# Bandlets and Model Selection for Image Estimation

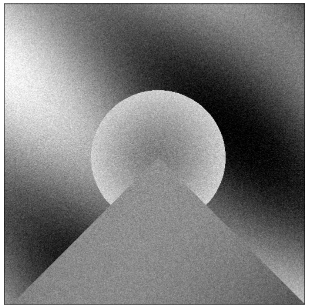
E. Le Pennec / LPMA / Université Paris Diderot (Paris 7)
S. Mallat, Ch. Dossal et G. Peyré,
F. Autin, J.-M. Loubes et V. Rivoirard

5 September 2007

# Outline

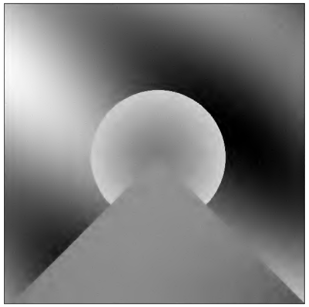
# Outline

Noisy

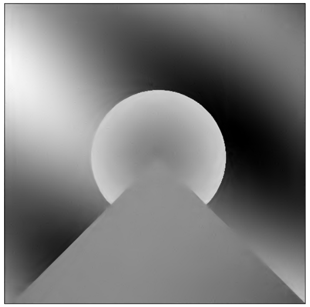


Estimation

Wavelets

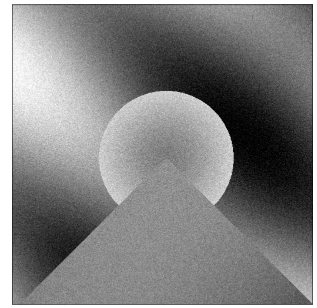


Bandlets

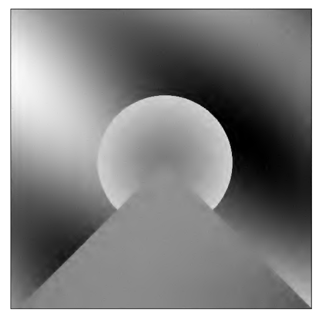


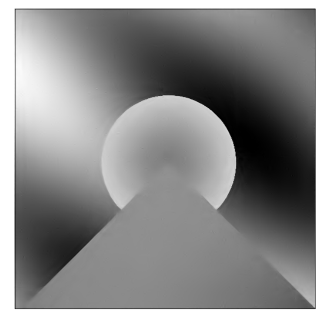
# Outline

Noisy



Estimation Wavelets Bandlets

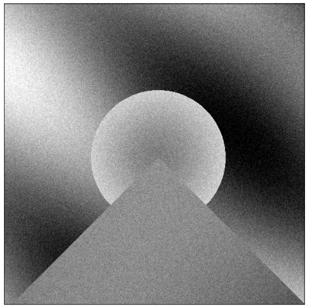




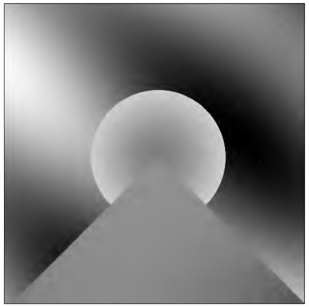
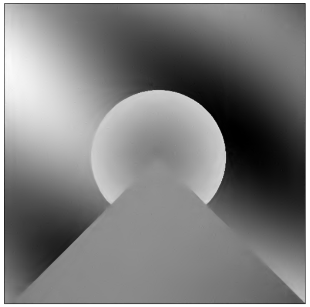
- Geometric image approximation.

# Outline

Noisy



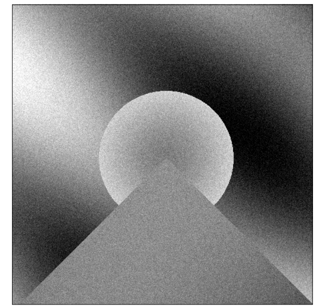
Estimation  Wavelets  Bandlets

- Geometric image approximation.
- Bandlet basis and bandlet estimation.

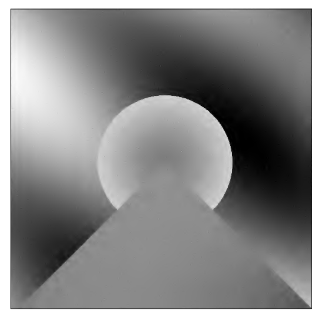
# Outline

Noisy

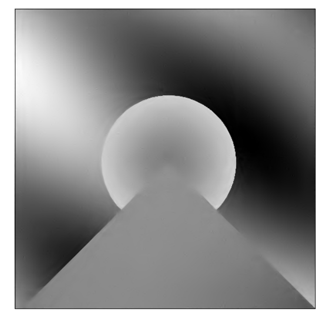


Estimation

Wavelets



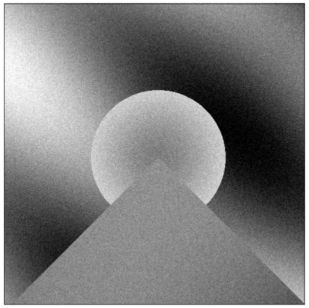
Bandlets



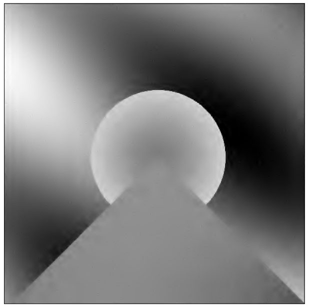
- Geometric image approximation.
- Bandlet basis and bandlet estimation.
- Thresholding, model selection, minimax and maxiset.

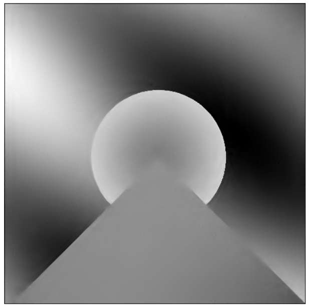
# Outline

Noisy



Estimation Wavelets Bandlets



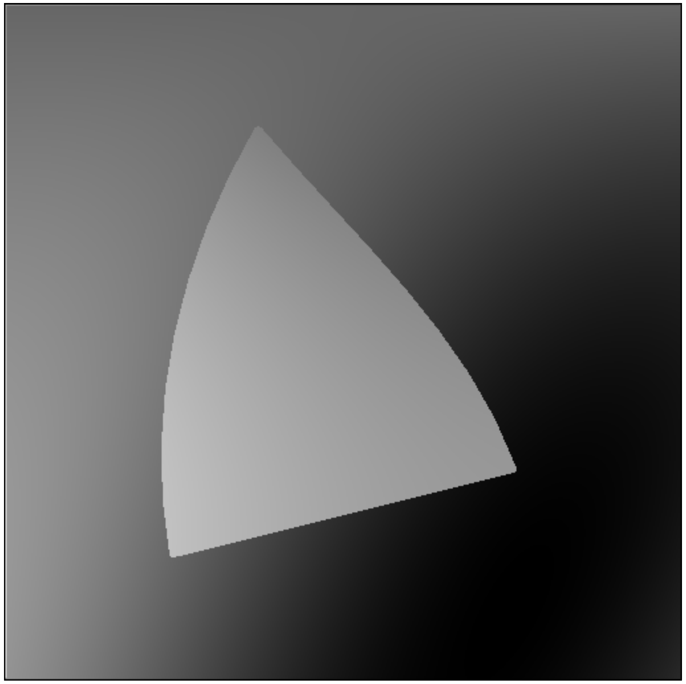


- Geometric image approximation.
- Bandlet basis and bandlet estimation.
- Thresholding, model selection, minimax and maxiset.
- Sketch of proof.

# Geometric Image and Wavelet Approximation

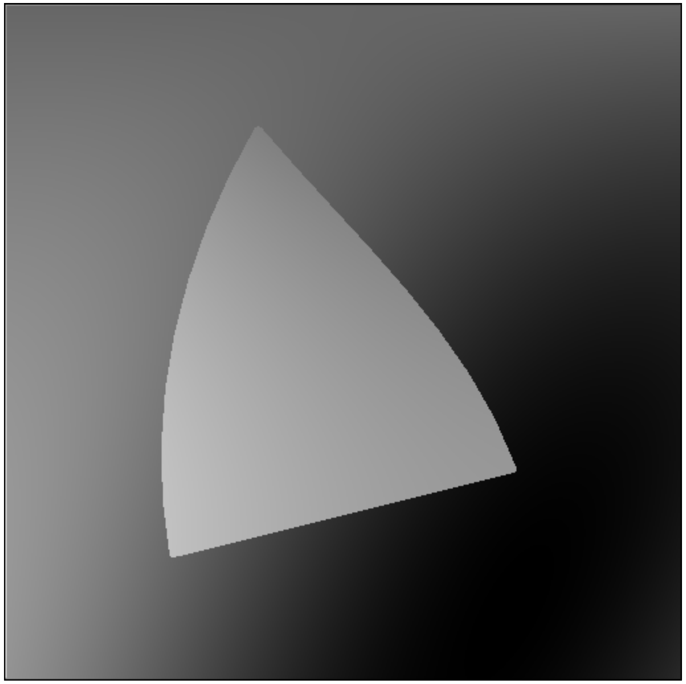
# Geometric Image and Wavelet Approximation

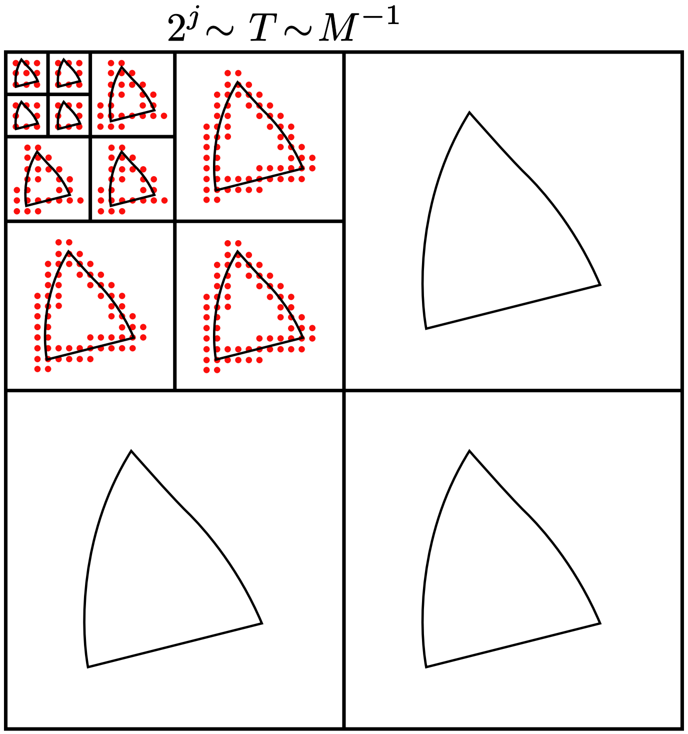
- Approximations of $f$ which is $\mathbf{C}^\alpha$ away from $\mathbf{C}^\alpha$ "edge" curves:



# Geometric Image and Wavelet Approximation

- Approximations of $f$ which is $\mathbf{C}^\alpha$ away from $\mathbf{C}^\alpha$ "edge" curves:

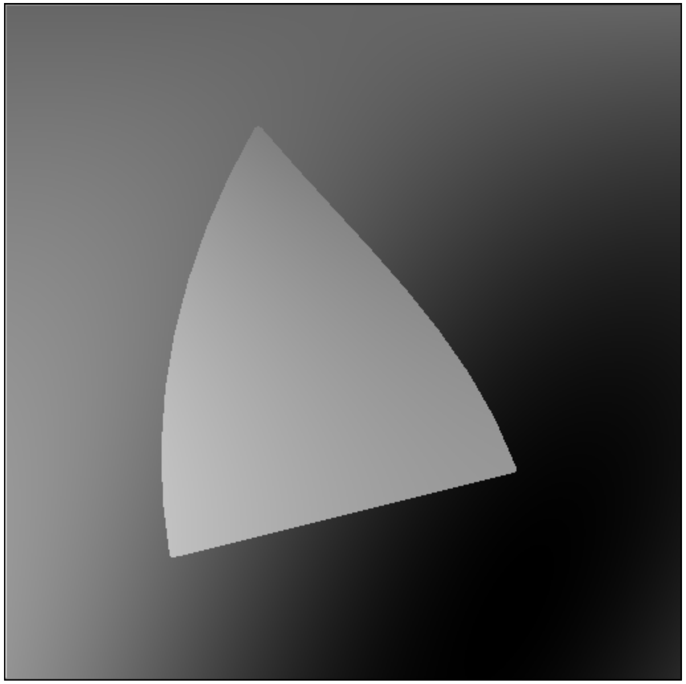


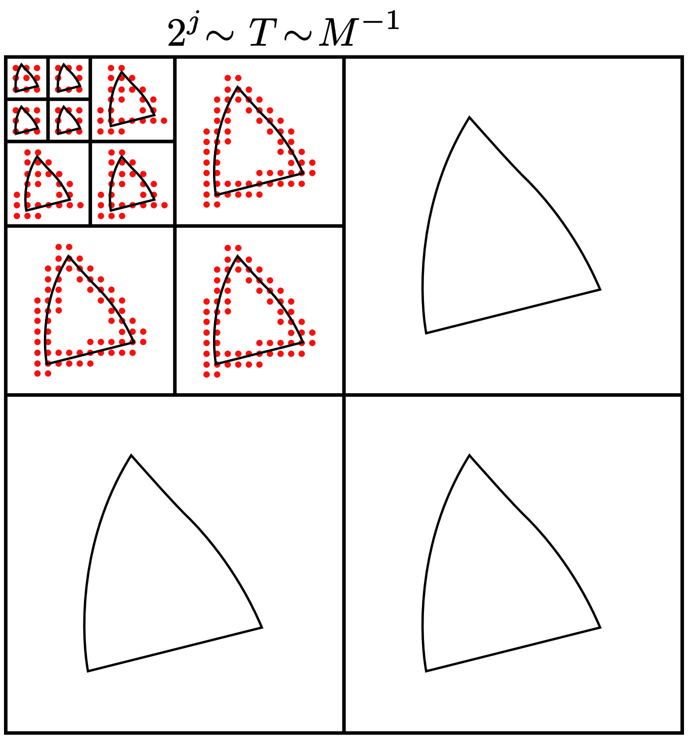



- With $M$ wavelets: $\|f - f_M\|^2 \leq C\, M^{-1}$.

# Geometric Image and Wavelet Approximation

- Approximations of $f$ which is $\mathbf{C}^\alpha$ away from $\mathbf{C}^\alpha$ "edge" curves:

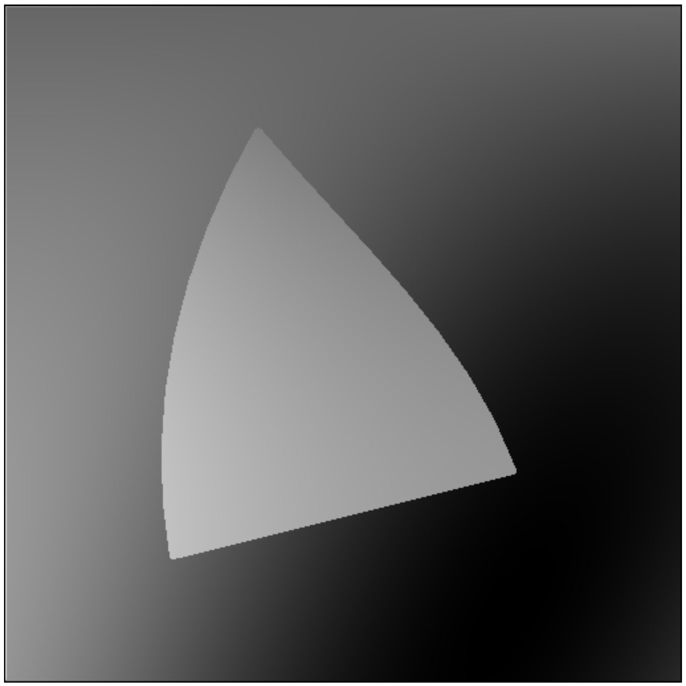


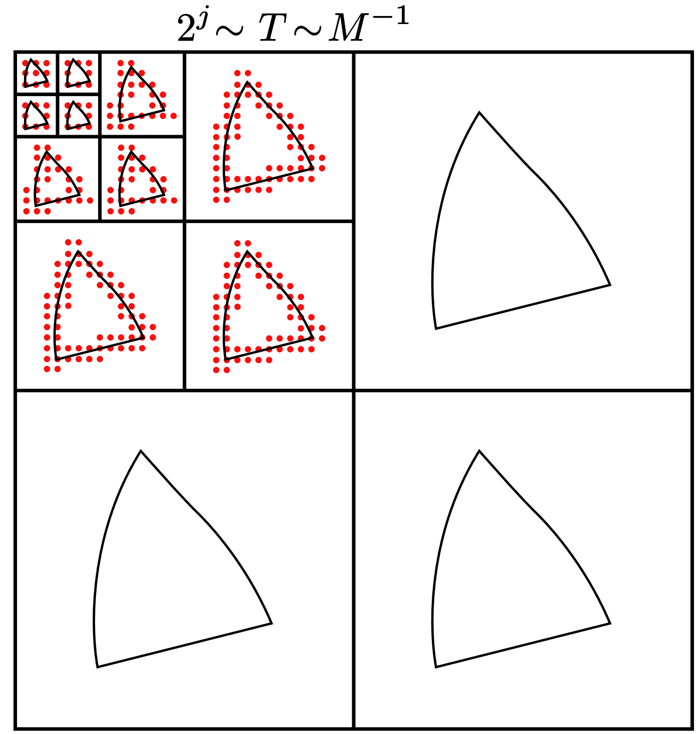



- With $M$ wavelets: $\|f - f_M\|^2 \leq C\, M^{-1}$.
- (*Cohen, DeVore, Petrushev, Xue*): Optimal for bounded variation functions: $\|f - f_M\|^2 \leq C\, M^{-1}$.

# Geometric Image and Wavelet Approximation

- Approximations of $f$ which is $\mathbf{C}^\alpha$ away from $\mathbf{C}^\alpha$ "edge" curves:



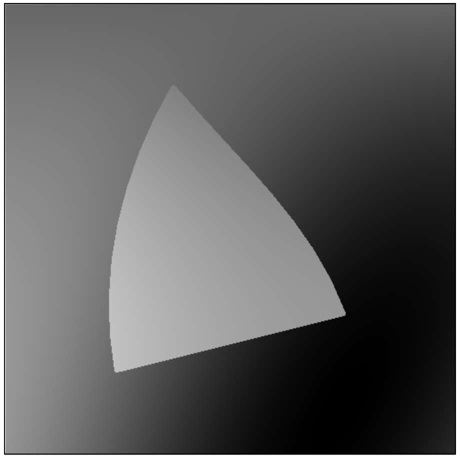



- With $M$ wavelets: $\|f - f_M\|^2 \leq C\, M^{-1}$.
- (*Cohen, DeVore, Petrushev, Xue*): Optimal for bounded variation functions: $\|f - f_M\|^2 \leq C\, M^{-1}$.
- But: does not take advantage of any geometric regularity.

# Geometric Elements for Contours

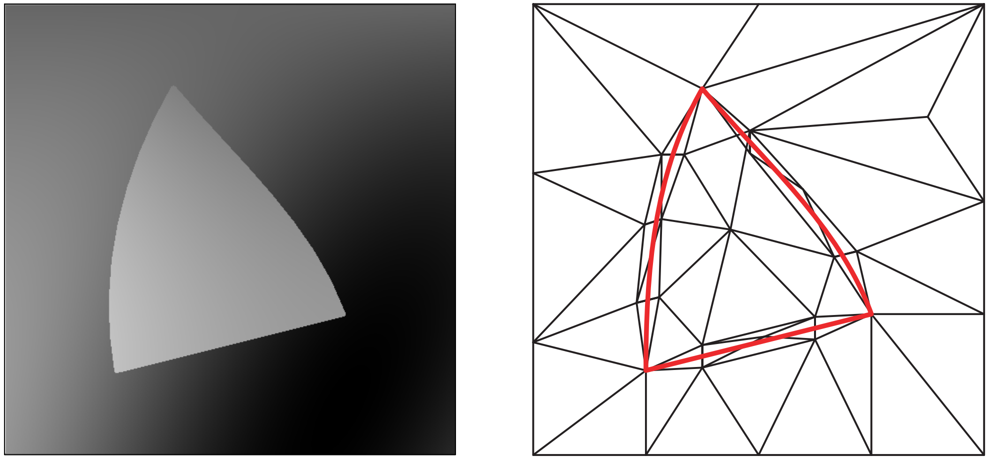
# Geometric Elements for Contours

- Approximation of $f$ that is $\mathbf{C}^\alpha$ outside $\mathbf{C}^\alpha$ contours:

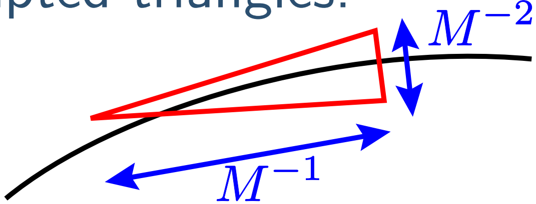


# Geometric Elements for Contours

- Approximation of $f$ that is $\mathbf{C}^\alpha$ outside $\mathbf{C}^\alpha$ contours:

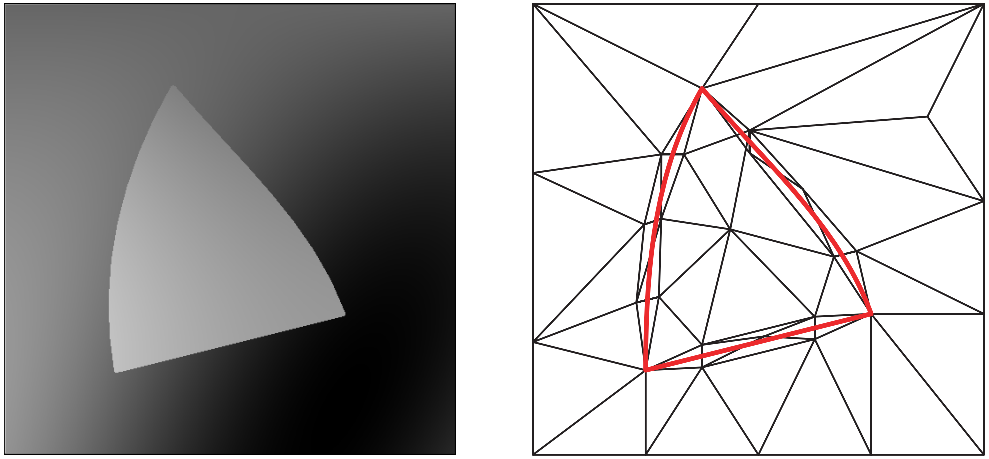


- Piecewise linear approximation with $M$ adapted triangles:
  if $\alpha \geq 2$ then $\|f - f_M\|^2 \leq C\, M^{-2}$.

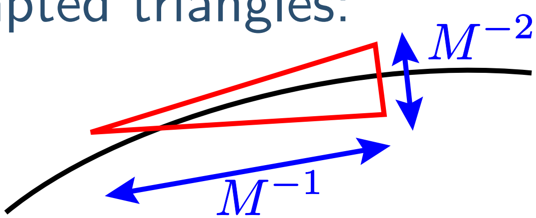


# Geometric Elements for Contours

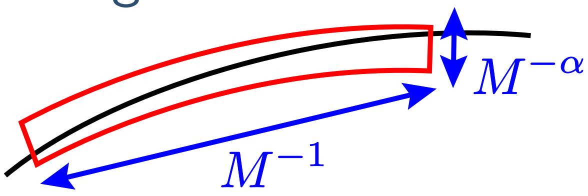
- Approximation of $f$ that is $\mathbf{C}^\alpha$ outside $\mathbf{C}^\alpha$ contours:



- Piecewise linear approximation with $M$ adapted triangles:
  if $\alpha \geq 2$ then $\|f - f_M\|^2 \leq C\, M^{-2}$.

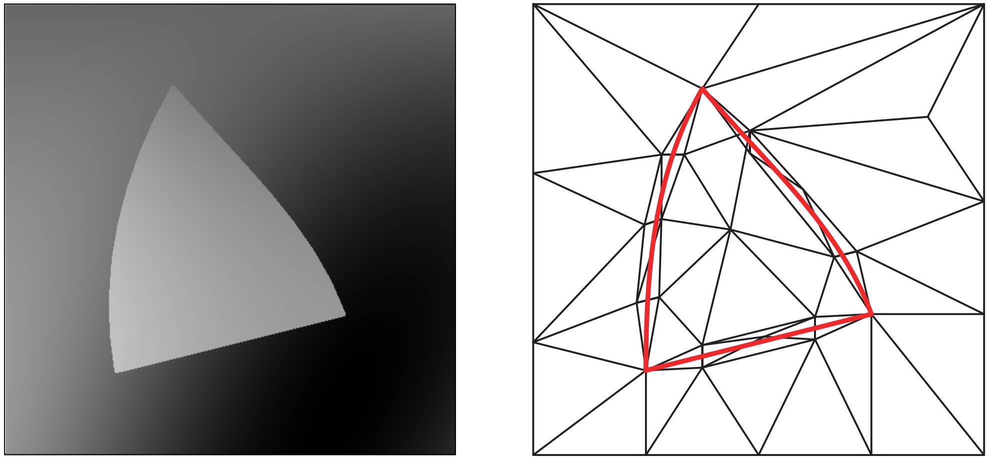



- Higher order approximation with $M$ adapted "higher order elements":
  $\|f - f_M\|^2 \leq C\, M^{-\alpha}$.

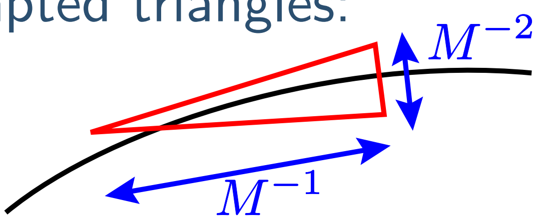


# Geometric Elements for Contours

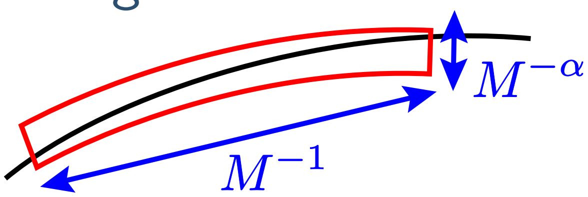
- Approximation of $f$ that is $\mathbf{C}^\alpha$ outside $\mathbf{C}^\alpha$ contours:



- Piecewise linear approximation with $M$ adapted triangles:
  if $\alpha \geq 2$ then $\|f - f_M\|^2 \leq C\, M^{-2}$.



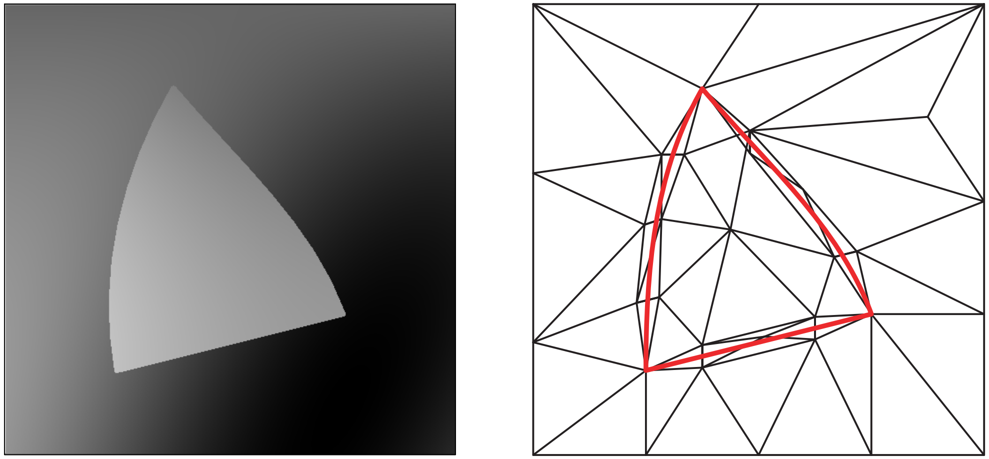

- Higher order approximation with $M$ adapted "higher order elements":
  $\|f - f_M\|^2 \leq C\, M^{-\alpha}$.



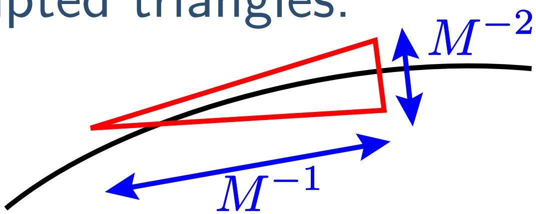

- Not a basis and difficult optimization.

# Geometric Elements for Contours

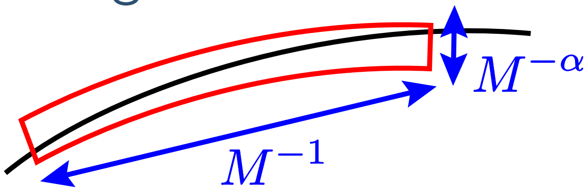
- Approximation of $f$ that is $\mathbf{C}^\alpha$ outside $\mathbf{C}^\alpha$ contours:



- Piecewise linear approximation with $M$ adapted triangles:
if $\alpha \geq 2$ then $\|f - f_M\|^2 \leq C\, M^{-2}$.



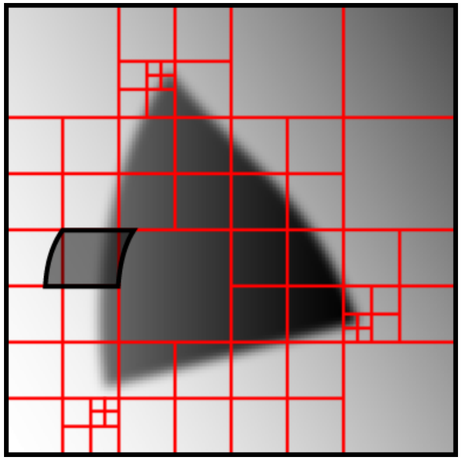

- Higher order approximation with $M$ adapted "higher order elements":
$\|f - f_M\|^2 \leq C\, M^{-\alpha}$.




- Not a basis and difficult optimization.
- If $\alpha = 2$ : curvelet tight frame is almost optimal.

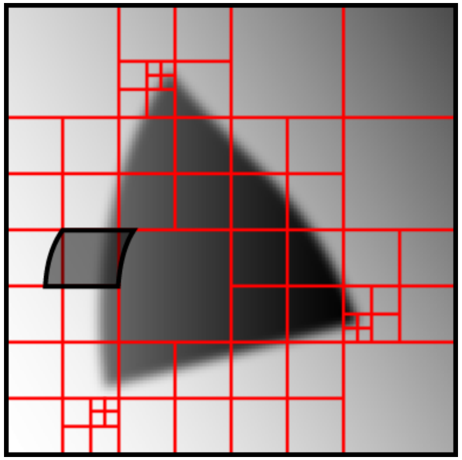
# Bandlets

# Bandlets



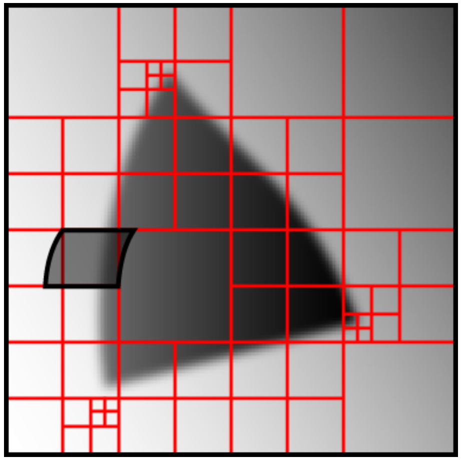
- Image $\mathbf{C}^{\alpha} - \mathbf{C}^{\alpha}$ simple

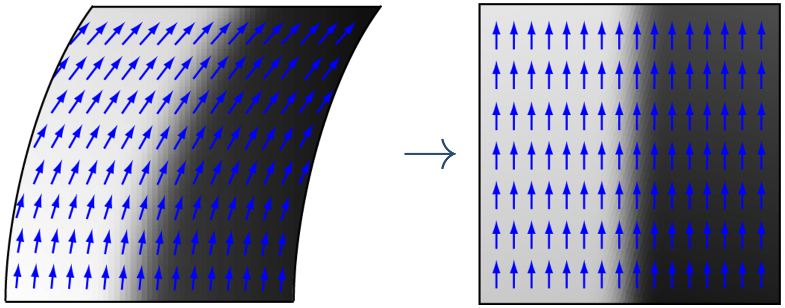
# Bandlets



- Image $\mathbf{C}^{\alpha} - \mathbf{C}^{\alpha}$ simple piecewise.

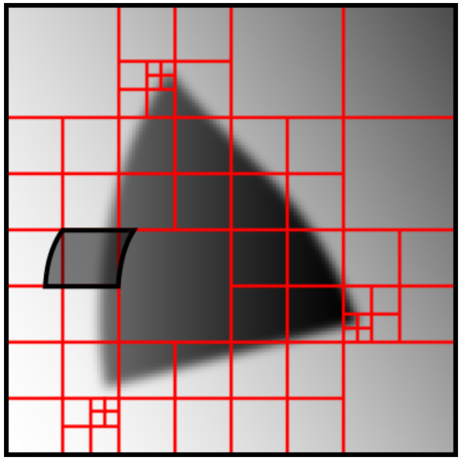
# Bandlets

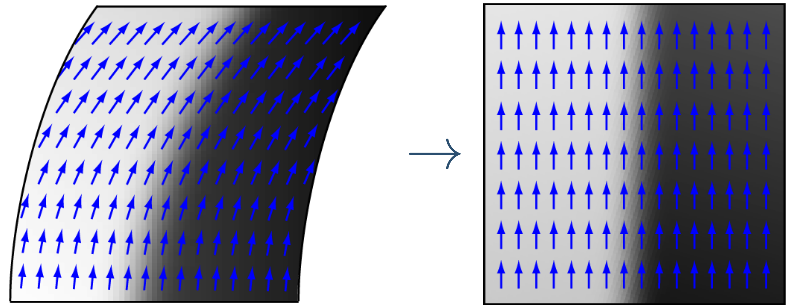




- Image $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ simple piecewise.
- Locale deformation $\implies$ vertical/horizontal singularities.

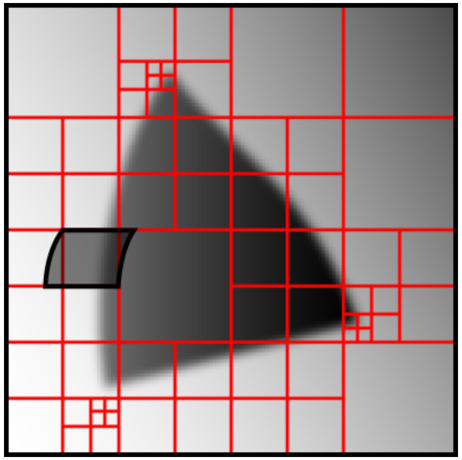
# Bandlets

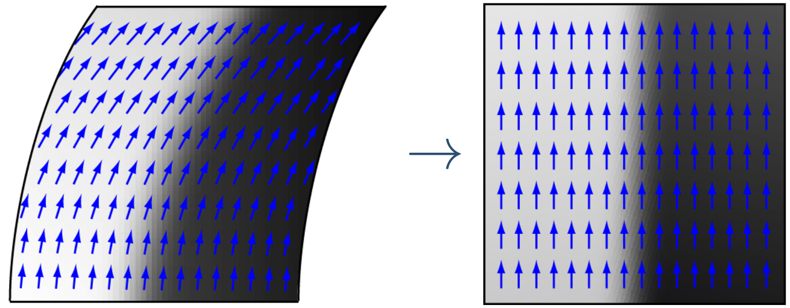


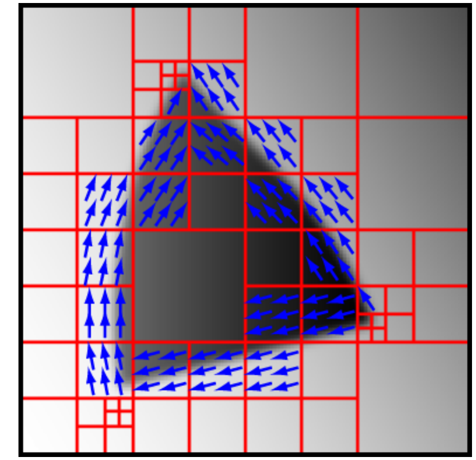


- Image $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ simple piecewise.
- Locale deformation $\Longrightarrow$ vertical/horizontal singularities.
- Local bandlets: preimage of an adapted basis.

# Bandlets

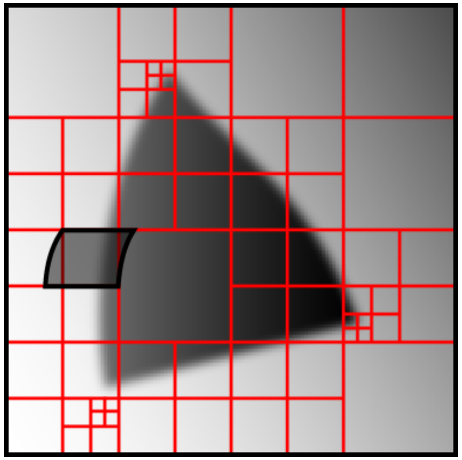


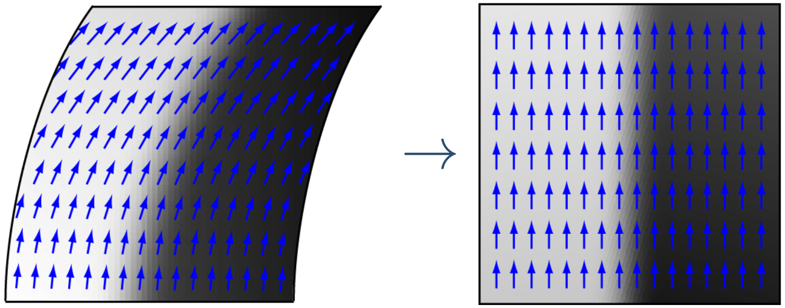


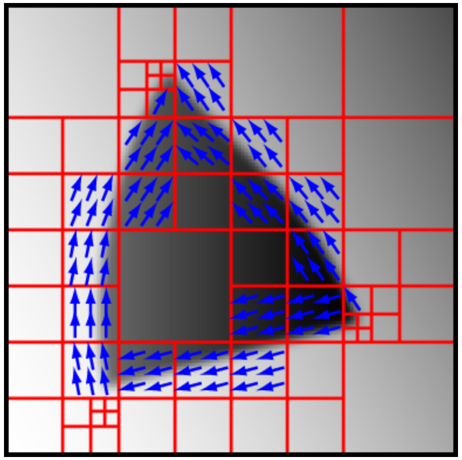


- Image $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ simple piecewise.
- Locale deformation $\Longrightarrow$ vertical/horizontal singularities.
- Local bandlets: preimage of an adapted basis.
- Bandlet basis: dyadic segmentation + a geometry per square.

# Bandlets



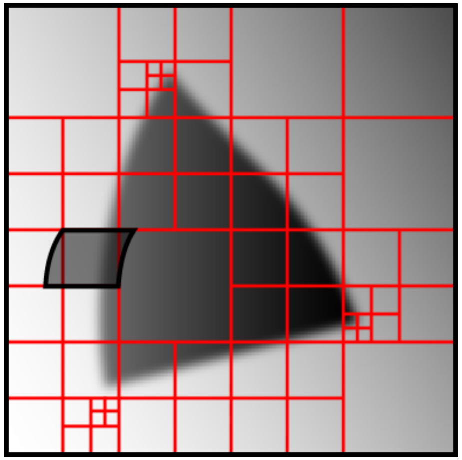


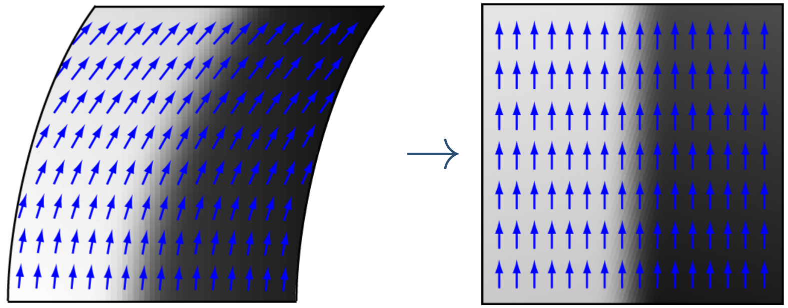


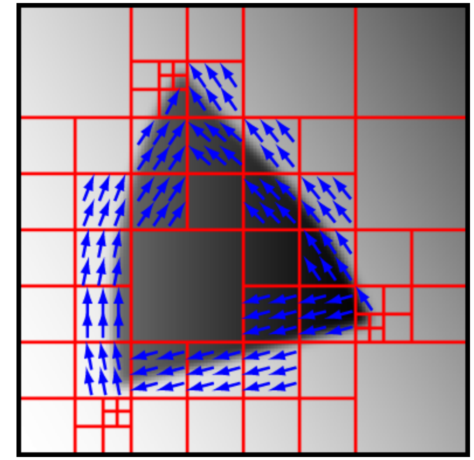
- Image $\mathbf{C}^{\alpha} - \mathbf{C}^{\alpha}$ simple piecewise.
- Locale deformation $\Longrightarrow$ vertical/horizontal singularities.
- Local bandlets: preimage of an adapted basis.
- Bandlet basis: dyadic segmentation + a geometry per square.
- **Theorem:** If $f$ is $\mathbf{C}^{\alpha} - \mathbf{C}^{\alpha}$, then, in a best basis,

$$\|f - f_M\|^2 \leq C(\log M) M^{-\alpha} \quad .$$

# Bandlets





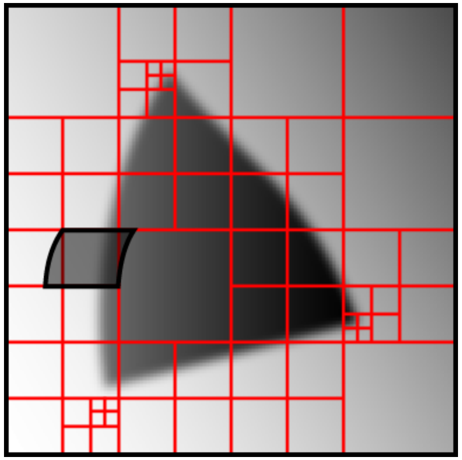


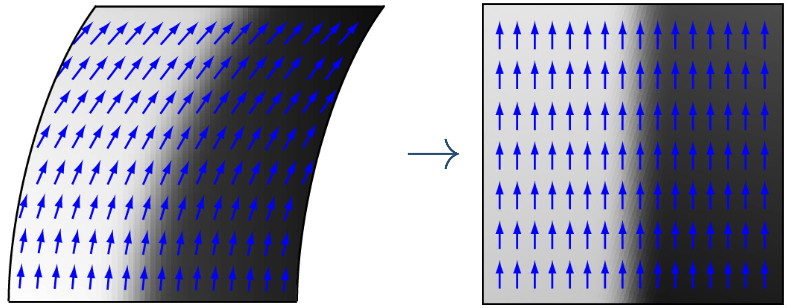
- Image $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ simple piecewise.
- Locale deformation $\Longrightarrow$ vertical/horizontal singularities.
- Local bandlets: preimage of an adapted basis.
- Bandlet basis: dyadic segmentation + a geometry per square.
- **Theorem:** If $f$ is $\mathbf{C}^\alpha - \mathbf{C}^\alpha$, then, in a best basis,

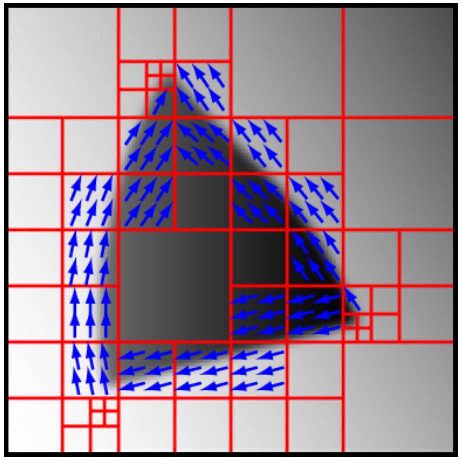
$$\|f - f_M\|^2 \leq C(\log M) M^{-\alpha} \quad .$$

- Lagrangian approach: minimization of $\|f - f_M\|^2 + T^2 M \Longrightarrow$ thresholding in a fixed basis (easy) and best basis search (difficult).

# Bandlets





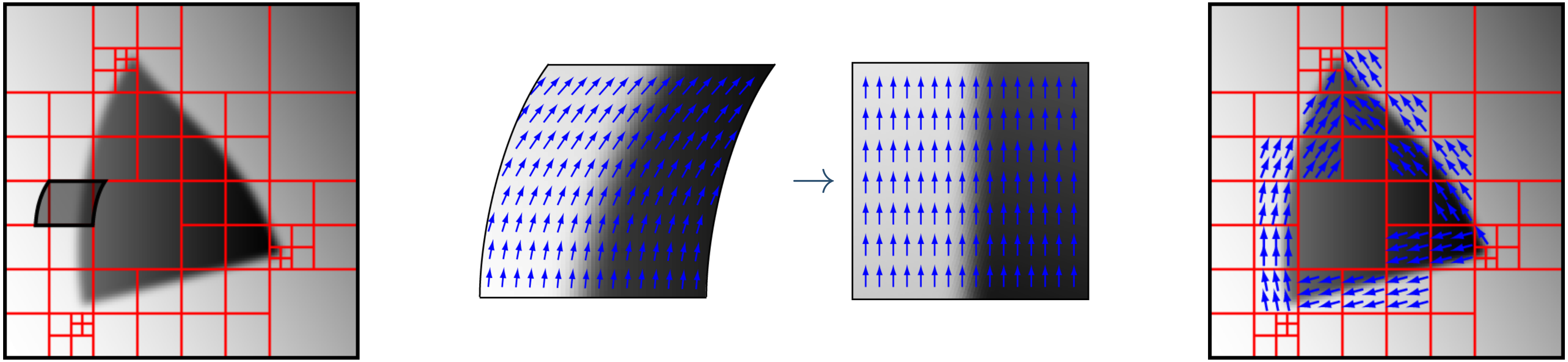


- Image $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ simple piecewise.
- Locale deformation $\Longrightarrow$ vertical/horizontal singularities.
- Local bandlets: preimage of an adapted basis.
- Bandlet basis: dyadic segmentation + a geometry per square.
- **Theorem:** If $f$ is $\mathbf{C}^\alpha - \mathbf{C}^\alpha$, then, in a best basis,

$$\|f - f_M\|^2 \leq C(\log M) M^{-\alpha} \quad .$$

- Lagrangian approach: minimization of $\|f - f_M\|^2 + T^2 M \Longrightarrow$ thresholding in a fixed basis (easy) and best basis search (difficult).
- Hierarchical structure of the segmentation and additivity of the Lagrangian : Wickerhauser's best basis algorithm (CART).

# Bandlets



- Image $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ simple piecewise.
- Locale deformation $\Longrightarrow$ vertical/horizontal singularities.
- Local bandlets: preimage of an adapted basis.
- Bandlet basis: dyadic segmentation + a geometry per square.
- **Theorem:** If $f$ is $\mathbf{C}^\alpha - \mathbf{C}^\alpha$, then, in a best basis,

$$\|f - f_M\|^2 \leq C(\log M) M^{-\alpha} \quad .$$

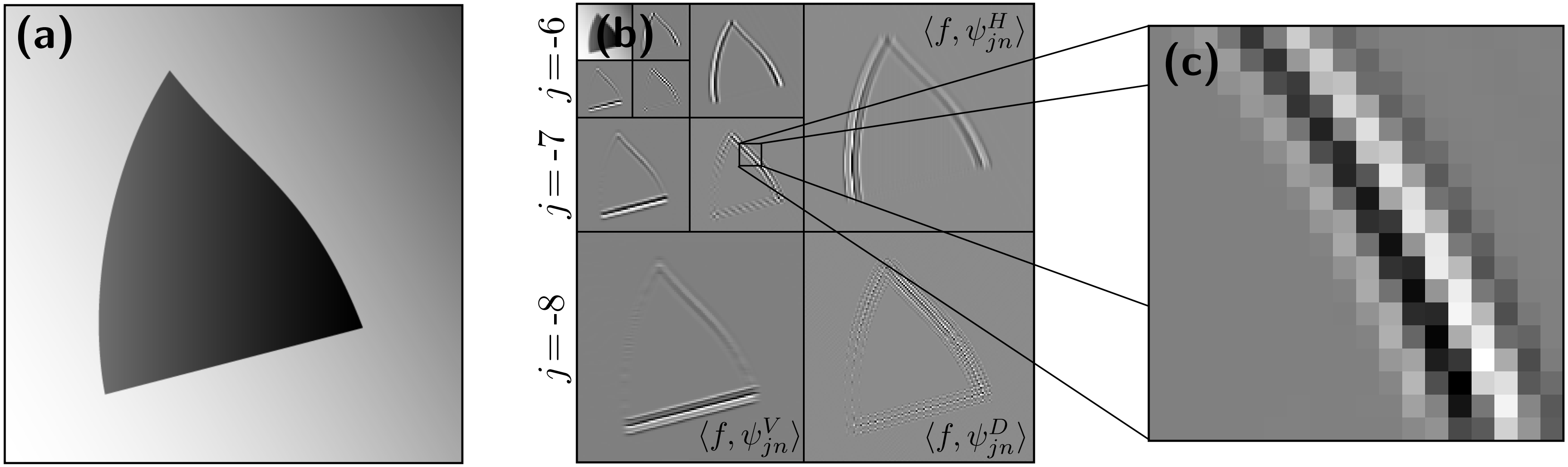
- Lagrangian approach: minimization of $\|f - f_M\|^2 + T^2 M \Longrightarrow$ thresholding in a fixed basis (easy) and best basis search (difficult).
- Hierarchical structure of the segmentation and additivity of the Lagrangian : Wickerhauser's best basis algorithm (CART).
- Exhaustive exploration of the geometries in each square.

# Bandlets and Multiscale Geometry

# Bandlets and Multiscale Geometry

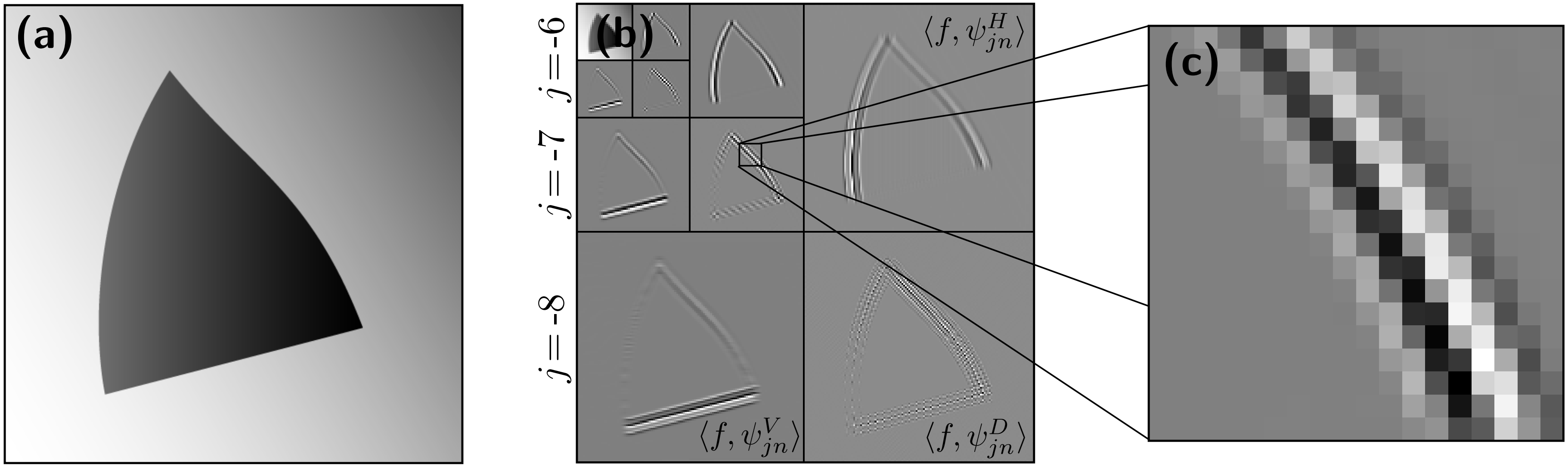
- Bandlets: monoscale geometric representation.

# Bandlets and Multiscale Geometry



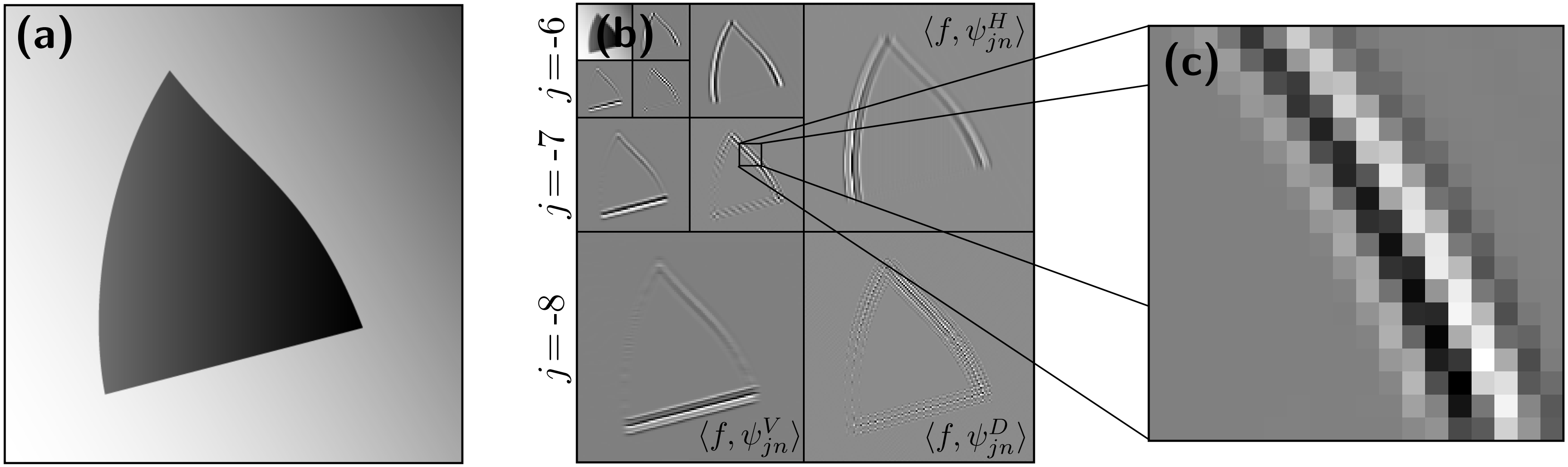

- Bandlets: monoscale geometric representation.
- Wavelets: good multiscale representation.

# Bandlets and Multiscale Geometry



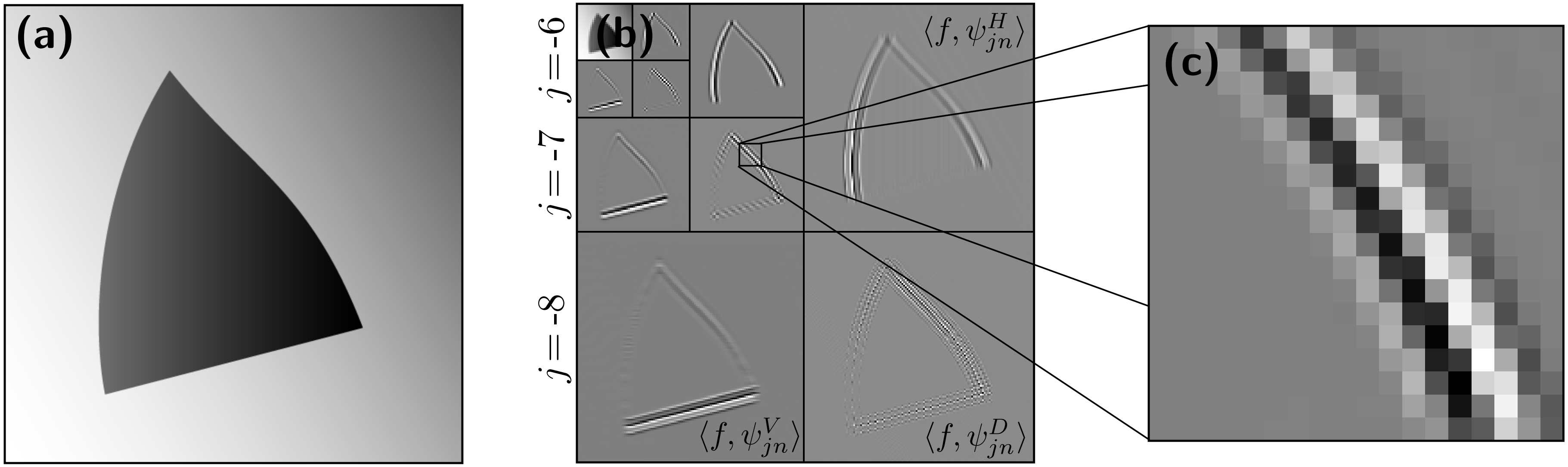

- Bandlets: monoscale geometric representation.
- Wavelets: good multiscale representation.
- How to combine both?

# Bandlets and Multiscale Geometry



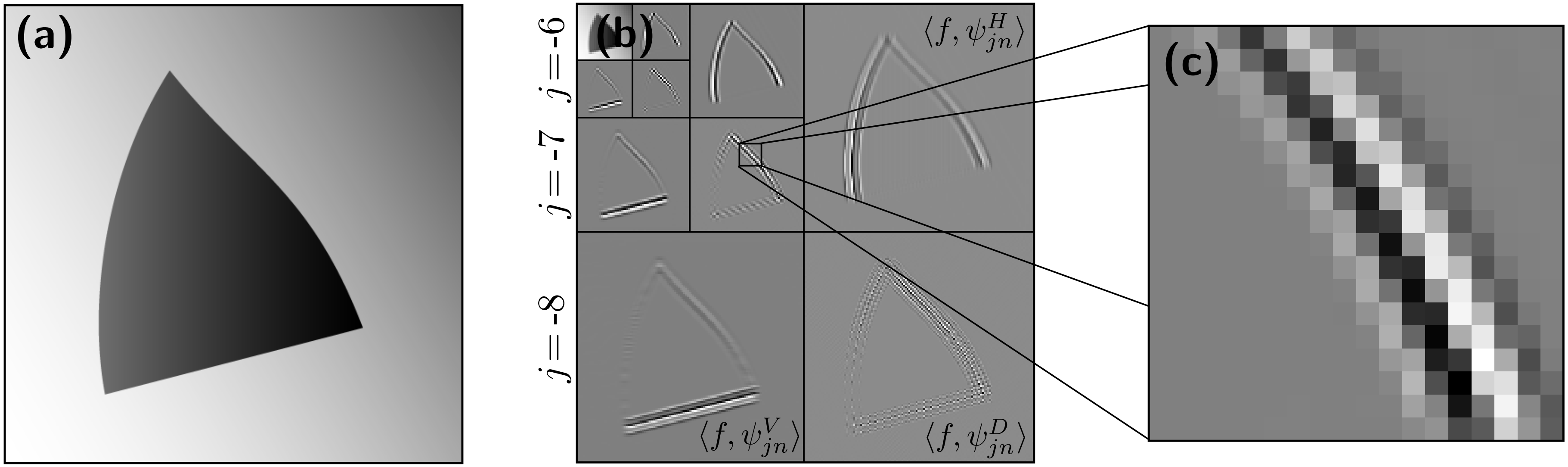

- Bandlets: monoscale geometric representation.
- Wavelets: good multiscale representation.
- How to combine both?
- Bandlets on the wavelet coefficients!

# Bandlets and Multiscale Geometry




- Bandlets: monoscale geometric representation.
- Wavelets: good multiscale representation.
- How to combine both?
- Bandlets on the wavelet coefficients!
- Analogy with the visual system.

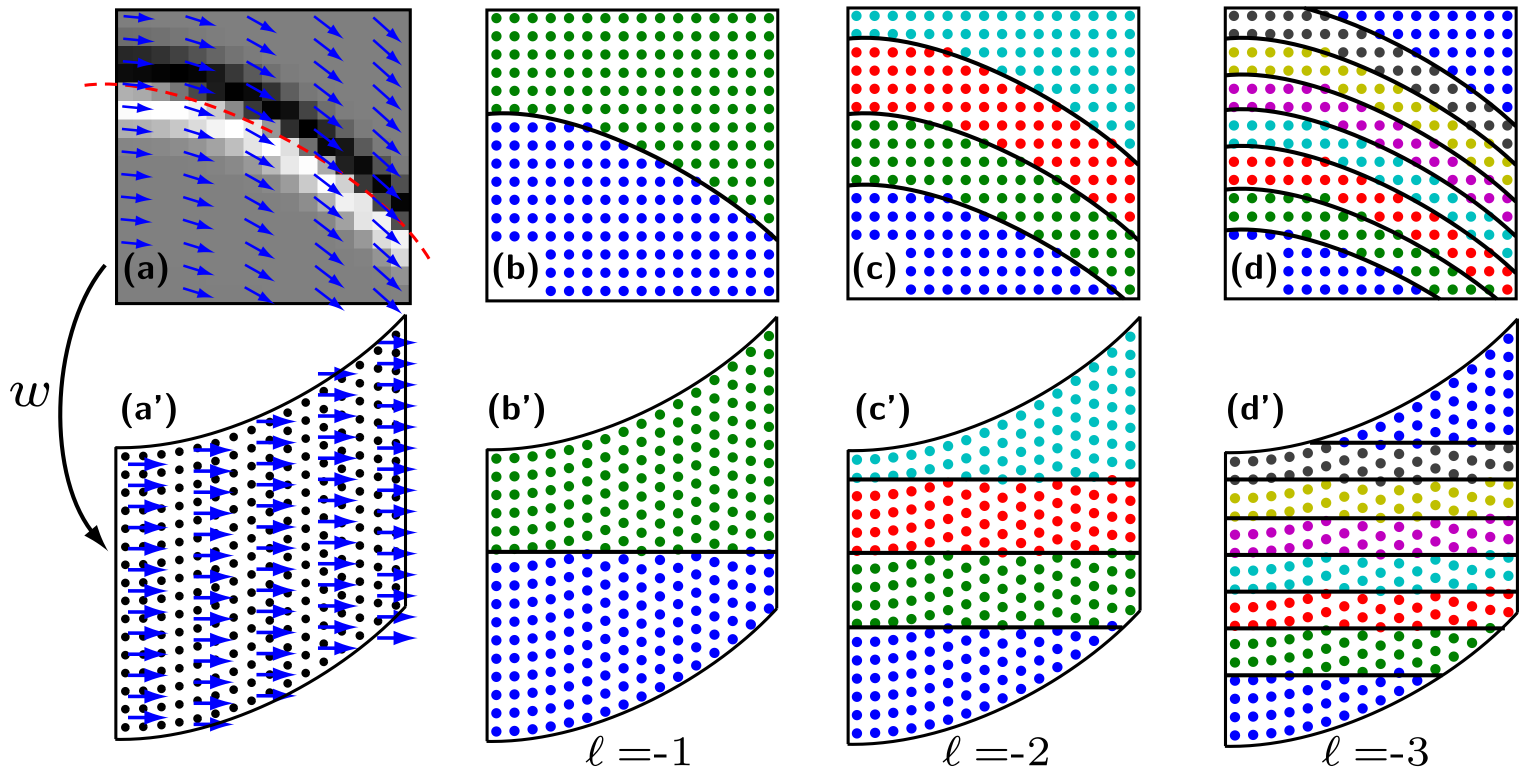
# Bandlets and Multiscale Geometry




- Bandlets: monoscale geometric representation.
- Wavelets: good multiscale representation.
- How to combine both?
- Bandlets on the wavelet coefficients!
- Analogy with the visual system.
- Similar ideas in JPEG2000, Edgeprint,...

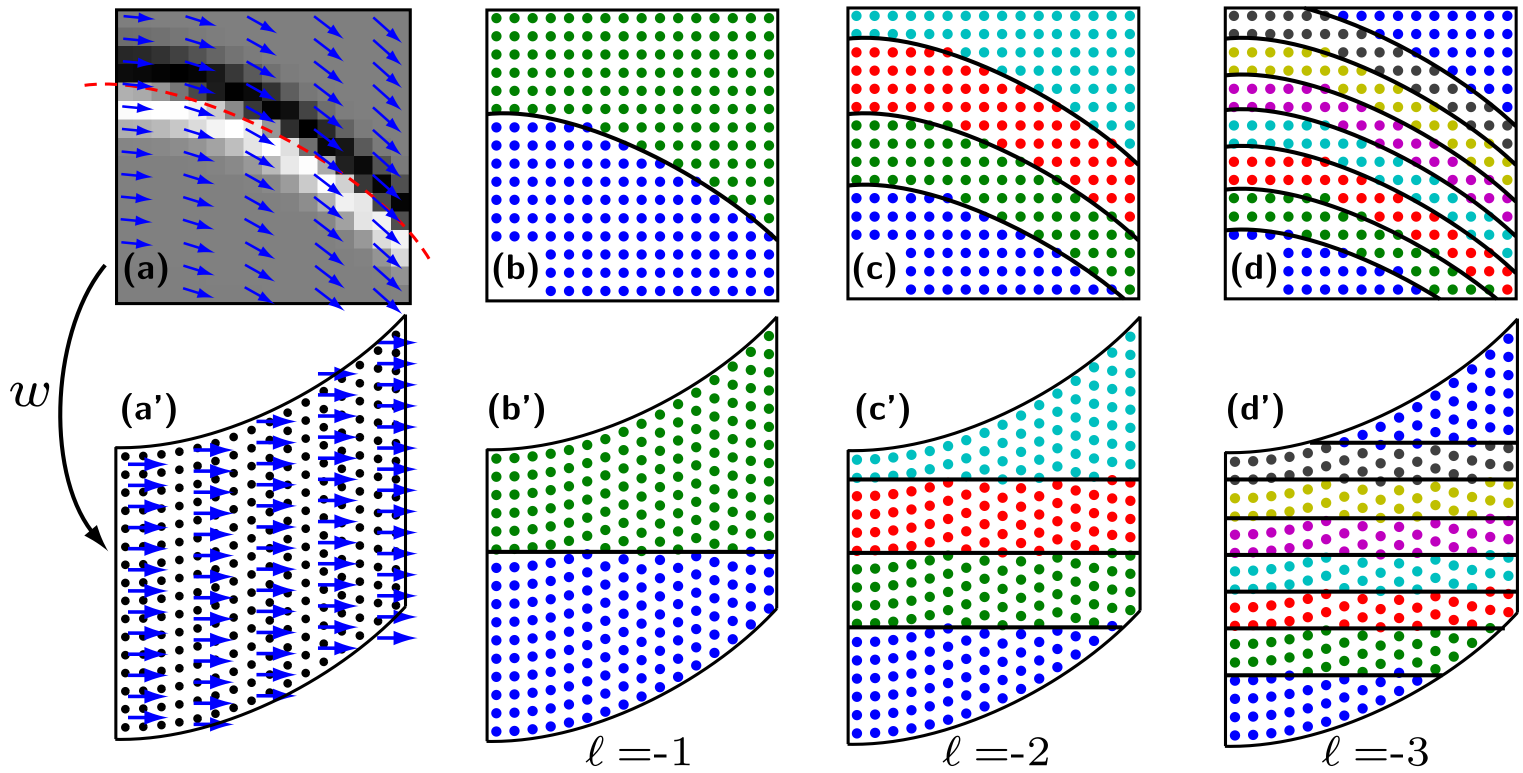
# Local Bandlet Basis

# Local Bandlet Basis



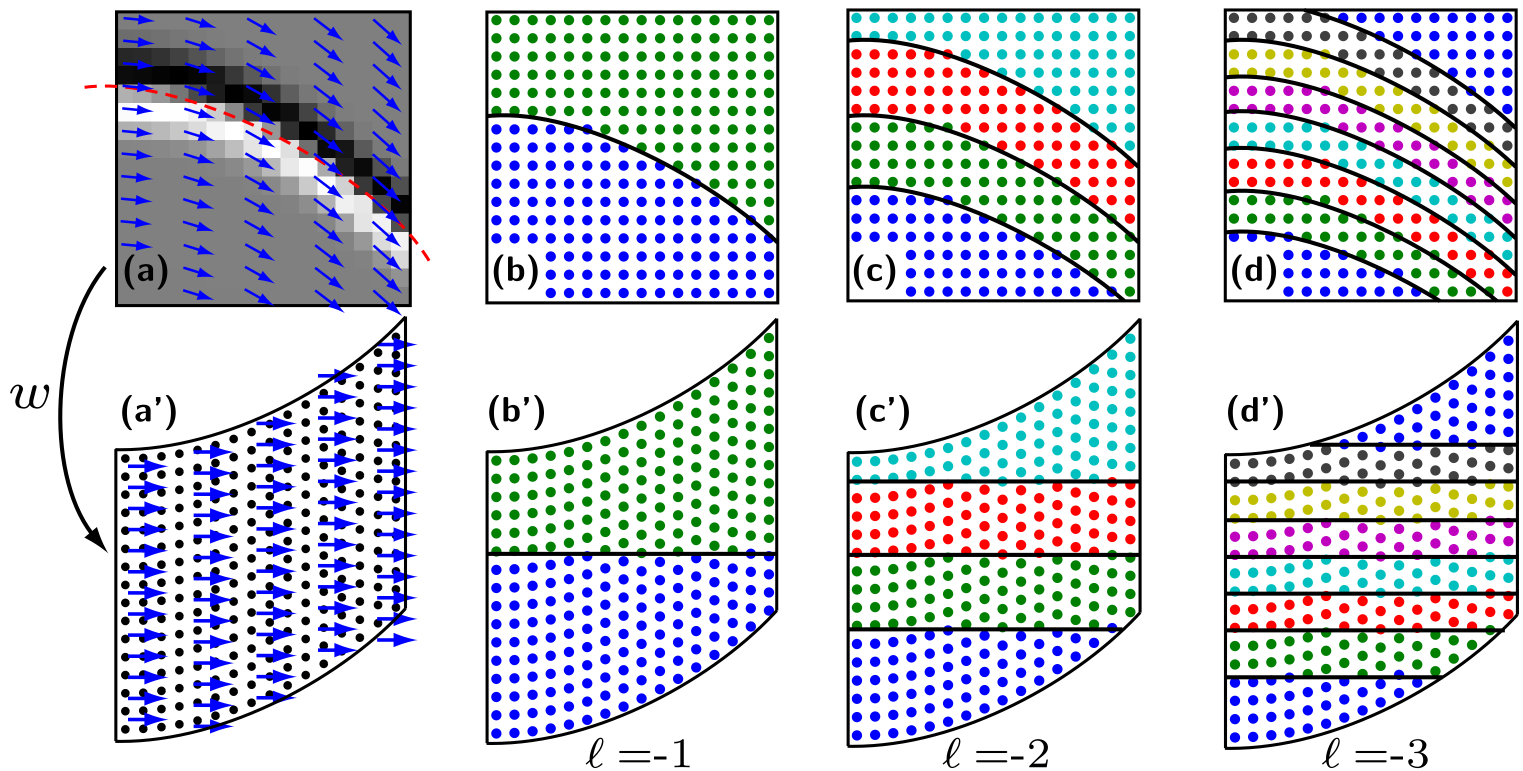

- Bandlets 2G (*Peyré*) : orthogonal change of basis on the wavelet coefficients adapted to the geometry.

# Local Bandlet Basis



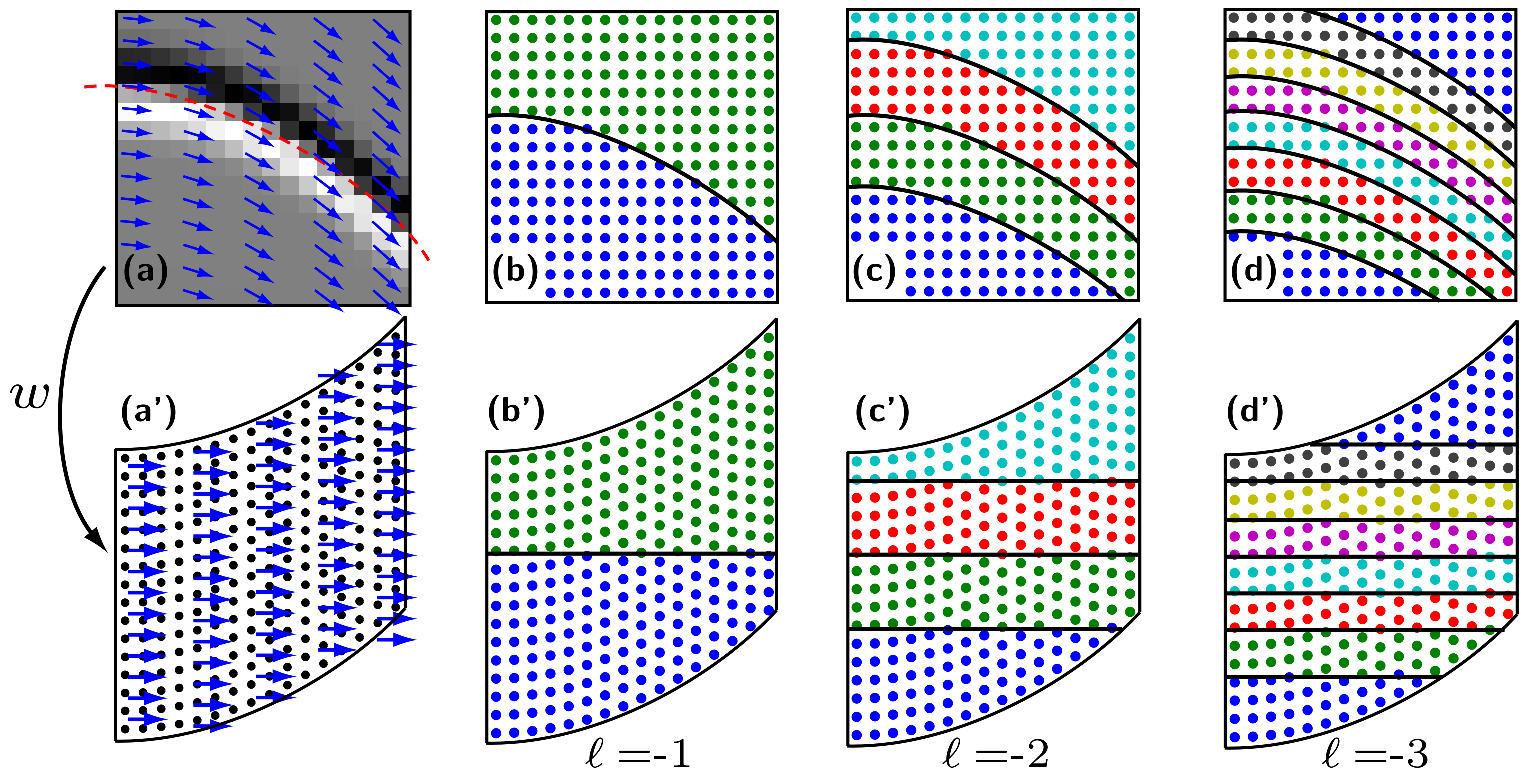

- Bandlets 2G (*Peyré*) : orthogonal change of basis on the wavelet coefficients adapted to the geometry.
- Multiresolution space of piecewise polynomial approximation.

# Local Bandlet Basis




- Bandlets 2G (*Peyré*) : orthogonal change of basis on the wavelet coefficients adapted to the geometry.
- Multiresolution space of piecewise polynomial approximation.
- Basis of the orthogonal complement spaces (*Alpert*).

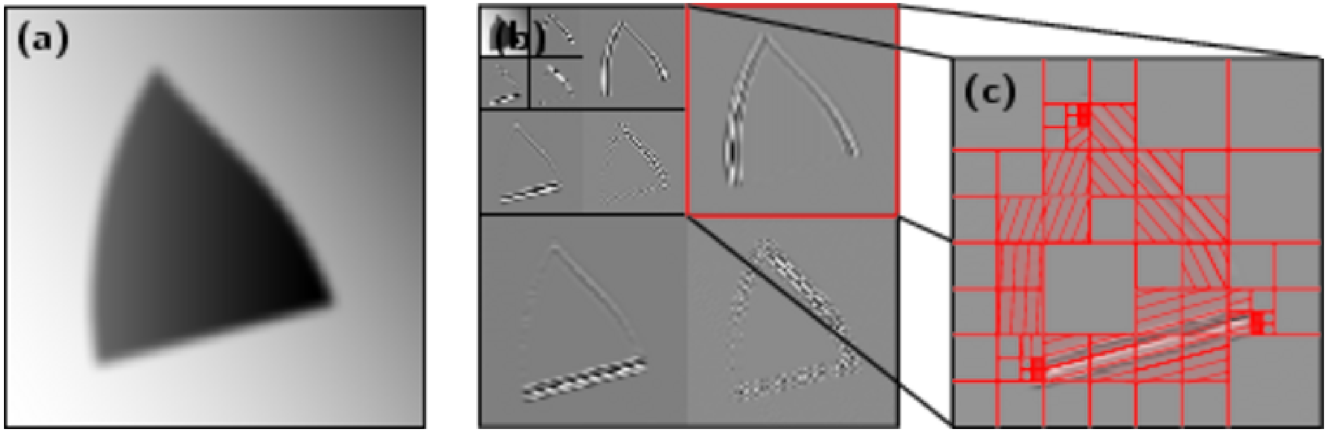
# Local Bandlet Basis




- Bandlets 2G (*Peyré*) : orthogonal change of basis on the wavelet coefficients adapted to the geometry.
- Multiresolution space of piecewise polynomial approximation.
- Basis of the orthogonal complement spaces (*Alpert*).
- Image of the wavelets through this change of basis: bandlets 2G.

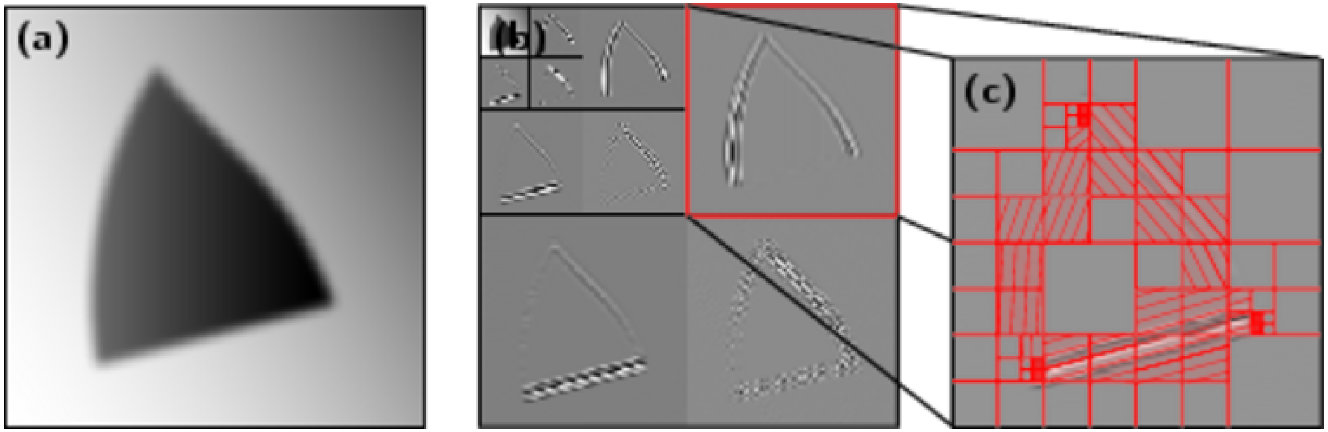
# Bandlet 2G Basis

# Bandlet 2G Basis




- Bandlet basis:
  - dyadic segmentation of the subbands,
  - geometry in each square.

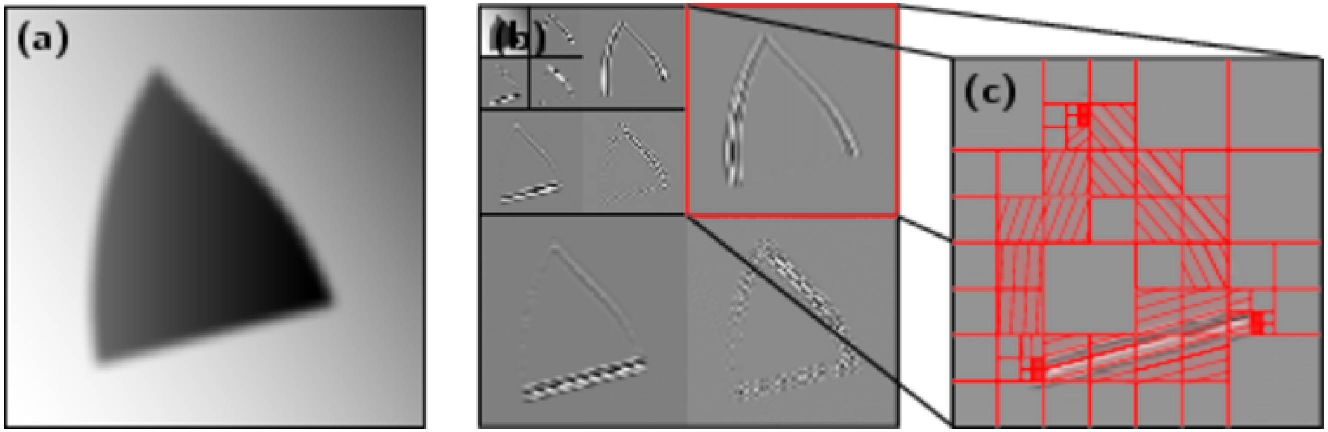
# Bandlet 2G Basis




- Bandlet basis:
  - dyadic segmentation of the subbands,
  - geometry in each square.
- **Theorem:** If $f$ est $\mathbf{C}^\alpha - \mathbf{C}^\alpha$, then, in the best basis,

$$\|f - f_M\|^2 \leq C M^{-\alpha}$$

# Bandlet 2G Basis



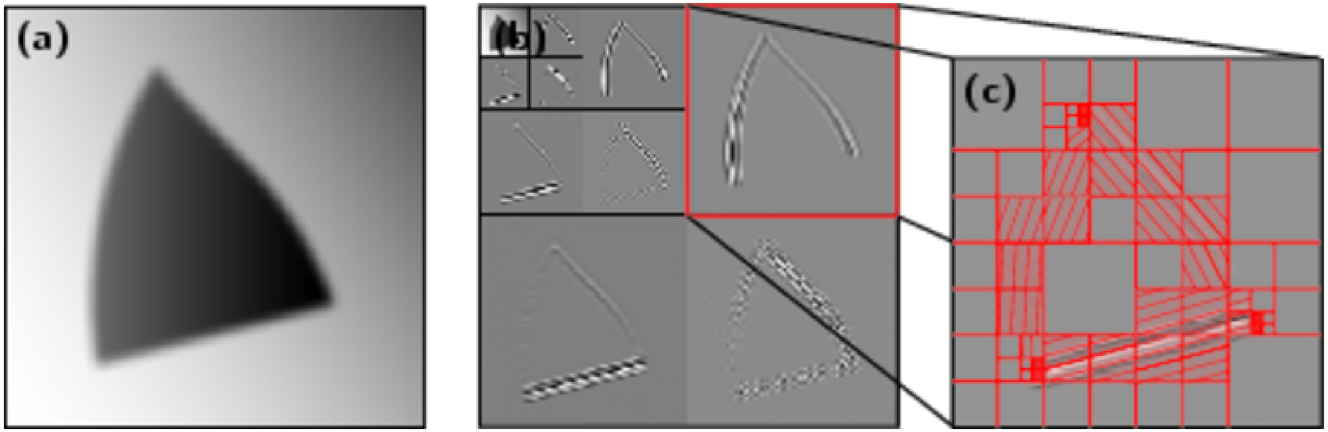

- Bandlet basis:
  - dyadic segmentation of the subbands,
  - geometry in each square.
- **Theorem:** If $f$ est $\mathbf{C}^\alpha - \mathbf{C}^\alpha$, then, in the best basis,

$$\|f - f_M\|^2 \leq C M^{-\alpha}$$

- Basis structure and polynomial number of bandlets: "fast" optimization algorithm (dynamic programming/CART) of

$$\|f - f_M\|^2 + T^2 M \quad .$$

# Bandlet 2G Basis




- Bandlet basis:
  - dyadic segmentation of the subbands,
  - geometry in each square.
- **Theorem:** If $f$ est $\mathbf{C}^\alpha - \mathbf{C}^\alpha$, then, in the best basis,

$$\|f - f_M\|^2 \leq C M^{-\alpha}$$

- Basis structure and polynomial number of bandlets: "fast" optimization algorithm (dynamic programming/CART) of

$$\|f - f_M\|^2 + T^2 M \quad .$$

- How to use these bases for estimation?

# Bandlets for Geometric Image Estimation

- Estimation in the white noise model: $Y = f + \epsilon W$ with $W$ a standard gaussian white noise and $\epsilon$ the known standard deviation.

# Bandlets for Geometric Image Estimation

- Estimation in the white noise model: $Y = f + \epsilon W$ with $W$ a standard gaussian white noise and $\epsilon$ the known standard deviation.
- Question: how to find an estimate $F$ of $f$ from $Y$ so that the quadratic risk $E(\|f - F\|^2)$ is small?

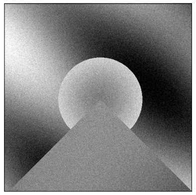
# Bandlets for Geometric Image Estimation

- Estimation in the white noise model: $Y = f + \epsilon W$ with $W$ a standard gaussian white noise and $\epsilon$ the known standard deviation.
- Question: how to find an estimate $F$ of $f$ from $Y$ so that the quadratic risk $E(\|f - F\|^2)$ is small?
- Answer: approximate $Y$ with the best bandlet basis algorithm with a threshold $T = \lambda\sqrt{\frac{\log \epsilon}{\epsilon^2}}$!

# Bandlets for Geometric Image Estimation

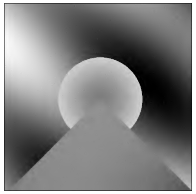
- Estimation in the white noise model: $Y = f + \epsilon W$ with $W$ a standard gaussian white noise and $\epsilon$ the known standard deviation.
- Question: how to find an estimate $F$ of $f$ from $Y$ so that the quadratic risk $E(\|f - F\|^2)$ is small?
- Answer: approximate $Y$ with the best bandlet basis algorithm with a threshold $T = \lambda\sqrt{\frac{\log \epsilon}{\epsilon^2}}$!

Noisy

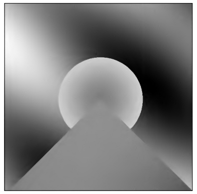


Estimation

Wavelets



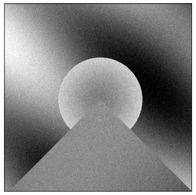
Bandlets



# Bandlets for Geometric Image Estimation

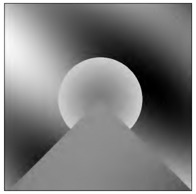
- Estimation in the white noise model: $Y = f + \epsilon W$ with $W$ a standard gaussian white noise and $\epsilon$ the known standard deviation.
- Question: how to find an estimate $F$ of $f$ from $Y$ so that the quadratic risk $E(\|f - F\|^2)$ is small?
- Answer: approximate $Y$ with the best bandlet basis algorithm with a threshold $T = \lambda\sqrt{\frac{\log \epsilon}{\epsilon^2}}$!

Noisy

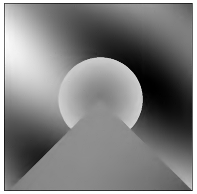


Estimation

Wavelets



Bandlets



- Questions:
  - How is this working?
  - Why is it almost optimal for $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ functions? (Minimax)
  - For which functions this method works efficiently? (Maxiset)

# Oracle Estimation in a Basis

# Oracle Estimation in a Basis

- Decomposition of $Y = f + \epsilon W$ in an orthogonal basis

$$Y = \sum_{b_n} \langle Y, b_n \rangle b_n = \sum_{b_n} \left( \langle f, b_n \rangle + \epsilon \langle W, b_n \rangle \right) b_n \quad .$$

# Oracle Estimation in a Basis

- Decomposition of $Y = f + \epsilon W$ in an orthogonal basis

$$Y = \sum_{b_n} \langle Y, b_n \rangle b_n = \sum_{b_n} \left( \langle f, b_n \rangle + \epsilon \langle W, b_n \rangle \right) b_n \quad .$$

- Estimate $F$ by a simple coordinate selection process

$$F = Y_\Gamma = \sum_{n \in \Gamma} \langle Y, b_n \rangle b_n \quad .$$

# Oracle Estimation in a Basis

- Decomposition of $Y = f + \epsilon W$ in an orthogonal basis

$$Y = \sum_{b_n} \langle Y, b_n \rangle b_n = \sum_{b_n} \left( \langle f, b_n \rangle + \epsilon \langle W, b_n \rangle \right) b_n \quad .$$

- Estimate $F$ by a simple coordinate selection process

$$F = Y_\Gamma = \sum_{n \in \Gamma} \langle Y, b_n \rangle b_n \quad .$$

- Minimization of the quadratic risk:

$$E(\|f - F\|^2) = \sum_{n \notin \Gamma} |\langle f, b_n \rangle|^2 + \sum_{n \in \Gamma} \epsilon^2 \quad .$$

# Oracle Estimation in a Basis

- Decomposition of $Y = f + \epsilon W$ in an orthogonal basis

$$Y = \sum_{b_n} \langle Y, b_n \rangle b_n = \sum_{b_n} \left( \langle f, b_n \rangle + \epsilon \langle W, b_n \rangle \right) b_n \quad .$$

- Estimate $F$ by a simple coordinate selection process

$$F = Y_\Gamma = \sum_{n \in \Gamma} \langle Y, b_n \rangle b_n \quad .$$

- Minimization of the quadratic risk:

$$E(\|f - F\|^2) = \sum_{n \notin \Gamma} |\langle f, b_n \rangle|^2 + \sum_{n \in \Gamma} \epsilon^2 \quad .$$

- Solution: $\Gamma_O = \{n, |\langle f, b_n \rangle| \geq \epsilon\}$ et $F_O = Y_{\Gamma_0}$.

# Oracle Estimation in a Basis

- Decomposition of $Y = f + \epsilon W$ in an orthogonal basis

$$Y = \sum_{b_n} \langle Y, b_n \rangle b_n = \sum_{b_n} \left( \langle f, b_n \rangle + \epsilon \langle W, b_n \rangle \right) b_n \quad .$$

- Estimate $F$ by a simple coordinate selection process

$$F = Y_\Gamma = \sum_{n \in \Gamma} \langle Y, b_n \rangle b_n \quad .$$

- Minimization of the quadratic risk:

$$E(\|f - F\|^2) = \sum_{n \notin \Gamma} |\langle f, b_n \rangle|^2 + \sum_{n \in \Gamma} \epsilon^2 \quad .$$

- Solution: $\Gamma_O = \{n, |\langle f, b_n \rangle| \geq \epsilon\}$ et $F_O = Y_{\Gamma_0}$.
- Big issue: requires the knowledge of $f$! (Oracle$\neq$estimate)

# Oracle, Risk and Approximation

# Oracle, Risk and Approximation

- Quadratic risk of the oracle estimator $F_O$:

$$E(\|f - F_O\|^2) = \sum_{n \notin \Gamma_O} |\langle f, b_n \rangle|^2 + \sum_{n \in \Gamma_O} \epsilon^2$$

$$E(\|f - F_O\|^2) = \|f - f_{\Gamma_O}\|^2 + \epsilon^2 |\Gamma_O| \quad .$$

# Oracle, Risk and Approximation

- Quadratic risk of the oracle estimator $F_O$:

$$E(\|f - F_O\|^2) = \sum_{n \notin \Gamma_O} |\langle f, b_n \rangle|^2 + \sum_{n \in \Gamma_O} \epsilon^2$$

$$E(\|f - F_O\|^2) = \|f - f_{\Gamma_O}\|^2 + \epsilon^2 |\Gamma_O| \quad .$$

- Trade-off between approximation error and number of terms.

# Oracle, Risk and Approximation

- Quadratic risk of the oracle estimator $F_O$:

$$E(\|f - F_O\|^2) = \sum_{n \notin \Gamma_O} |\langle f, b_n \rangle|^2 + \sum_{n \in \Gamma_O} \epsilon^2$$

$$E(\|f - F_O\|^2) = \|f - f_{\Gamma_O}\|^2 + \epsilon^2 |\Gamma_O| \quad .$$

- Trade-off between approximation error and number of terms.
- Approximation theory:

$$E(\|f - F_O\|^2) = \|f - f_{\Gamma_O}\|^2 + \epsilon^2 |\Gamma_O| \leq C \left(\epsilon^2\right)^{\frac{\beta}{\beta+1}}$$

$$\Leftrightarrow \|f - f_M\|^2 \leq C M^{-\beta} \Leftrightarrow f \in \mathcal{A}^{\beta} \quad .$$

# Oracle, Risk and Approximation

- Quadratic risk of the oracle estimator $F_O$:

$$E(\|f - F_O\|^2) = \sum_{n \notin \Gamma_O} |\langle f, b_n \rangle|^2 + \sum_{n \in \Gamma_O} \epsilon^2$$
$$E(\|f - F_O\|^2) = \|f - f_{\Gamma_O}\|^2 + \epsilon^2 |\Gamma_O| \quad .$$

- Trade-off between approximation error and number of terms.
- Approximation theory:

$$E(\|f - F_O\|^2) = \|f - f_{\Gamma_O}\|^2 + \epsilon^2 |\Gamma_O| \leq C \left(\epsilon^2\right)^{\frac{\beta}{\beta+1}}$$
$$\Leftrightarrow \|f - f_M\|^2 \leq C M^{-\beta} \Leftrightarrow f \in \mathcal{A}^\beta \quad .$$

- How to measure the performance of a given estimator? Minimax or Maxiset?

# Minimax or Maxiset

# Minimax or Maxiset

- Minimax: for a given function class $\Theta$ and a given estimator $F$, what is the largest $\beta$ such that

$$\forall f \in \Theta, \exists C, \forall \epsilon, E(\|f - F\|^2) \leq C \left(\epsilon^2\right)^{\frac{\beta}{\beta+1}} \quad .$$

# Minimax or Maxiset

- Minimax: for a given function class $\Theta$ and a given estimator $F$, what is the largest $\beta$ such that
$$\forall f \in \Theta, \exists C, \forall \epsilon, E(\|f - F\|^2) \leq C \left(\epsilon^2\right)^{\frac{\beta}{\beta+1}} \quad .$$
- Minimax rate : best possible rate over all estimator = largest possible $\beta$ (Complexity issue)

# Minimax or Maxiset

- Minimax: for a given function class $\Theta$ and a given estimator $F$, what is the largest $\beta$ such that
$$\forall f \in \Theta, \exists C, \forall \epsilon, E(\|f - F\|^2) \leq C \left(\epsilon^2\right)^{\frac{\beta}{\beta+1}} \quad .$$
- Minimax rate : best possible rate over all estimator $=$ largest possible $\beta$ (Complexity issue)
- Maxiset: for a given $\beta$, and a given estimator $F$, which are the functions such that
$$\exists C, \forall \epsilon, E(\|f - F\|^2) \leq C \left(\epsilon^2\right)^{\frac{\beta}{\beta+1}} \quad .$$

# Minimax or Maxiset

- Minimax: for a given function class $\Theta$ and a given estimator $F$, what is the largest $\beta$ such that
$$\forall f \in \Theta, \exists C, \forall \epsilon, E(\|f - F\|^2) \leq C \left(\epsilon^2\right)^{\frac{\beta}{\beta+1}} \quad .$$
- Minimax rate : best possible rate over all estimator = largest possible $\beta$ (Complexity issue)
- Maxiset: for a given $\beta$, and a given estimator $F$, which are the functions such that
$$\exists C, \forall \epsilon, E(\|f - F\|^2) \leq C \left(\epsilon^2\right)^{\frac{\beta}{\beta+1}} \quad .$$
- Approximation theory:
$$E(\|f - F_O\|^2) = \|f - f_{\Gamma_O}\|^2 + \epsilon^2 |\Gamma_O| \leq C \left(\epsilon^2\right)^{\frac{\beta}{\beta+1}}$$
$$\Leftrightarrow \|f - f_M\|^2 \leq C M^{-\beta} \Leftrightarrow f \in \mathcal{A}^\beta \quad .$$

# Minimax or Maxiset

- Minimax: for a given function class $\Theta$ and a given estimator $F$, what is the largest $\beta$ such that
$$\forall f \in \Theta, \exists C, \forall \epsilon, E(\|f - F\|^2) \leq C \left(\epsilon^2\right)^{\frac{\beta}{\beta+1}} \quad .$$
- Minimax rate : best possible rate over all estimator = largest possible $\beta$ (Complexity issue)
- Maxiset: for a given $\beta$, and a given estimator $F$, which are the functions such that
$$\exists C, \forall \epsilon, E(\|f - F\|^2) \leq C \left(\epsilon^2\right)^{\frac{\beta}{\beta+1}} \quad .$$
- Approximation theory:
$$E(\|f - F_O\|^2) = \|f - f_{\Gamma_O}\|^2 + \epsilon^2 |\Gamma_O| \leq C \left(\epsilon^2\right)^{\frac{\beta}{\beta+1}}$$
$$\Leftrightarrow \|f - f_M\|^2 \leq C M^{-\beta} \Leftrightarrow f \in \mathcal{A}^\beta \quad .$$
- Oracle minimax : for $\Theta$, a given function class, which basis gives $\Theta \subset \mathcal{A}^\beta$ with a large $\beta$ ? / Is $\left(\epsilon^2\right)^{\frac{\beta}{\beta+1}}$ the minimax rate?.

# Minimax or Maxiset

- Minimax: for a given function class $\Theta$ and a given estimator $F$, what is the largest $\beta$ such that
$$\forall f \in \Theta, \exists C, \forall \epsilon, E(\|f - F\|^2) \leq C \left(\epsilon^2\right)^{\frac{\beta}{\beta+1}} \quad .$$
- Minimax rate : best possible rate over all estimator = largest possible $\beta$ (Complexity issue)
- Maxiset: for a given $\beta$, and a given estimator $F$, which are the functions such that
$$\exists C, \forall \epsilon, E(\|f - F\|^2) \leq C \left(\epsilon^2\right)^{\frac{\beta}{\beta+1}} \quad .$$
- Approximation theory:
$$E(\|f - F_O\|^2) = \|f - f_{\Gamma_O}\|^2 + \epsilon^2 |\Gamma_O| \leq C \left(\epsilon^2\right)^{\frac{\beta}{\beta+1}}$$
$$\Leftrightarrow \|f - f_M\|^2 \leq C M^{-\beta} \Leftrightarrow f \in \mathcal{A}^\beta \quad .$$
- Oracle minimax : for $\Theta$, a given function class, which basis gives $\Theta \subset \mathcal{A}^\beta$ with a large $\beta$ ? / Is $\left(\epsilon^2\right)^{\frac{\beta}{\beta+1}}$ the minimax rate?.
- Oracle maxiset : The set of functions that are estimated with the rate $\left(\epsilon^2\right)^{\frac{\beta}{\beta+1}}$ is $\mathcal{A}^\beta$.

# Thresholding estimator

# Thresholding estimator

- Oracle : $\Gamma_O = \{n, |\langle f, b_n \rangle| \geq \epsilon\}$ and $F_O = Y_{\Gamma_0}$.

# Thresholding estimator

- Oracle : $\Gamma_O = \{n, |\langle f, b_n \rangle| \geq \epsilon\}$ and $F_O = Y_{\Gamma_0}$.
- Strategy: keep the largest coefficients!

# Thresholding estimator

- Oracle : $\Gamma_O = \{n, |\langle f, b_n\rangle| \geq \epsilon\}$ and $F_O = Y_{\Gamma_0}$.
- Strategy: keep the largest coefficients!
- Thresholding: $\Gamma_S = \{n, |\langle Y, b_n\rangle| \geq T(\epsilon)$ and $n \leq C\epsilon^{-\gamma}\}$ and $F_S = Y_{\Gamma_S}$.

# Thresholding estimator

- Oracle : $\Gamma_O = \{n, |\langle f, b_n \rangle| \geq \epsilon\}$ and $F_O = Y_{\Gamma_0}$.
- Strategy: keep the largest coefficients!
- Thresholding: $\Gamma_S = \{n, |\langle Y, b_n \rangle| \geq T(\epsilon)$ and $n \leq C\epsilon^{-\gamma}\}$ and $F_S = Y_{\Gamma_S}$.
- **Theorem (*Donoho, Johnstone*):** If $T(\epsilon) = \lambda\sqrt{|\log(\epsilon)|}\epsilon$, then

$$E(\|f - F_S\|^2) \leq C|\log(\epsilon)|E(\|f - F_O\|^2)$$

$$E(\|f - F_S\|^2) \leq C \min_{\Gamma} \|f - f_\Gamma\|^2 + \lambda^2|\log(\epsilon)|\epsilon^2|\Gamma| \quad \text{more precise.}$$

# Thresholding estimator

- Oracle : $\Gamma_O = \{n, |\langle f, b_n \rangle| \geq \epsilon\}$ and $F_O = Y_{\Gamma_0}$.
- Strategy: keep the largest coefficients!
- Thresholding: $\Gamma_S = \{n, |\langle Y, b_n \rangle| \geq T(\epsilon)$ and $n \leq C\epsilon^{-\gamma}\}$ and $F_S = Y_{\Gamma_S}$.
- **Theorem (*Donoho, Johnstone*):** If $T(\epsilon) = \lambda\sqrt{|\log(\epsilon)|}\epsilon$, then

$$E(\|f - F_S\|^2) \leq C|\log(\epsilon)|E(\|f - F_O\|^2)$$

$$E(\|f - F_S\|^2) \leq C \min_{\Gamma} \|f - f_\Gamma\|^2 + \lambda^2|\log(\epsilon)|\epsilon^2|\Gamma| \quad \text{more precise.}$$

- **Theorem (Maxiset) (*Cohen, DeVore, Kerkyacharian, Picard*):**

$$E(\|f - F_S\|^2) \leq C\left(|\log(\epsilon)|\epsilon^2\right)^{\frac{\beta}{\beta+1}} \Leftrightarrow f \in V^*_{\frac{2\beta}{\beta+1}}$$

$$\Leftrightarrow \min_{\Gamma} \|f - f_\Gamma\|^2 + \lambda^2 T^2|\Gamma| \leq C(T^2)^{\frac{\beta}{\beta+1}}$$

$$\Leftrightarrow \|f - f_M\|^2 \leq CM^{-\beta} \text{ plus linear approximation property } (\gamma!)$$

$$\Leftrightarrow f \in \mathcal{A}^{\beta}{}_{\gamma} \quad .$$

# Thresholding estimator

- Oracle : $\Gamma_O = \{n, |\langle f, b_n \rangle| \geq \epsilon\}$ and $F_O = Y_{\Gamma_0}$.
- Strategy: keep the largest coefficients!
- Thresholding: $\Gamma_S = \{n, |\langle Y, b_n \rangle| \geq T(\epsilon)$ and $n \leq C\epsilon^{-\gamma}\}$ and $F_S = Y_{\Gamma_S}$.
- **Theorem (*Donoho, Johnstone*):** If $T(\epsilon) = \lambda\sqrt{|\log(\epsilon)|}\epsilon$, then

$$E(\|f - F_S\|^2) \leq C|\log(\epsilon)|E(\|f - F_O\|^2)$$

$$E(\|f - F_S\|^2) \leq C \min_{\Gamma} \|f - f_\Gamma\|^2 + \lambda^2 |\log(\epsilon)|\epsilon^2 |\Gamma| \quad \text{more precise.}$$

- **Theorem (Maxiset) (*Cohen, DeVore, Kerkyacharian, Picard*):**

$$E(\|f - F_S\|^2) \leq C \left(|\log(\epsilon)|\epsilon^2\right)^{\frac{\beta}{\beta+1}} \Leftrightarrow f \in V^*_{\frac{2\beta}{\beta+1}}$$

$$\Leftrightarrow \min_{\Gamma} \|f - f_\Gamma\|^2 + \lambda^2 T^2 |\Gamma| \leq C(T^2)^{\frac{\beta}{\beta+1}}$$

$$\Leftrightarrow \|f - f_M\|^2 \leq CM^{-\beta} \text{ plus linear approximation property } (\gamma!)$$

$$\Leftrightarrow f \in \mathcal{A}^{\beta}{}_{\gamma} \quad .$$

- Key factor: Approximation properties of the basis!

# Limits of 2D Wavelet Thresholding

# Limits of 2D Wavelet Thresholding

- For $f$ $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ ($\mathbf{C}^\alpha$ outside $\mathbf{C}^\alpha$contours) (*Korostelev, Tsybakov*): minimax rate = best possible estimation rate = $\left(\epsilon^2\right)^{\frac{\alpha}{\alpha+1}}$ ($\beta = \alpha$).

# Limits of 2D Wavelet Thresholding

- For $f$ $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ ($\mathbf{C}^\alpha$ outside $\mathbf{C}^\alpha$contours) (*Korostelev, Tsybakov*): minimax rate $=$ best possible estimation rate $= \left(\epsilon^2\right)^{\frac{\alpha}{\alpha+1}}$ ($\beta = \alpha$).
- Maxiset associated to the rate $\left(|\log(\epsilon)|\epsilon^2\right)^{\frac{\beta}{\beta+1}}$ :

$$\mathcal{A}^\beta = \mathcal{W}B^\beta{}_{2/(2\beta+1),2/(2\beta+1)} = \text{Weak version of } B^\beta{}_{2/(2\beta+1),2/(2\beta+1)} \quad .$$

# Limits of 2D Wavelet Thresholding

- For $f$ $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ ($\mathbf{C}^\alpha$ outside $\mathbf{C}^\alpha$contours) (*Korostelev, Tsybakov*): minimax rate = best possible estimation rate = $\left(\epsilon^2\right)^{\frac{\alpha}{\alpha+1}}$ ($\beta = \alpha$).
- Maxiset associated to the rate $\left(|\log(\epsilon)|\epsilon^2\right)^{\frac{\beta}{\beta+1}}$ :

$$\mathcal{A}^\beta = \mathcal{W}B^\beta{}_{2/(2\beta+1),2/(2\beta+1)} = \text{Weak version of } B^\beta{}_{2/(2\beta+1),2/(2\beta+1)} \quad .$$

- Natural question: Is $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ in $\mathcal{W}B^\alpha{}_{2/(2\alpha+1),2/(2\alpha+1)}$ ?

# Limits of 2D Wavelet Thresholding

- For $f$ $\mathbf{C}^{\alpha} - \mathbf{C}^{\alpha}$ ($\mathbf{C}^{\alpha}$ outside $\mathbf{C}^{\alpha}$contours) (*Korostelev, Tsybakov*): minimax rate = best possible estimation rate $= \left(\epsilon^2\right)^{\frac{\alpha}{\alpha+1}}$ ($\beta = \alpha$).
- Maxiset associated to the rate $\left(|\log(\epsilon)|\epsilon^2\right)^{\frac{\beta}{\beta+1}}$ :

$$\mathcal{A}^{\beta} = \mathcal{W}B^{\beta}{}_{2/(2\beta+1),2/(2\beta+1)} = \text{Weak version of } B^{\beta}{}_{2/(2\beta+1),2/(2\beta+1)} \quad .$$

- Natural question: Is $\mathbf{C}^{\alpha} - \mathbf{C}^{\alpha}$ in $\mathcal{W}B^{\alpha}{}_{2/(2\alpha+1),2/(2\alpha+1)}$ ?
- Rephrased question: Are the function of $\mathbf{C}^{\alpha} - \mathbf{C}^{\alpha}$ approximable at the rate $M^{-\alpha}$?

# Limits of 2D Wavelet Thresholding

- For $f$ $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ ($\mathbf{C}^\alpha$ outside $\mathbf{C}^\alpha$contours) (*Korostelev, Tsybakov*): minimax rate = best possible estimation rate $= \left(\epsilon^2\right)^{\frac{\alpha}{\alpha+1}}$ ($\beta = \alpha$).
- Maxiset associated to the rate $\left(|\log(\epsilon)|\epsilon^2\right)^{\frac{\beta}{\beta+1}}$ :

$$\mathcal{A}^\beta = \mathcal{W}B^\beta{}_{2/(2\beta+1),2/(2\beta+1)} = \text{Weak version of } B^\beta{}_{2/(2\beta+1),2/(2\beta+1)} \quad .$$

- Natural question: Is $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ in $\mathcal{W}B^\alpha{}_{2/(2\alpha+1),2/(2\alpha+1)}$ ?
- Rephrased question: Are the function of $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ approximable at the rate $M^{-\alpha}$?
- Answer: No, $\mathbf{C}^\alpha - \mathbf{C}^\alpha \not\subset \mathcal{W}B^\alpha{}_{2/(2\alpha+1),2/(2\alpha+1)}$!

# Limits of 2D Wavelet Thresholding

- For $f$ $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ ($\mathbf{C}^\alpha$ outside $\mathbf{C}^\alpha$contours) (*Korostelev, Tsybakov*): minimax rate = best possible estimation rate = $\left(\epsilon^2\right)^{\frac{\alpha}{\alpha+1}}$ ($\beta = \alpha$).
- Maxiset associated to the rate $\left(|\log(\epsilon)|\epsilon^2\right)^{\frac{\beta}{\beta+1}}$ :

$$\mathcal{A}^\beta = \mathcal{W}B^\beta{}_{2/(2\beta+1),2/(2\beta+1)} = \text{Weak version of } B^\beta{}_{2/(2\beta+1),2/(2\beta+1)} \quad .$$

- Natural question: Is $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ in $\mathcal{W}B^\alpha{}_{2/(2\alpha+1),2/(2\alpha+1)}$ ?
- Rephrased question: Are the function of $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ approximable at the rate $M^{-\alpha}$?
- Answer: No, $\mathbf{C}^\alpha - \mathbf{C}^\alpha \not\subset \mathcal{W}B^\alpha{}_{2/(2\alpha+1),2/(2\alpha+1)}$!
- Well known approximation fact: with $M$ wavelets $\|f - f_M\|^2 \leq C\, M^{-1}$.

# Limits of 2D Wavelet Thresholding

- For $f$ $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ ($\mathbf{C}^\alpha$ outside $\mathbf{C}^\alpha$contours) (*Korostelev, Tsybakov*): minimax rate = best possible estimation rate $= \left(\epsilon^2\right)^{\frac{\alpha}{\alpha+1}}$ $(\beta = \alpha)$.
- Maxiset associated to the rate $\left(|\log(\epsilon)|\epsilon^2\right)^{\frac{\beta}{\beta+1}}$ :

$$\mathcal{A}^\beta = \mathcal{W}B^\beta{}_{2/(2\beta+1),2/(2\beta+1)} = \text{Weak version of } B^\beta{}_{2/(2\beta+1),2/(2\beta+1)} \quad .$$

- Natural question: Is $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ in $\mathcal{W}B^\alpha{}_{2/(2\alpha+1),2/(2\alpha+1)}$ ?
- Rephrased question: Are the function of $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ approximable at the rate $M^{-\alpha}$?
- Answer: No, $\mathbf{C}^\alpha - \mathbf{C}^\alpha \not\subset \mathcal{W}B^\alpha{}_{2/(2\alpha+1),2/(2\alpha+1)}$!
- Well known approximation fact: with $M$ wavelets $\|f - f_M\|^2 \leq C\, M^{-1}$.
- Minimax risk requires $\|f - f_M\|^2 \leq C\, M^{-\alpha}$.

# Limits of 2D Wavelet Thresholding

- For $f$ $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ ($\mathbf{C}^\alpha$ outside $\mathbf{C}^\alpha$contours) (*Korostelev, Tsybakov*): minimax rate = best possible estimation rate $= \left(\epsilon^2\right)^{\frac{\alpha}{\alpha+1}}$ ($\beta = \alpha$).
- Maxiset associated to the rate $\left(|\log(\epsilon)|\epsilon^2\right)^{\frac{\beta}{\beta+1}}$ :

$$\mathcal{A}^\beta = \mathcal{W}B^\beta{}_{2/(2\beta+1),2/(2\beta+1)} = \text{Weak version of } B^\beta{}_{2/(2\beta+1),2/(2\beta+1)} \quad .$$

- Natural question: Is $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ in $\mathcal{W}B^\alpha{}_{2/(2\alpha+1),2/(2\alpha+1)}$ ?
- Rephrased question: Are the function of $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ approximable at the rate $M^{-\alpha}$?
- Answer: No, $\mathbf{C}^\alpha - \mathbf{C}^\alpha \not\subset \mathcal{W}B^\alpha{}_{2/(2\alpha+1),2/(2\alpha+1)}$!
- Well known approximation fact: with $M$ wavelets $\|f - f_M\|^2 \leq C\, M^{-1}$.
- Minimax risk requires $\|f - f_M\|^2 \leq C\, M^{-\alpha}$.
- Bandlets gives the correct rate $\Longrightarrow$ bandlets thresholding?

# Limits of 2D Wavelet Thresholding

- For $f$ $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ ($\mathbf{C}^\alpha$ outside $\mathbf{C}^\alpha$contours) (*Korostelev, Tsybakov*): minimax rate = best possible estimation rate = $\left(\epsilon^2\right)^{\frac{\alpha}{\alpha+1}}$ ($\beta = \alpha$).
- Maxiset associated to the rate $\left(|\log(\epsilon)|\epsilon^2\right)^{\frac{\beta}{\beta+1}}$ :

$$\mathcal{A}^\beta = \mathcal{W}B^\beta{}_{2/(2\beta+1),2/(2\beta+1)} = \text{Weak version of } B^\beta{}_{2/(2\beta+1),2/(2\beta+1)} \quad .$$

- Natural question: Is $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ in $\mathcal{W}B^\alpha{}_{2/(2\alpha+1),2/(2\alpha+1)}$ ?
- Rephrased question: Are the function of $\mathbf{C}^\alpha - \mathbf{C}^\alpha$ approximable at the rate $M^{-\alpha}$?
- Answer: No, $\mathbf{C}^\alpha - \mathbf{C}^\alpha \not\subset \mathcal{W}B^\alpha{}_{2/(2\alpha+1),2/(2\alpha+1)}$!
- Well known approximation fact: with $M$ wavelets $\|f - f_M\|^2 \leq C\, M^{-1}$.
- Minimax risk requires $\|f - f_M\|^2 \leq C\, M^{-\alpha}$.
- Bandlets gives the correct rate $\Longrightarrow$ bandlets thresholding?
- Big issue: how to threshold in the “best” bandlet basis without knowing $f$?

# Thresholding and Model Selection

# Thresholding and Model Selection

- Back to the oracle estimator.

# Thresholding and Model Selection

- Back to the oracle estimator.
- Oracle risk: $\Gamma_0 = \mathrm{argmin}_{\Gamma} \|f - f_{\Gamma}\|^2 + \epsilon^2 |\Gamma| \quad .$

# Thresholding and Model Selection

- Back to the oracle estimator.
- Oracle risk: $\Gamma_0 = \mathrm{argmin}_\Gamma \|f - f_\Gamma\|^2 + \epsilon^2|\Gamma| \quad .$
- Empirical risk: $\Gamma_S = \mathrm{argmin}_{\Gamma\subset[1,\epsilon^{-\gamma}]} \|Y - Y_\Gamma\|^2 + \lambda^2|\log(\epsilon)|\epsilon^2|\Gamma| \quad .$

# Thresholding and Model Selection

- Back to the oracle estimator.
- Oracle risk: $\Gamma_0 = \text{argmin}_\Gamma \|f - f_\Gamma\|^2 + \epsilon^2|\Gamma| \quad .$
- Empirical risk: $\Gamma_S = \text{argmin}_{\Gamma\subset[1,\epsilon^{-\gamma}]} \|Y - Y_\Gamma\|^2 + \lambda^2|\log(\epsilon)|\epsilon^2|\Gamma| \quad .$
- Minimization: $\Gamma_S = \{n \leq \epsilon^{-\gamma}, |\langle Y, b_n\rangle| \geq \lambda\sqrt{|\log(\epsilon)|}\epsilon\}$ (thresholding) and $F_S = Y_{\Gamma_S}$.

# Thresholding and Model Selection

- Back to the oracle estimator.
- Oracle risk: $\Gamma_0 = \mathrm{argmin}_\Gamma \|f - f_\Gamma\|^2 + \epsilon^2|\Gamma| \quad .$
- Empirical risk: $\Gamma_S = \mathrm{argmin}_{\Gamma \subset [1,\epsilon^{-\gamma}]} \|Y - Y_\Gamma\|^2 + \lambda^2 |\log(\epsilon)|\epsilon^2|\Gamma| \quad .$
- Minimization: $\Gamma_S = \{n \leq \epsilon^{-\gamma}, |\langle Y, b_n\rangle| \geq \lambda\sqrt{|\log(\epsilon)|}\epsilon\}$ (thresholding) and $F_S = Y_{\Gamma_S}$.
- Model selection:

$$F_S = \underset{P_m Y,\ m \in \mathcal{M}_\epsilon}{\mathrm{argmin}} \|Y - P_m Y\|^2 + \lambda^2 |\log(\epsilon)|\epsilon^2 \dim(m) \quad .$$

with $\mathcal{M}_\epsilon$, model collection, with models $m =$ subspaces spanned by some of the $\epsilon^{-\gamma}$ first basis vector.

# Thresholding and Model Selection

- Back to the oracle estimator.
- Oracle risk: $\Gamma_0 = \operatorname{argmin}_\Gamma \|f - f_\Gamma\|^2 + \epsilon^2|\Gamma| \quad .$
- Empirical risk: $\Gamma_S = \operatorname{argmin}_{\Gamma\subset[1,\epsilon^{-\gamma}]} \|Y - Y_\Gamma\|^2 + \lambda^2|\log(\epsilon)|\epsilon^2|\Gamma| \quad .$
- Minimization: $\Gamma_S = \{n \leq \epsilon^{-\gamma}, |\langle Y, b_n\rangle| \geq \lambda\sqrt{|\log(\epsilon)|}\epsilon\}$ (thresholding) and $F_S = Y_{\Gamma_S}$.
- Model selection:

$$F_S = \operatorname*{argmin}_{P_m Y,\ m\in\mathcal{M}_\epsilon} \|Y - P_m Y\|^2 + \lambda^2|\log(\epsilon)|\epsilon^2 \dim(m) \quad .$$

  with $\mathcal{M}_\epsilon$, model collection, with models $m =$ subspaces spanned by some of the $\epsilon^{-\gamma}$ first basis vector.
- **Theorem (*Barron, Birgé, Massart*) $\simeq$ (*Donoho, Johnstone*):** For $\lambda$ large enough,

$$E(\|f - F_S\|^2) \leq C \min_{m\in\mathcal{M}_\epsilon} \|f - P_m f\|^2 + \lambda^2|\log(\epsilon)|\epsilon^2 \dim(m) \quad .$$

# Thresholding and Model Selection

- Back to the oracle estimator.
- Oracle risk: $\Gamma_0 = \mathrm{argmin}_\Gamma \|f - f_\Gamma\|^2 + \epsilon^2|\Gamma|$ .
- Empirical risk: $\Gamma_S = \mathrm{argmin}_{\Gamma \subset [1,\epsilon^{-\gamma}]} \|Y - Y_\Gamma\|^2 + \lambda^2 |\log(\epsilon)|\epsilon^2|\Gamma|$ .
- Minimization: $\Gamma_S = \{n \leq \epsilon^{-\gamma}, |\langle Y, b_n\rangle| \geq \lambda\sqrt{|\log(\epsilon)|}\epsilon\}$ (thresholding) and $F_S = Y_{\Gamma_S}$.
- Model selection:

$$F_S = \underset{P_m Y,\ m\in\mathcal{M}_\epsilon}{\mathrm{argmin}} \|Y - P_m Y\|^2 + \lambda^2|\log(\epsilon)|\epsilon^2 \dim(m) \quad .$$

  with $\mathcal{M}_\epsilon$, model collection, with models $m =$ subspaces spanned by some of the $\epsilon^{-\gamma}$ first basis vector.
- **Theorem (*Barron, Birgé, Massart*) $\simeq$ (*Donoho, Johnstone*):** For $\lambda$ large enough,

$$E(\|f - F_S\|^2) \leq C \min_{m\in\mathcal{M}_\epsilon} \|f - P_m f\|^2 + \lambda^2|\log(\epsilon)|\epsilon^2 \dim(m) \quad .$$

- Suitable approach when more than one basis is considered.

# Model Selection

# Model Selection

- Simplified setting: $\mathcal{M}_\epsilon$ model collection with each model $m$ spanned by some vectors chosen amongst $\kappa$ different vectors.

# Model Selection

- Simplified setting: $\mathcal{M}_\epsilon$ model collection with each model $m$ spanned by some vectors chosen amongst $\kappa$ different vectors.
- **Theorem (*Barron, Birgé, Massart*):**

$$F_S = \mathop{\mathrm{argmin}}_{P_m Y,\ m \in \mathcal{M}_\epsilon} \|Y - P_m Y\|^2 + \lambda^2 \log(\kappa)\epsilon^2 \dim(m)$$

satisfies

$$E(\|f - F_S\|^2) \leq C \min_{m \in \mathcal{M}_\epsilon} \|f - P_m f\|^2 + \lambda^2 \log(\kappa)\epsilon^2 \dim(m) \quad .$$

# Model Selection

- Simplified setting: $\mathcal{M}_\epsilon$ model collection with each model $m$ spanned by some vectors chosen amongst $\kappa$ different vectors.
- **Theorem (*Barron, Birgé, Massart*):**

$$F_S = \underset{P_m Y,\ m \in \mathcal{M}_\epsilon}{\operatorname{argmin}} \|Y - P_m Y\|^2 + \lambda^2 \log(\kappa) \epsilon^2 \dim(m)$$

satisfies

$$E(\|f - F_S\|^2) \leq C \min_{m \in \mathcal{M}_\epsilon} \|f - P_m f\|^2 + \lambda^2 \log(\kappa) \epsilon^2 \dim(m) \quad .$$

- **Theorem (Maxiset):**

$E(\|f - F_S\|^2) \leq C \left(\log(\kappa)\epsilon^2\right)^{\frac{\beta}{\beta+1}}$

$$\Leftrightarrow \min_{m \in \mathcal{M}_\epsilon} \|f - P_m f\|^2 + \lambda^2 \log(\kappa) \epsilon^2 \dim(m) \leq C \left(\log(\kappa)\epsilon^2\right)^{\frac{\beta}{\beta+1}}$$

$$\Leftrightarrow \min_{m \in \mathcal{M}_\epsilon} \|f - P_m f\|^2 + T^2 \dim(m) \leq C \left(T^2\right)^{\frac{\beta}{\beta+1}}$$

$\Leftrightarrow \|f - f_M\|^2 \leq C M^{-\beta}$ plus linear approximation property

$$\Leftrightarrow f \in \mathcal{A}^{\beta}{}_{\kappa} \quad .$$

# Geometric Estimation with Bandlets

# Geometric Estimation with Bandlets

- $\mathcal{M}_\epsilon$: collection of subspaces spanned by some of the first $\epsilon^{-\gamma}$ basis vector of a suitable collection of bandelet bases.

# Geometric Estimation with Bandlets

- $\mathcal{M}_\epsilon$: collection of subspaces spanned by some of the first $\epsilon^{-\gamma}$ basis vector of a suitable collection of bandelet bases.
- Polynomial bound on the number $\kappa$ of bandlets in all the considered bandlet basis: $\log \kappa = \gamma |\log \epsilon|$.

# Geometric Estimation with Bandlets

- $\mathcal{M}_\epsilon$: collection of subspaces spanned by some of the first $\epsilon^{-\gamma}$ basis vector of a suitable collection of bandelet bases.
- Polynomial bound on the number $\kappa$ of bandlets in all the considered bandlet basis: $\log \kappa = \gamma |\log \epsilon|$.
- Bandlet model selection:

$$F_S = \operatorname*{argmin}_{m \in \mathcal{M}} \|Y - P_m Y\|^2 + \lambda^2 |\log(\epsilon)| \epsilon^2 \mathsf{dim}(m)$$

# Geometric Estimation with Bandlets

- $\mathcal{M}_\epsilon$: collection of subspaces spanned by some of the first $\epsilon^{-\gamma}$ basis vector of a suitable collection of bandelet bases.
- Polynomial bound on the number $\kappa$ of bandlets in all the considered bandlet basis: $\log \kappa = \gamma |\log \epsilon|$.
- Bandlet model selection:

$$F_S = \operatorname*{argmin}_{m \in \mathcal{M}} \|Y - P_m Y\|^2 + \lambda^2 |\log(\epsilon)| \epsilon^2 \mathsf{dim}(m)$$

- Same optimization algorithm than for approximation.

# Geometric Estimation with Bandlets

- $\mathcal{M}_\epsilon$: collection of subspaces spanned by some of the first $\epsilon^{-\gamma}$ basis vector of a suitable collection of bandelet bases.
- Polynomial bound on the number $\kappa$ of bandlets in all the considered bandlet basis: $\log \kappa = \gamma |\log \epsilon|$.
- Bandlet model selection:

$$F_S = \underset{m \in \mathcal{M}}{\operatorname{argmin}} \|Y - P_m Y\|^2 + \lambda^2 |\log(\epsilon)| \epsilon^2 \mathrm{dim}(m)$$

- Same optimization algorithm than for approximation.
- **Theorem (Almost minimax optimality)**: if $f \in \mathbf{C}^\alpha - \mathbf{C}^\alpha$ the

$$E(\|f - F_S\|^2) \leq C \left(|\log(\epsilon)|\right)^{\frac{\alpha}{\alpha+1}} \quad .$$

# Geometric Estimation with Bandlets

- $\mathcal{M}_\epsilon$: collection of subspaces spanned by some of the first $\epsilon^{-\gamma}$ basis vector of a suitable collection of bandelet bases.
- Polynomial bound on the number $\kappa$ of bandlets in all the considered bandlet basis: $\log \kappa = \gamma |\log \epsilon|$.
- Bandlet model selection:

$$F_S = \underset{m \in \mathcal{M}}{\operatorname{argmin}} \|Y - P_m Y\|^2 + \lambda^2 |\log(\epsilon)| \epsilon^2 \mathrm{dim}(m)$$

- Same optimization algorithm than for approximation.
- **Theorem (Almost minimax optimality)**: if $f \in \mathbf{C}^\alpha - \mathbf{C}^\alpha$ the

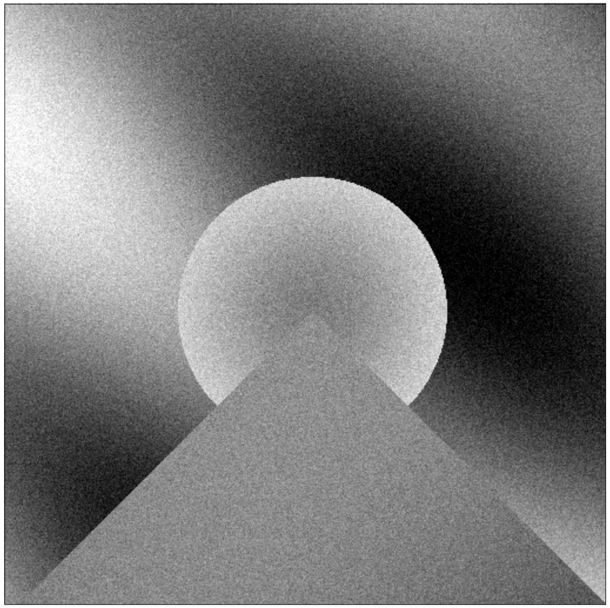
$$E(\|f - F_S\|^2) \leq C \left(|\log(\epsilon)|\right)^{\frac{\alpha}{\alpha+1}} \quad .$$

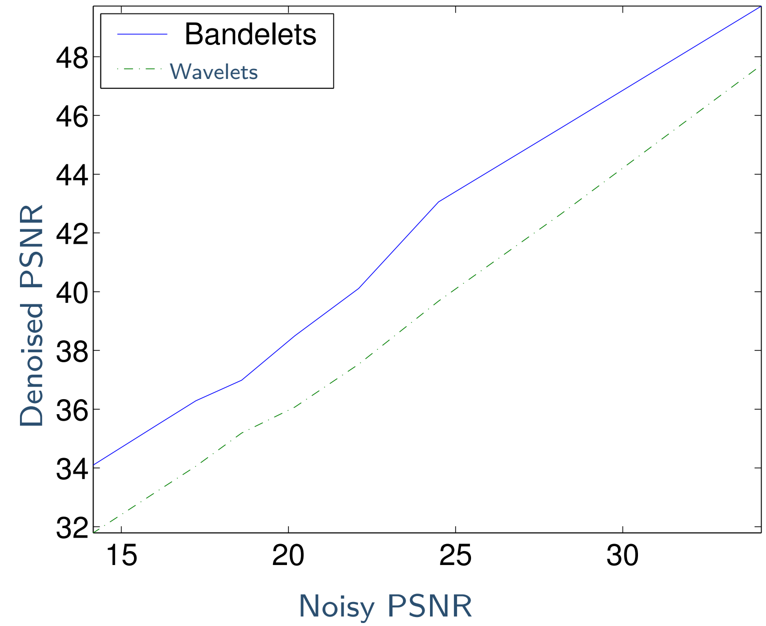
- **Theorem (Maxiset)**:

$$E(\|f - F_S\|^2) \leq C \left(|\log(\epsilon)|\epsilon^2\right)^{\frac{\alpha}{\alpha+1}} \Leftrightarrow f \in \mathcal{A}^{\alpha}{}_{\gamma}$$
$$\Leftrightarrow \forall M, \exists \mathcal{B}, \|f - f_M\|^2 \leq C M^{-\alpha}$$

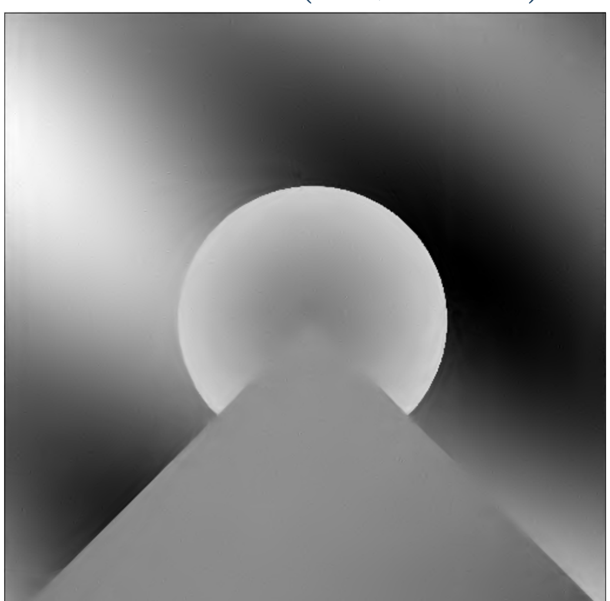
plus linear approximation property .

Noisy (20,19 dB)

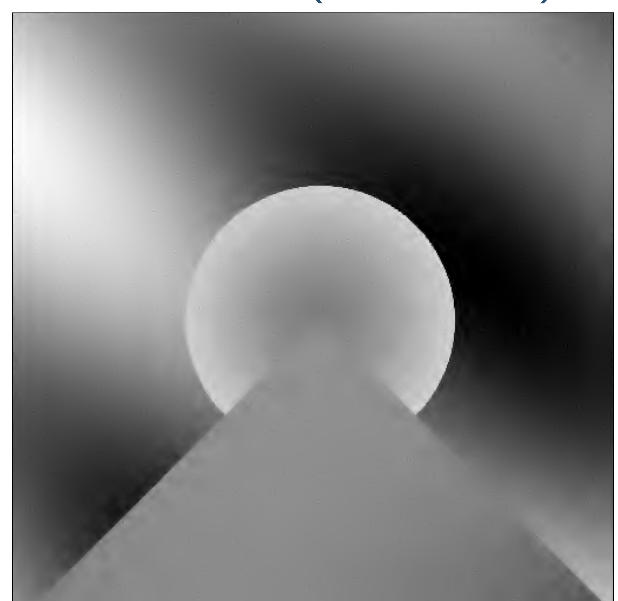




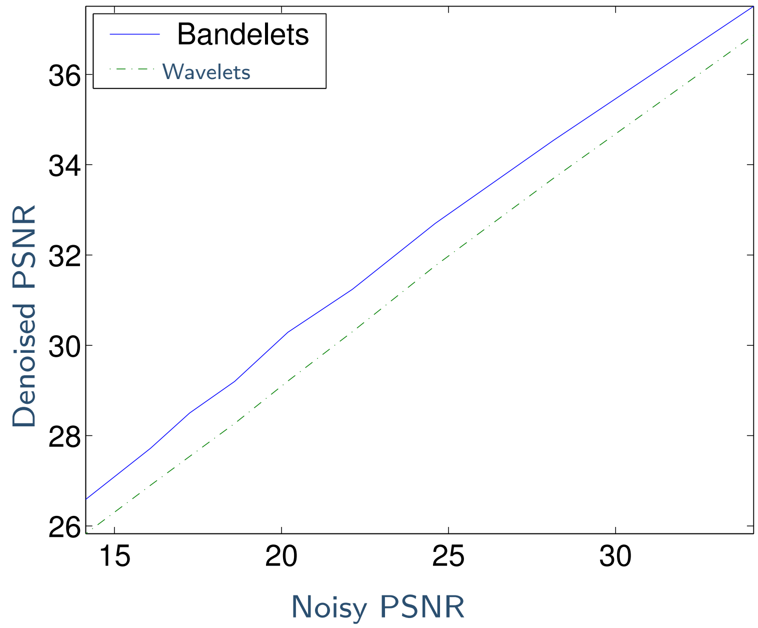

Bandlets (30,29 dB)



Wavelets (28,21 dB)

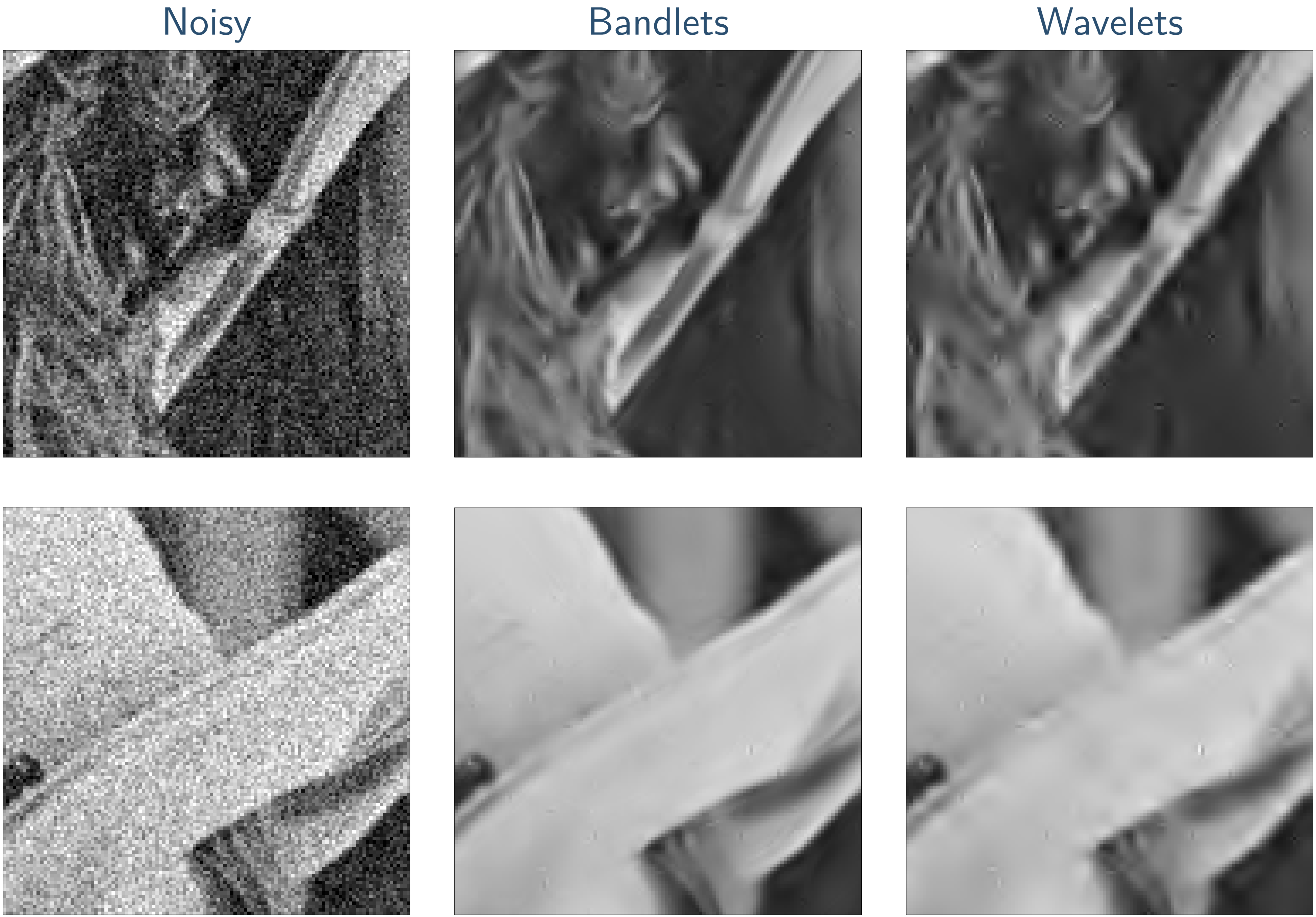


Noisy $(20{,}19\,\mathrm{dB})$

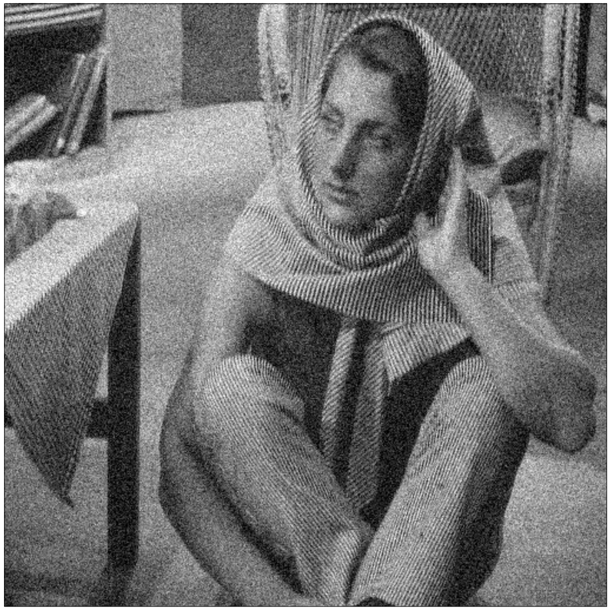


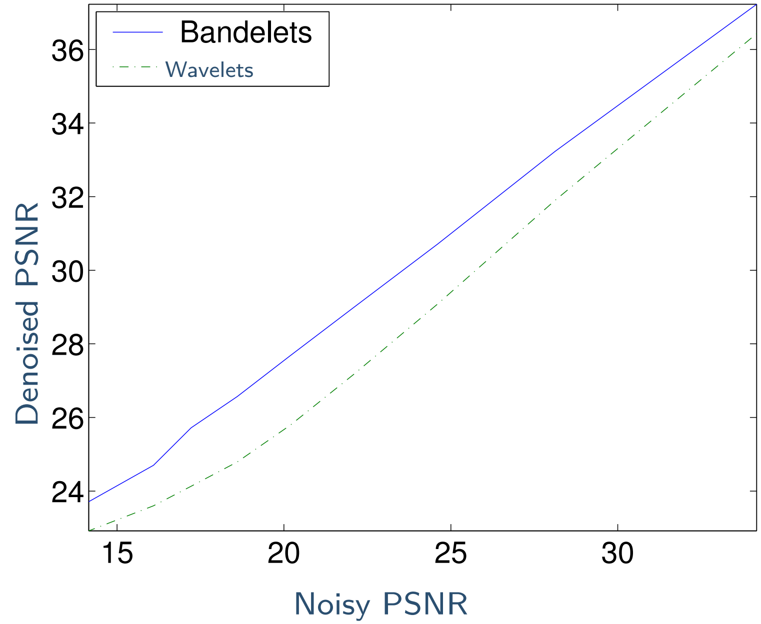

Bandlets $(30{,}29\,\mathrm{dB})$

Wavelets $(28{,}21\,\mathrm{dB})$

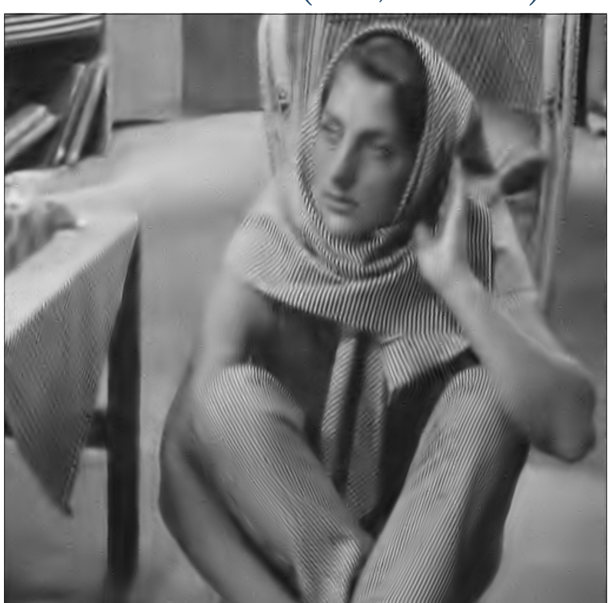

Noisy
Bandlets
Wavelets

Noisy (20,19 dB)

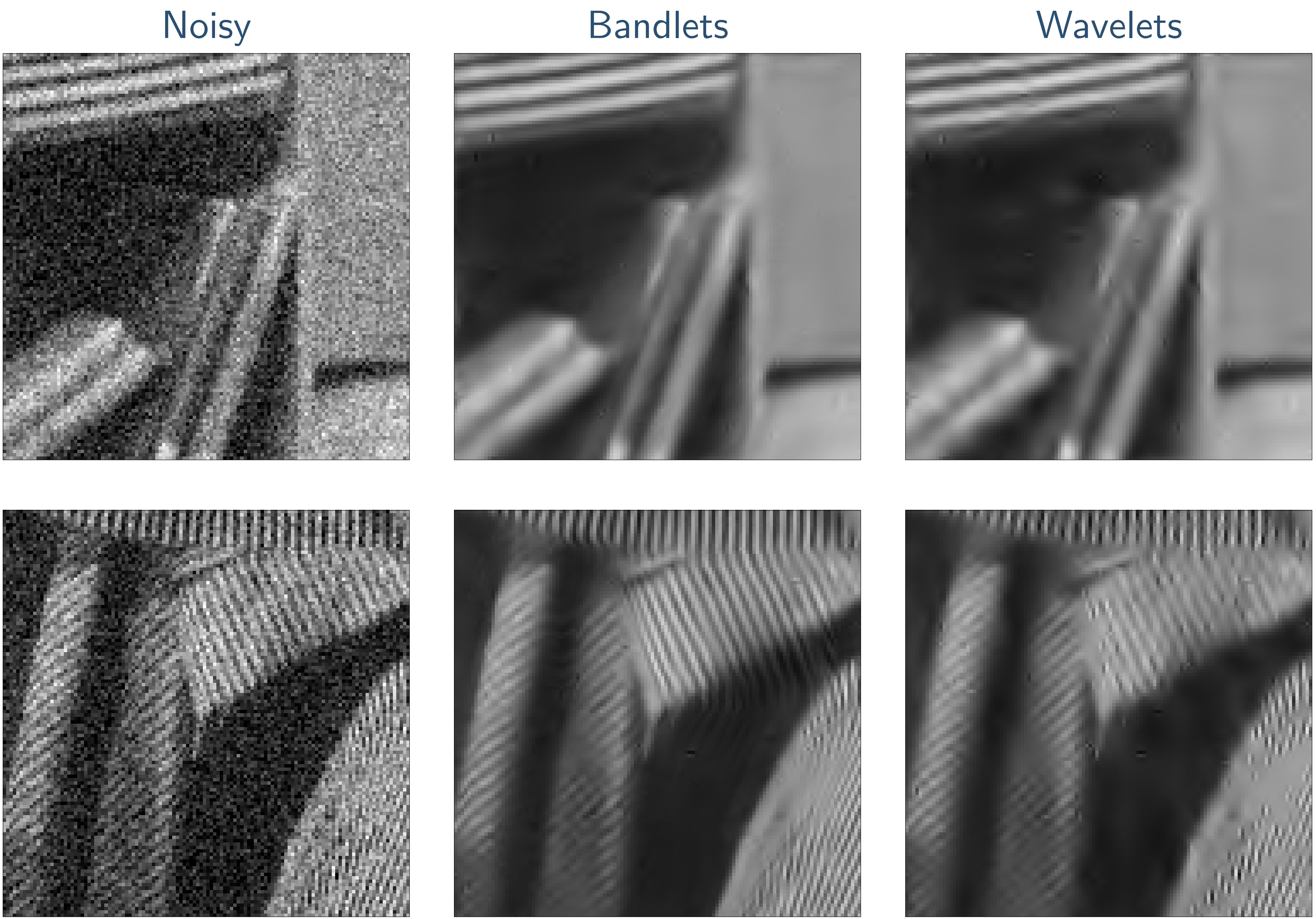





Bandlets (27,68 dB)



Wavelets (25,79 dB)


Noisy
Bandlets
Wavelets

# Conclusion

- Bandelets are suitable for good geometrical image representation.

# Conclusion

- Bandelets are suitable for good geometrical image representation.
- Bandelets are suitable for good geometrical image estimation.

# Conclusion

- Bandelets are suitable for good geometrical image representation.
- Bandelets are suitable for good geometrical image estimation.
- Strong connection between these two properties.

# Conclusion

- Bandelets are suitable for good geometrical image representation.
- Bandelets are suitable for good geometrical image estimation.
- Strong connection between these two properties.
- Very strong link between model selection estimation convergence rate and approximation rate.

# Conclusion

- Bandelets are suitable for good geometrical image representation.
- Bandelets are suitable for good geometrical image estimation.
- Strong connection between these two properties.
- Very strong link between model selection estimation convergence rate and approximation rate.
- What are the images well estimated by the bandelet estimator?

# Conclusion

- Bandelets are suitable for good geometrical image representation.
- Bandelets are suitable for good geometrical image estimation.
- Strong connection between these two properties.
- Very strong link between model selection estimation convergence rate and approximation rate.
- What are the images well estimated by the bandelet estimator?
  - Maxiset answer: the images well approximated by the bandelets!.

# Conclusion

- Bandelets are suitable for good geometrical image representation.
- Bandelets are suitable for good geometrical image estimation.
- Strong connection between these two properties.
- Very strong link between model selection estimation convergence rate and approximation rate.
- What are the images well estimated by the bandelet estimator?
  - Maxiset answer: the images well approximated by the bandelets!.
  - Are the natural images within this set?

# Conclusion

- Bandelets are suitable for good geometrical image representation.
- Bandelets are suitable for good geometrical image estimation.
- Strong connection between these two properties.
- Very strong link between model selection estimation convergence rate and approximation rate.
- What are the images well estimated by the bandelet estimator?
  - Maxiset answer: the images well approximated by the bandelets!.
  - Are the natural images within this set? Probably not (Texture?)

# Conclusion

- Bandelets are suitable for good geometrical image representation.
- Bandelets are suitable for good geometrical image estimation.
- Strong connection between these two properties.
- Very strong link between model selection estimation convergence rate and approximation rate.
- What are the images well estimated by the bandelet estimator?
  - Maxiset answer: the images well approximated by the bandelets!.
  - Are the natural images within this set? Probably not (Texture?)
- Open issues:
  - Functional characterization of the bandlet approximation spaces.

# Conclusion

- Bandelets are suitable for good geometrical image representation.
- Bandelets are suitable for good geometrical image estimation.
- Strong connection between these two properties.
- Very strong link between model selection estimation convergence rate and approximation rate.
- What are the images well estimated by the bandelet estimator?
  - Maxiset answer: the images well approximated by the bandelets!.
  - Are the natural images within this set? Probably not (Texture?)
- Open issues:
  - Functional characterization of the bandlet approximation spaces.
  - What about other penalizations?

# Conclusion

- Bandelets are suitable for good geometrical image representation.
- Bandelets are suitable for good geometrical image estimation.
- Strong connection between these two properties.
- Very strong link between model selection estimation convergence rate and approximation rate.
- What are the images well estimated by the bandelet estimator?
  - Maxiset answer: the images well approximated by the bandelets!.
  - Are the natural images within this set? Probably not (Texture?)
- Open issues:
  - Functional characterization of the bandlet approximation spaces.
  - What about other penalizations?
  - Is the minimax approach too pessimistic?

# Conclusion

- Bandelets are suitable for good geometrical image representation.
- Bandelets are suitable for good geometrical image estimation.
- Strong connection between these two properties.
- Very strong link between model selection estimation convergence rate and approximation rate.
- What are the images well estimated by the bandelet estimator?
  - Maxiset answer: the images well approximated by the bandelets!.
  - Are the natural images within this set? Probably not (Texture?)
- Open issues:
  - Functional characterization of the bandlet approximation spaces.
  - What about other penalizations?
  - Is the minimax approach too pessimistic? Probably not (work in progress...)

# Conclusion

- Bandelets are suitable for good geometrical image representation.
- Bandelets are suitable for good geometrical image estimation.
- Strong connection between these two properties.
- Very strong link between model selection estimation convergence rate and approximation rate.
- What are the images well estimated by the bandelet estimator?
  - Maxiset answer: the images well approximated by the bandelets!.
  - Are the natural images within this set? Probably not (Texture?)
- Open issues:
  - Functional characterization of the bandlet approximation spaces.
  - What about other penalizations?
  - Is the minimax approach too pessimistic? Probably not (work in progress...)

# Conclusion

- Bandelets are suitable for good geometrical image representation.
- Bandelets are suitable for good geometrical image estimation.
- Strong connection between these two properties.
- Very strong link between model selection estimation convergence rate and approximation rate.
- What are the images well estimated by the bandelet estimator?
  - Maxiset answer: the images well approximated by the bandelets!.
  - Are the natural images within this set? Probably not (Texture?)
- Open issues:
  - Functional characterization of the bandlet approximation spaces.
  - What about other penalizations?
  - Is the minimax approach too pessimistic? Probably not (work in progress...)

# Sketch of proof

- Model selection and Maxiset Theorems sketch of proof.
- $\epsilon^2 = \frac{1}{N}$
- Players:
  - Estimate: $F_S = \operatorname{argmin}_{m \in \mathcal{M}} \|Y - P_m Y\|^2 + \lambda \frac{\log N}{N} \dim(m)$
  - Model Selection Oracle:
    $f_O = \operatorname{argmin}_{m \in \mathcal{M}} \|f - P_m f\|^2 + \lambda \frac{\log N}{N} \dim(m)$
  - Maxiset Oracle: $f_{O'} = \operatorname{argmin}_{m \in \mathcal{M}} \|Y - P_m Y\|^2 + \lambda \frac{\log N}{4N} \dim(m)$
- Model selection:

$$E(\|f - F_S\|^2 + \lambda \frac{\log N}{N} \dim(m_F)) \leq C \left( \|f - f_O\|^2 + \lambda \frac{\log N}{N} \dim(m_O) \right) ??$$

- Maxiset:

$$\|f - f_{O'}\|^2 + \lambda \frac{\log N}{N} \dim(m_{O'}) \leq C E(\|f - F_S\|^2) \quad ??$$

- Model selection: general case but in probability.
- Maxiset: simple case (embedded models).

# Model Selection – 1

- Players:
  - Estimate: $F_S = \mathrm{argmin}_{m \in \mathcal{M}} \|Y - P_m Y\|^2 + \lambda \frac{\log N}{N} \dim(m)$
  - Model Selection Oracle:
    $f_O = \mathrm{argmin}_{m \in \mathcal{M}} \|f - P_m f\|^2 + \lambda \frac{\log N}{N} \dim(m)$
- Model selection: With large probability

$$\|f - F_S\|^2 + \lambda \frac{\log N}{N} \dim(m_F) \leq C \left( \|f - f_O\|^2 + \lambda \frac{\log N}{N} \dim(m_O) \right)$$

- Uniform noise control over all models $m$: Gaussian concentration inequality

$$P\left( \forall m, \quad \|P_m W\| \leq \sqrt{12 \log N \dim(m)} \right) \geq 1 - \frac{1}{N}$$

# Model Selection – 2

- By definition:
$$\|Y - F_S\|^2 + \lambda \frac{\log N}{N} \dim(m_F) \leq \|Y - f_O\|^2 + \lambda \frac{\log N}{N} \dim(m_O)$$
and using $\|Y-g\|^2 = \|Y-f\|^2 + 2\langle Y-f, f-g\rangle + \|f-g\|^2$
$$\|f - F_S\|^2 + \lambda \frac{\log N}{N} \dim(m_F) \leq \|f - f_O\|^2 + \lambda \frac{\log N}{N} \dim(m_O) + \frac{2}{\sqrt{N}}\langle W, f_O - F_S\rangle$$
- $|\langle W, f_O - F_S\rangle| \leq \|P_{m_O \cup m_F} W\| \, \|f_P - F_S\|$ .
- $\|f_O - F_S\| \leq \|f_O - f\| + \|f - F_S\| \leq 2(\|f - F_S\|^2 + \lambda \frac{\log N}{N} \dim(m_F))^{1/2}$ .
- Concentration inequality:
$$P\left(\forall m, \quad \|P_m W\| \leq \sqrt{12 \log N \dim(m)}\right) \geq 1 - \frac{1}{N}$$
- Thus with $P \geq 1 - \frac{1}{N}$,
$$\|P_{m_O \cup m_F} W\| \leq \sqrt{12 \log N (\dim m_O + \dim(m_F))}$$
$$|\langle W, f_O - F_S\rangle| \leq \sqrt{48/\lambda^2}(\|f - F_S\|^2 + \lambda \frac{\log N}{N} \dim(m_F)) \quad .$$

# Maxiset – 1

- Maxiset:

$$\|f - f_{O'}\|^2 + \lambda \frac{\log N}{N} \dim(m_{O'}) \leq CE(\|f - F_S\|^2)???$$

- No but

$$E(\|f - F_S\|^2) \leq C \left( \frac{\log N)}{N} \right)^{\frac{\beta}{\beta+1}}$$

$$\implies \min_{m \in \mathcal{M}_N} \|f - P_m f\|^2 + \log N \frac{\dim(m)}{N} \leq C \left( \frac{\log N}{N} \right)^{\frac{\beta}{\beta+1}}$$

$$\implies f \in \mathcal{A}^{\beta}$$

- Simple proof when the models are embedded ($m_1 \subset m_2$ or $m_1 \supset m_2$ for all $m_1, m_2$).
- General case much more complex...
- Thresholding is a simple extension of the embedded model case.

# Maxiset – 2

- We are going to prove that $\|f - F_S\|^2 \geq \|f - f_O\|^2$ !
- If $m_F \subset m_O$, $\|f - F_S\|^2 \geq \|f - P_{m_F} f\|^2 \geq \|f - f_O\|^2$.
- Otherwise $m_F \supset m_O$ and $\|f - F_S\|^2 = \|f - P_{m_F} f|^2 + \|P_{m_F} f - F_S\|^2 = \|f - P_{m_O} f|^2 + \|P_{m_O} W|^2 + \|P_{m_F \setminus m_O} W\|^2 - \|P_{m_F \setminus m_O} f\|^2$ .
- Recall that

$$\|Y - P_{m_F} Y\|^2 + \lambda \frac{\log N}{N} \dim(m_F) \leq \|Y - P_{m_O} Y\|^2 + \lambda \frac{\log N}{N} \dim(m_O)$$

$$\|f - P_{m_O} f\|^2 + \lambda \frac{\log N}{4N} \dim(m_O) \leq \|f - P_{m_F} f\|^2 + \lambda \frac{\log N}{4N} \dim(m_F)$$

- Thus

$$\|P_{m_F \setminus m_O} f\|^2 \leq \frac{1}{4} \|P_{m_F \setminus m_O} Y\|^2$$

$$\|P_{m_F \setminus m_O} f\|^2 \leq \frac{1}{2} \left( \|P_{m_F \setminus m_O} W\|^2 + \|P_{m_F \setminus m_O} f\|^2 \right)$$

which leads to

$$\|P_{m_F \setminus m_O} f\|^2 \leq \|P_{m_F \setminus m_O} W\|^2$$

# Maxiset – 3

- We prove that $\|f - f_O\|^2 \leq \|f - F_S\|^2$ and thus that

$$\|f - f_O\|^2 \leq E(\|f - F_S\|^2) \leq C \left( \frac{\lambda_N}{N} \right)^{\frac{\beta}{\beta+1}}$$

- We should now work on the dependency on $N$ to control the number of term and prove that $\|f - f_O\|^2 + \lambda \frac{\log N}{N} \dim(m_O) \leq C' \left( \frac{\lambda_N}{N} \right)^{\frac{\beta}{\beta+1}}$
- Using that $m_{O(N/2)}$ is not as efficient as $m_{O(N)}$ so that

$$\begin{aligned}\|f - f_{O(N)}\|^2 + \lambda \frac{\log N}{N} \dim(m_{O(N)}) &\leq \|f - f_{O(N/2)}\|^2 + \lambda \frac{\log N}{N} \dim(m_{O(N/2)}) \\ &\leq \frac{1}{2} \|f - f_{O(N/2)}\|^2 \\ &\quad + \frac{1}{2} \left( \|f - f_{O(N/2)}\|^2 + \lambda \frac{\log N}{N/2} \dim(m_{O(N/2)}) \right)\end{aligned}$$

- Slightly more tricky because of the $\log N$...
- Conclusion obtained by recursion.