectifiers thus perform a *non-linear demodulation*:



sometimes called the *envelope*

Choice of Pointwise Nonlinearity

- Full rectification $\rho(z) = |z|$ preserves energy:
  - When the wavelet is complex, it produces smoother envelopes (thus more stable features).

- Half rectification (ReLU) $\rho(z) = \max(z, 0)$ captures half the energy, and it also creates *sparsity.*
  - We will see that this is important to perform *detection.*

- Sigmoid nonlinearity $\rho(z) = (1 + e^{-z})^{-1}.$
  - It is not homogeneous
  - Saturating regimes are problematic for learning via back propagation in deep models.

- "Leaky" ReLU [MSR'14]: parametrized half-rectifier.

Separable Scattering Operators

- Local averaging kernel: $x \star \phi_J$
  - locally translation invariant
  - stable to additive and geometric deformations
  - loss of high-frequency information.

Separable Scattering Operators

- Local averaging kernel: $x \star \phi_J$
  - locally translation invariant
  - stable to additive and geometric deformations
  - loss of high-frequency information.
- Recover lost information: $\mathcal{U}_J(x) = \{x \star \phi_J \,,\, |x \star \psi_\lambda|\}_{\lambda \in \Lambda_J}$.
  - Point-wise, non-expansive non-linearities: maintain stability.
  - Complex modulus maps energy towards low-frequencies.

Separable Scattering Operators

- Local averaging kernel: $x \star \phi_J$
  - locally translation invariant
  - stable to additive and geometric deformations
  - loss of high-frequency information.
- Recover lost information: $\mathcal{U}_J(x) = \{x \star \phi_J \,,\, |x \star \psi_\lambda|\}_{\lambda \in \Lambda_J} \,.$
  - Point-wise, non-expansive non-linearities: maintain stability.
  - Complex modulus maps energy towards low-frequencies.
- Cascade the "recovery" operator:

$$\mathcal{U}_J^2(x) = \{x \star \phi_J \,,\, |x \star \psi_\lambda| \star \phi_J,\, ||x \star \psi_\lambda| \star \psi_{\lambda'}|\}_{\lambda, \lambda' \in \Lambda_J} \,.$$

Separable Scattering Operators

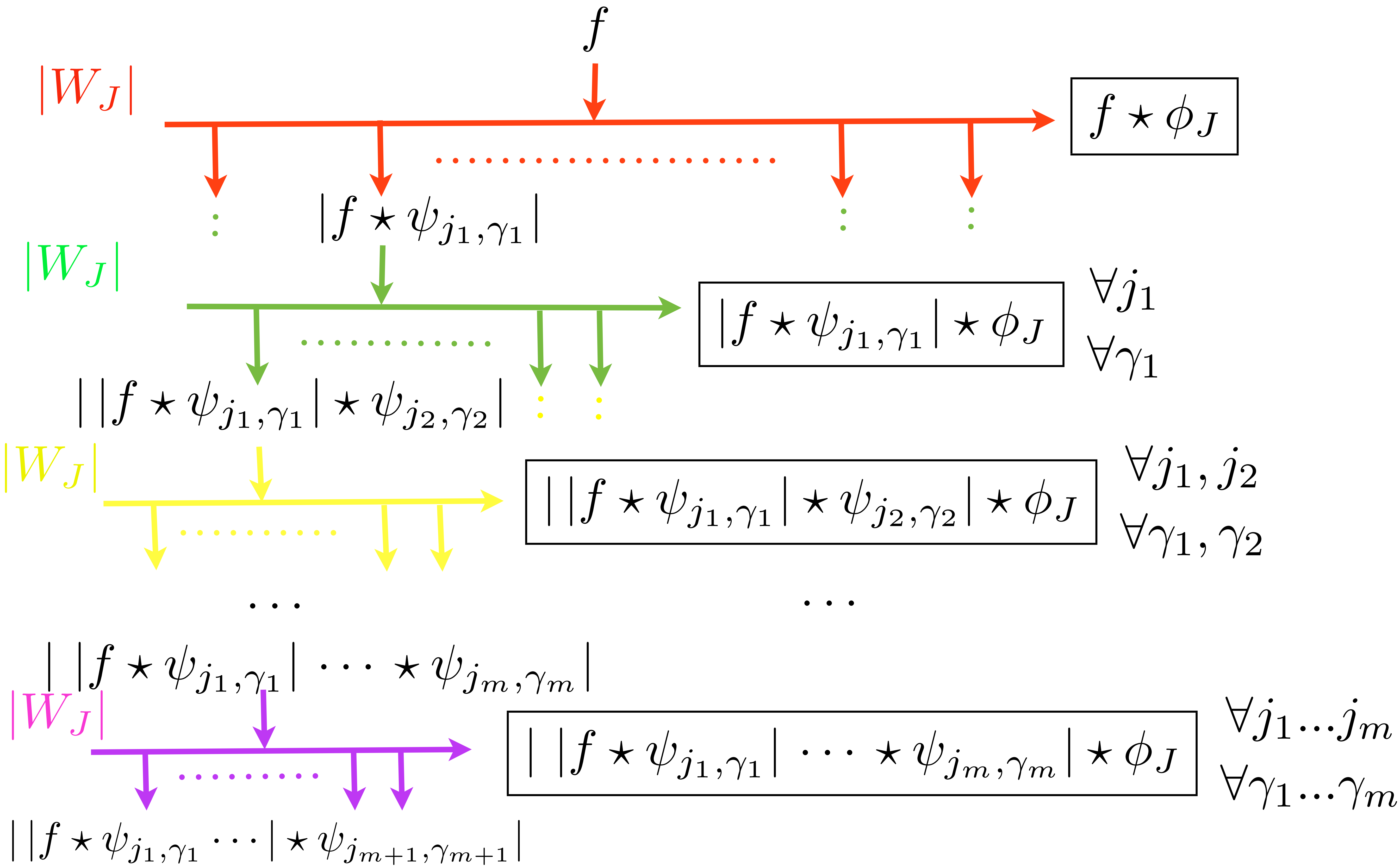
- Local averaging kernel: $x \star \phi_J$
  - locally translation invariant
  - stable to additive and geometric deformations
  - loss of high-frequency information.
- Recover lost information: $\mathcal{U}_J(x) = \{x \star \phi_J \,,\, |x \star \psi_\lambda|\}_{\lambda \in \Lambda_J}$.
  - Point-wise, non-expansive non-linearities: maintain stability.
  - Complex modulus maps energy towards low-frequencies.
- Cascade the "recovery" operator:

$$\mathcal{U}_J^2(x) = \{x \star \phi_J \,,\, |x \star \psi_\lambda| \star \phi_J, \, ||x \star \psi_\lambda| \star \psi_{\lambda'}|\}_{\lambda, \lambda' \in \Lambda_J} \ .$$

- Scattering coefficient along a path $p = (\lambda_1, \ldots, \lambda_m)$:

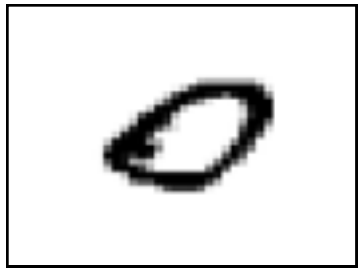
$$S_J[p]x(u) = |||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \ldots | \star \psi_{\lambda_m}| \star \phi_J(u) \ .$$

Scattering Convolutional Network



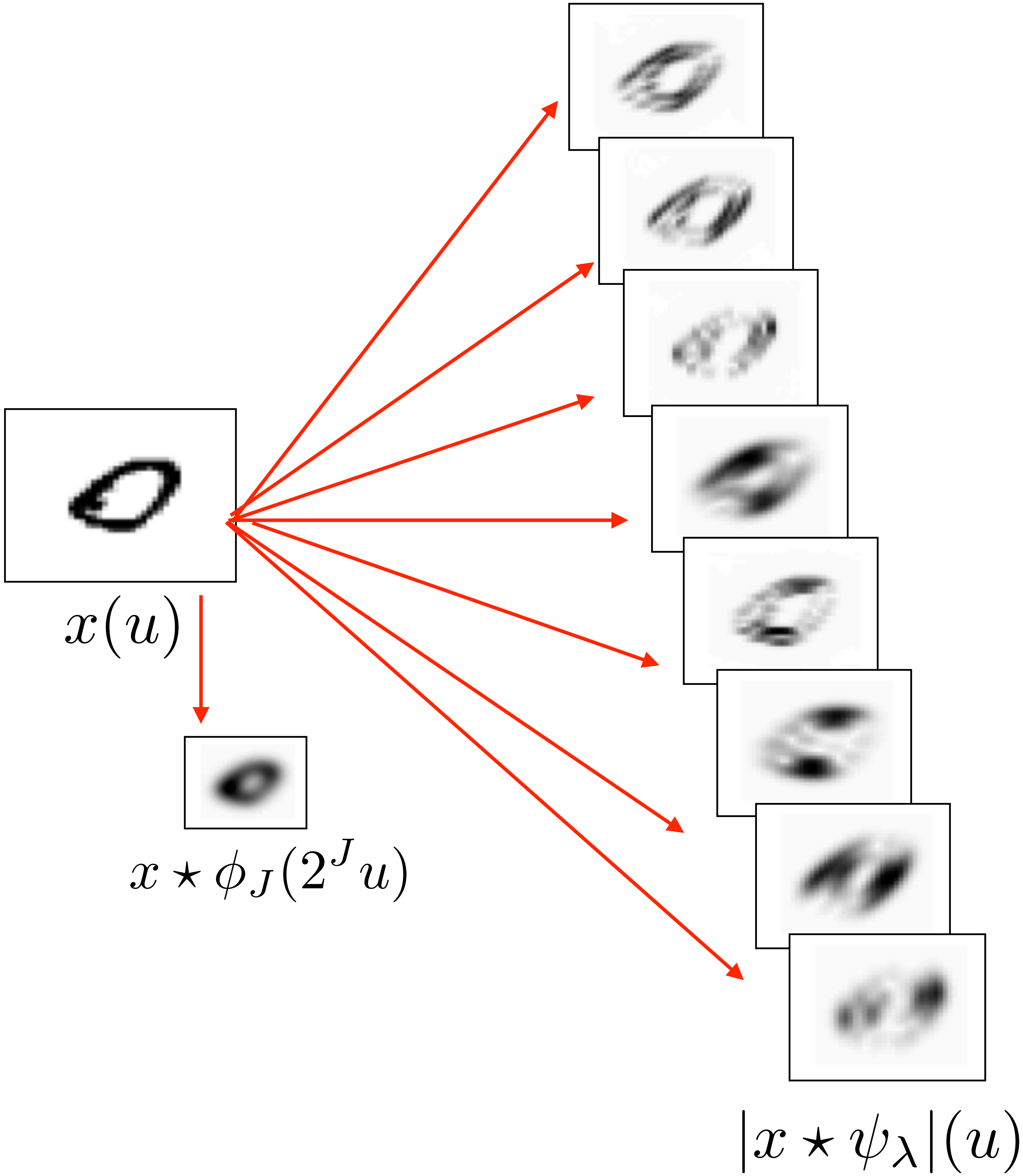
Cascade of contractive operators.

Scattering Example

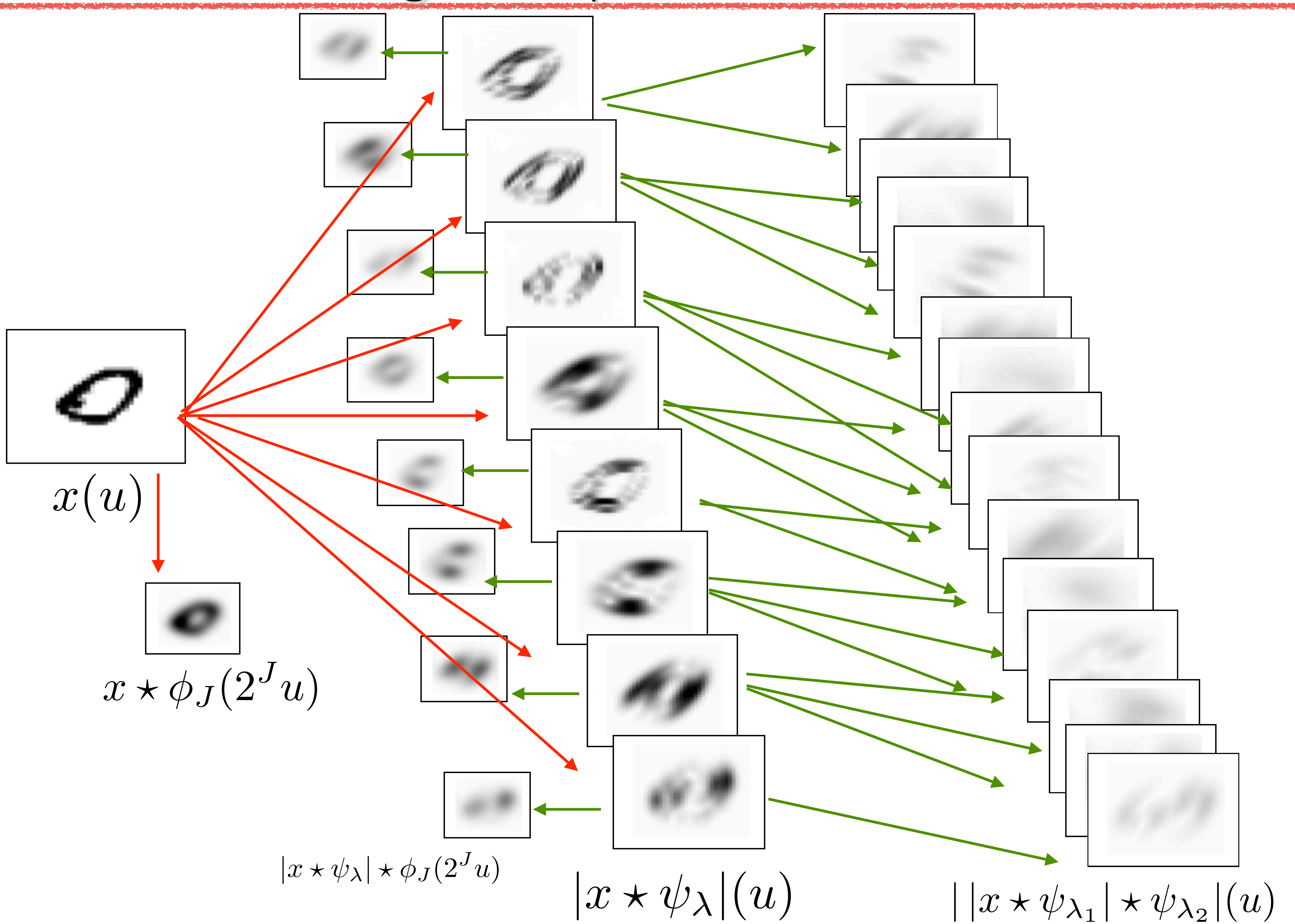


$x(u)$

Scattering Example



Scattering Example



Scattering Properties

- Additive stability and conservation of energy:

Theorem *(Mallat)*: For appropriate wavelets, the scattering representation is contractive, $\|S_J x - S_J x'\| \leq \|x - x'\|$,
and unitary, $\|S_J x\| = \|x\|$.

$$\|S_J x\|^2 = \sum_{p \in \mathcal{P}_J} \|S_J[p]x\|^2$$

Scattering Properties

- Additive stability and conservation of energy:

Theorem *(Mallat)*: For appropriate wavelets, the scattering representation is contractive, $\|S_J x - S_J x'\| \leq \|x - x'\|$, and unitary, $\|S_J x\| = \|x\|$.

$$\|S_J x\|^2 = \sum_{p \in \mathcal{P}_J} \|S_J[p]x\|^2$$

- In practice, the transform is limited to a finite number of layers m_{max}

Scattering Properties

- Additive stability and conservation of energy:

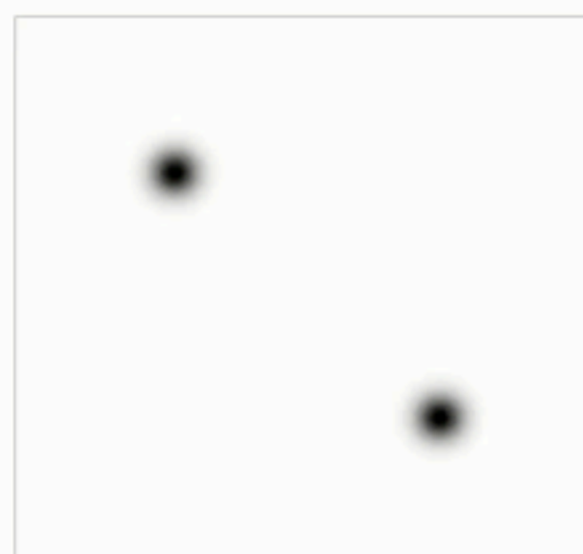
Theorem *(Mallat)*: For appropriate wavelets, the scattering representation is contractive, $\|S_J x - S_J x'\| \leq \|x - x'\|$, and unitary, $\|S_J x\| = \|x\|$.

$$\|S_J x\|^2 = \sum_{p \in \mathcal{P}_J} \|S_J[p]x\|^2$$

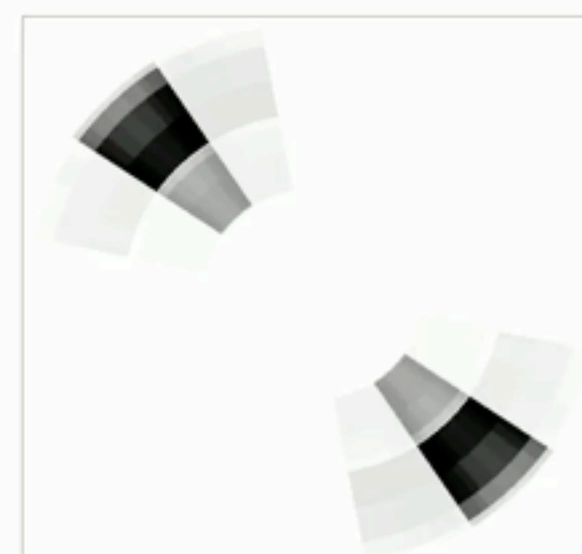
- Geometric Stability:

Theorem *(Mallat)*: There exists C such that $\forall x \in L^2$ and all m, the m-th order scattering satisfies $\|S_J L[\tau]x - S_J x\| \leq C m \|x\| \left(2^{-J}\|\tau\|_\infty + \|\nabla\tau\|_\infty + \|H\tau\|_\infty\right)$.

$L[\tau]x$



$|\widehat{L[\tau]x}|$



$S_J L[\tau]x$

Discriminability

- For appropriate wavelets, the information is preserved at each layer:

Theorem: (Waldspurger) For appropriate wavelets, the operator $Ux = \{x \star \phi_J, |x \star \psi_j|\}_{j \leq J}$ is injective.

- However, the inverse is unstable —> we might be contracting too much in general. How to prevent that?

- Sparsity In terms of contraction it is very intuitive.

Discriminability

- For appropriate wavelets, the information is preserved at each layer:

Theorem: (Waldspurger) For appropriate wavelets, the operator $Ux = \{x \star \phi_J, |x \star \psi_j|\}_{j \leq J}$ is injective.

Discriminability

- For appropriate wavelets, the information is preserved at each layer:

Theorem: (Waldspurger) For appropriate wavelets, the operator $Ux = \{x \star \phi_J, |x \star \psi_j|\}_{j \leq J}$ is injective.

- However, the inverse is unstable: we might be contracting too much in general. How to prevent that?

Discriminability and Sparsity

- Typical non-linearities are contractive:

$$\|\rho(x) - \rho(x')\| \leq \|x - x'\|$$

Discriminability and Sparsity

- Typical non-linearities are contractive:

$$\|\rho(x) - \rho(x')\| \leq \|x - x'\|$$

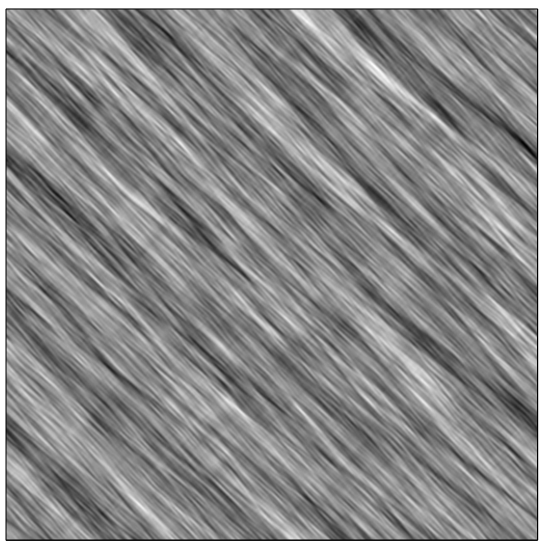
- However, if x, x' are sparse, this inequality is an equality in most of the signal domain.

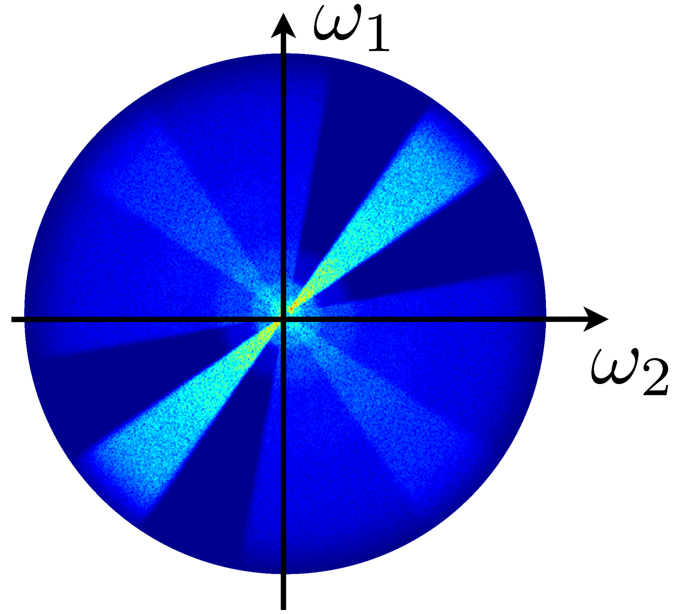
- Thus sparsity is a means to control and prevent excessive contraction of different signal classes.

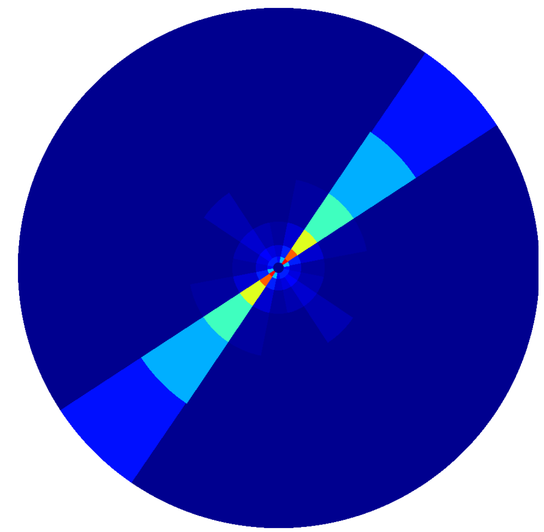
Image Examples

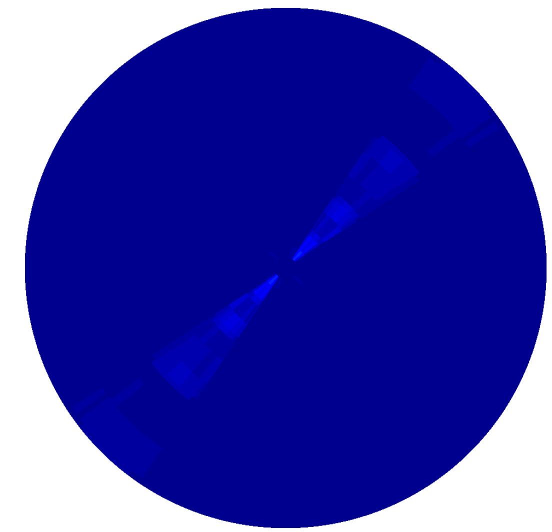
[Bruna, Mallat, '11,'12]

Images	Fourier	Wavelet Scattering	
f	$\hat{f}$	$\lvert f \star \psi_{\lambda_1} \rvert \star \phi$	$\lvert\lvert f \star \psi_{\lambda_1} \rvert \star \psi_{\lambda_2} \rvert \star \phi$

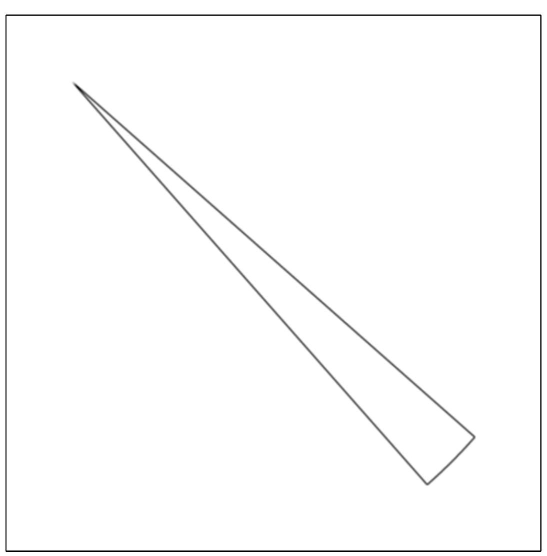


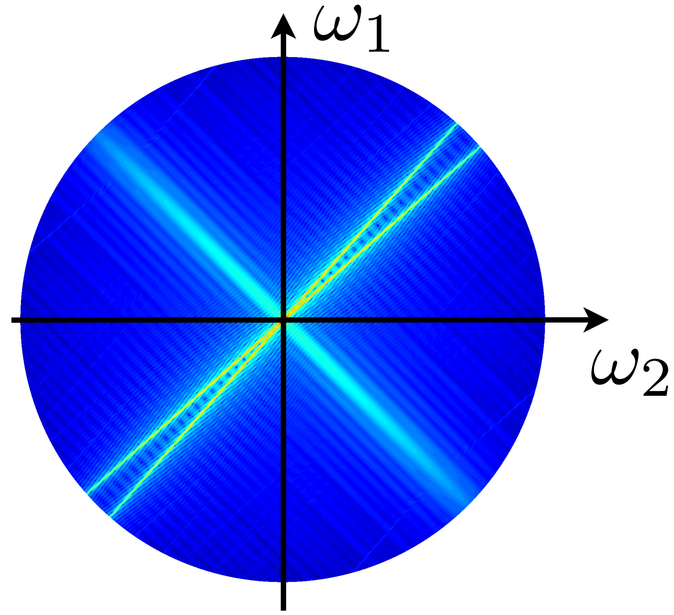


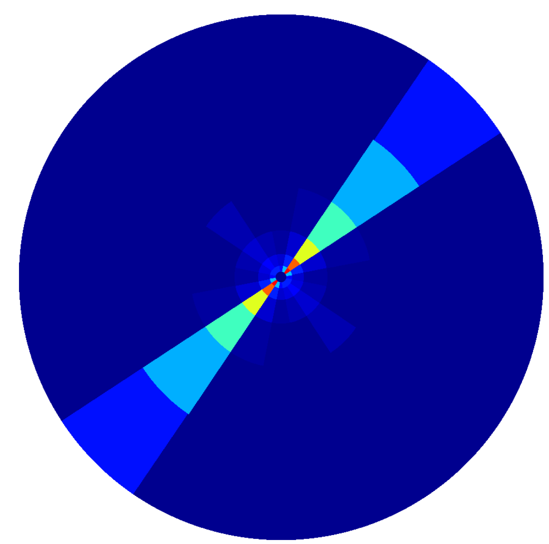


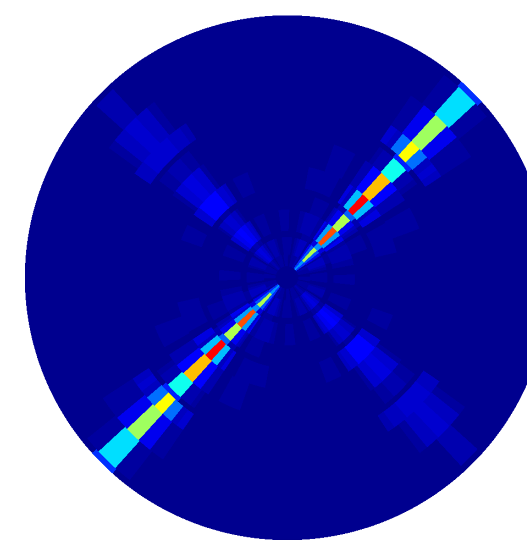


SIFT





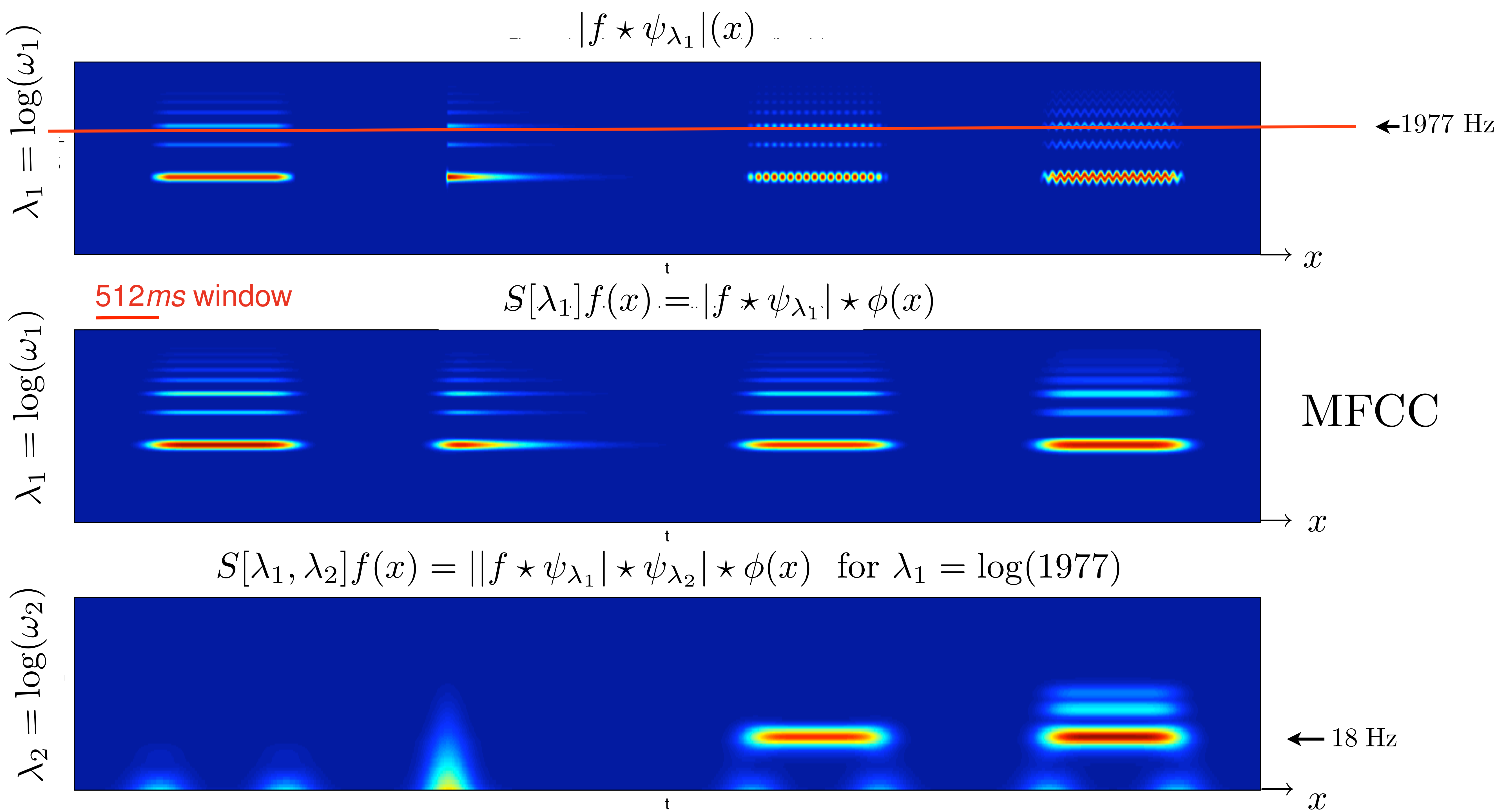




window size = image size

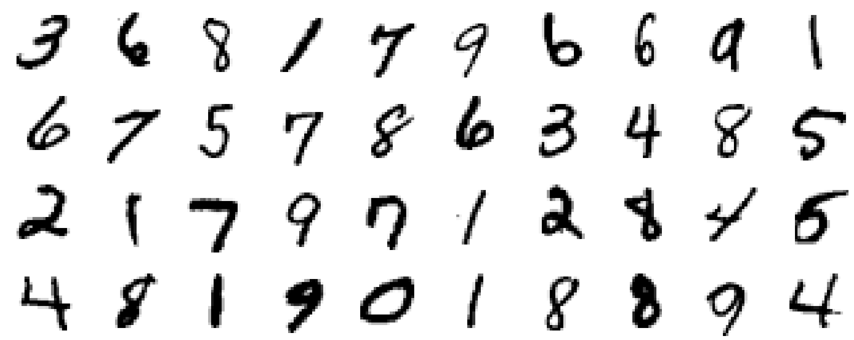
Sound Examples

(courtesy J. Anden)

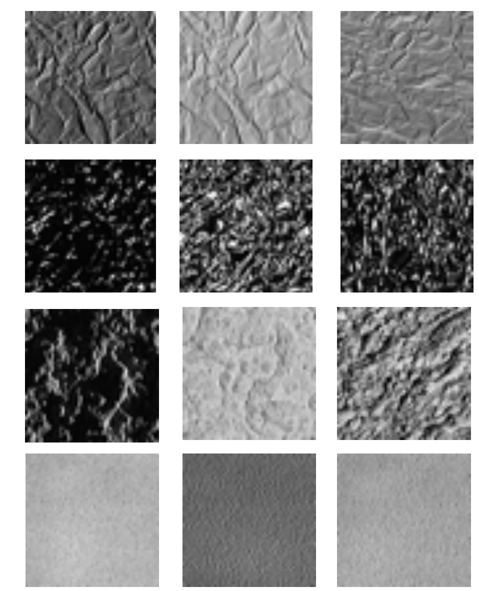


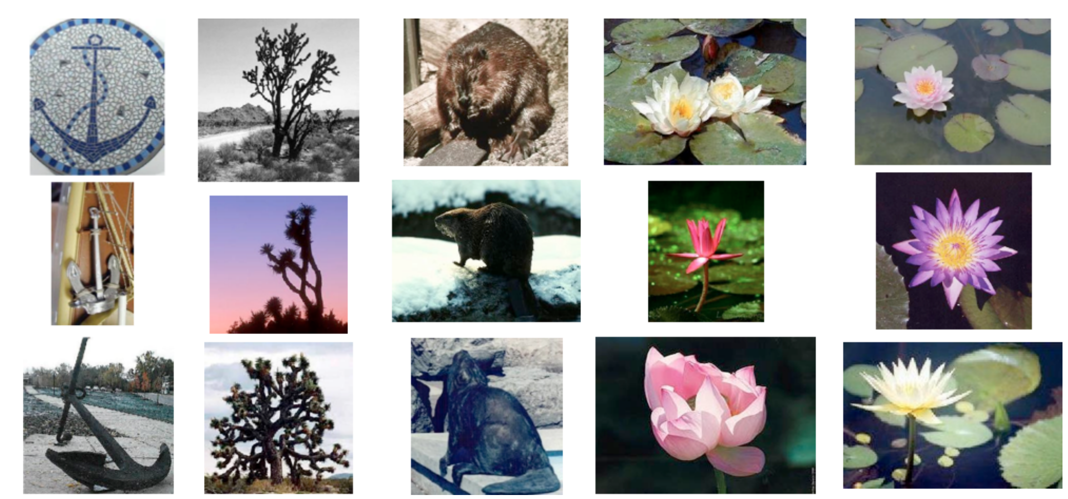
Classification with Scattering

- State-of-the art on pattern and texture recognition:
  - MNIST, USPS [Pami'13]



  - Texture (CUREt, UIUC) [Pami'13]





- Object Recognition:
  - ~17% error on Cifar-10 [Oyallon&Mallat, CVPR'15]

Limitations of Separable Scattering

- No feature dimensionality reduction
  - The number of features increases exponentially with depth
- Feature maps are not recombined
  - The deformation model is inherited from the input domain: we will see that recombining feature maps offers more powerful invariance.
- Feature maps are not learnt
  - We shall see that adapting the filters to object classes improves contraction AND discriminability.