Aisenstadt Chair Course
CRM September 2009

Part II
Redundant Dictionaries and Pursuit Algorithms

Stéphane Mallat

Centre de Mathématiques Appliquées
Ecole Polytechnique

Sparsity in Redundant Dictionaries

- Bases are minimum set to decompose signals.
- Natural languages use redundant dictionaries.
- Use of larger dictionaries incorporating more patterns to represent complex signals $f \in \mathbf{R}^N$

$$\mathcal{D} = \{\phi_p\}_{p\in\Gamma} \text{ with } \|\phi_p\| = 1 \text{ and } |\Gamma| = P > N.$$

- How to construct sparse representations in D ?
- What is the impact of redundancy on the mathematics and applications ?

Dictionary Approximation

- Let $\mathcal{D} = \{\phi_p\}_{p\in\Gamma}$ be a redundant dictionary of P vectors.

- The best approximation of f from a sub-family $\{\phi_p\}_{p\in\Lambda}$ is its orthogonal projection in $\mathbf{V}_\Lambda = \mathrm{Vect}\{\phi_p\}_{p\in\Lambda}$

$$f_\Lambda = \sum_{p\in\Lambda} a[p]\,\phi_p$$

- Stability: $\{\phi_p\}_{p\in\Lambda}$ must be a Riez basis of $\mathbf{V}_\Lambda$:

$$\text{there exists } \; 0 < A_\Lambda \leq B_\Lambda \; \text{ such that}$$

$$\forall a[p] \in \mathbf{R}^\Lambda \quad , \quad A_\Lambda \sum_{p\in\Lambda} |a[p]|^2 \leq \| \sum_{p\in\Lambda} a[p]\,\phi_p\|^2 \leq B_\Lambda \sum_{p\in\Lambda} |a[p]|^2 \; .$$

Best M-Term Approximation

- The best M-term approximation support Λ

$$\text{minimizes} \quad \|f - f_\Lambda\| \quad \text{with} \quad |\Lambda| = M \ .$$

- A best M-term approximation minimizes a Lagrangian:

$$\mathcal{L}_0 = \|f - f_\Lambda\|^2 + T^2 \, |\Lambda| \ .$$

- If $\mathcal{D}$ is an orthonormal basis then $\Lambda = \{p \ : \ |\langle f, \phi_p \rangle| \geq T\}$

- In general, finding Λ is an NP-hard problem.

Compression Applications

- Compute a best M-term approximation

$$f_\Lambda = \sum_{p \in \Lambda} a[p]\, \phi_p \quad \text{with} \quad |\Lambda| = M \;.$$

- Compression with uniform quantization

$$\tilde{f}_\Lambda = \sum_{p \in \Lambda} Q(a[p])\, \phi_p$$

- Total bit budget:

$$R = \log_2 \binom{P}{M} + \mu\, M$$

$$R \sim M \log_2(P/M)$$

- Increasing P reduces $D = \|f - \tilde{f}_\Lambda\|$ but increases R.

If $\|f - f_\Lambda\|^2 = O(M^{-\alpha})$ then $D(R) = R^{-\alpha}\, |\log(P/R)|^{-\alpha}$.

Wavelets for Cartoon Images

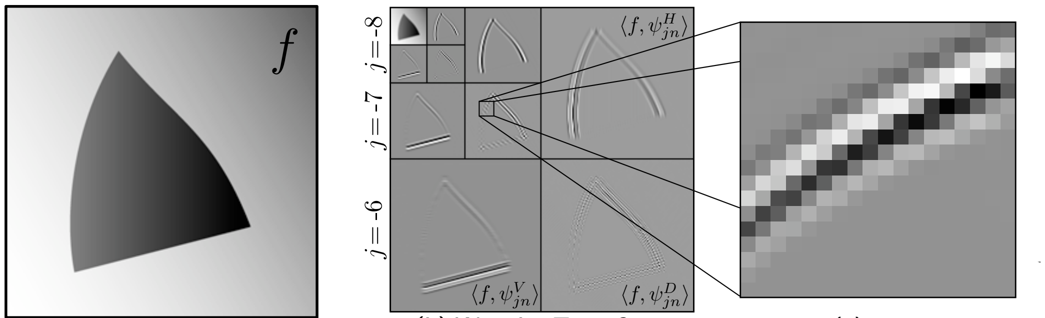
- **Theorem:** If f is uniformly $\mathbf{C}^\alpha$ then an M-term wavelet approximation gives

$$\|f - f_M\|^2 = O(M^{-\alpha})$$

- **Theorem:** If f is piecewise $\mathbf{C}^\alpha$ with finite length contours then

$$\|f - f_M\|^2 = O(M^{-1}) \quad \text{so} \quad D(R) = R^{-1}\,|\log(N/R)|$$

- Result valid for all bounded variation functions.

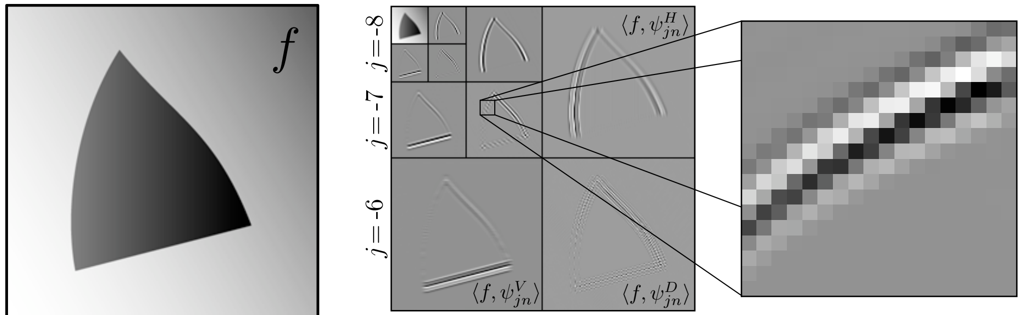


(b) Wavelet Transform

(c) Zoom

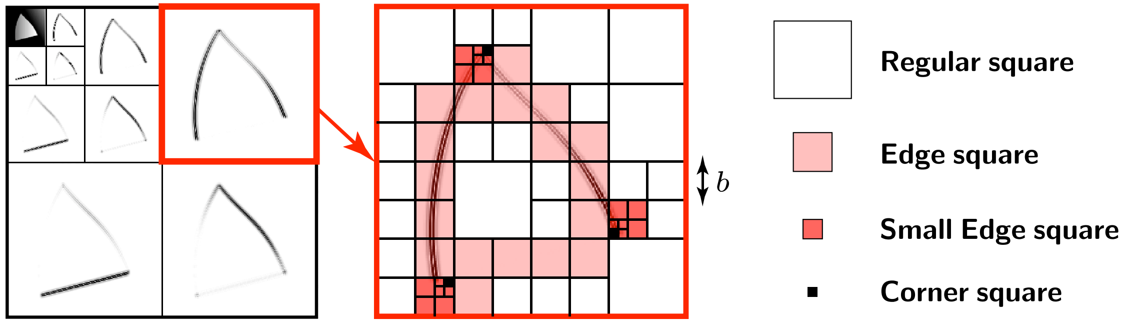
Bandlets for Geometric Regularity

Wavelet coefficients inherit the geometric regularity



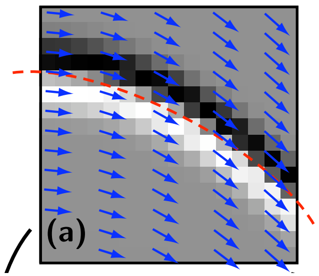
(b) Wavelet Transform (c) Zoom

1. Segmentation of wavelet coefficients.

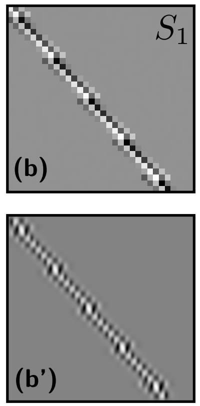


Wavelet Transform Quadtree

2. Geometric flow in edge squares along the direction of regularity.



3. 1D wavelet transform along the flow = bandlet transform



Bandlet Approximations

- A bandlet dictionary is a union of orthonormal bases.

- The best bandlet approximation (best geometry) which minimizes

$$\mathcal{L}_0 = \|f - f_M\|^2 + T^2\, M$$

can be computed with $O(N \log N)$ operations.

- **Theorem** If f is piecewise $\mathbf{C}^\alpha$ with piecewise $\mathbf{C}^\alpha$ contours

$$\|f - f_M\|^2 = O(M^{-\alpha}) \;\; \text{so} \;\; D(R) = R^{-\alpha}\, |\log(P/R)|^\alpha$$

Application of Photo ID Compression

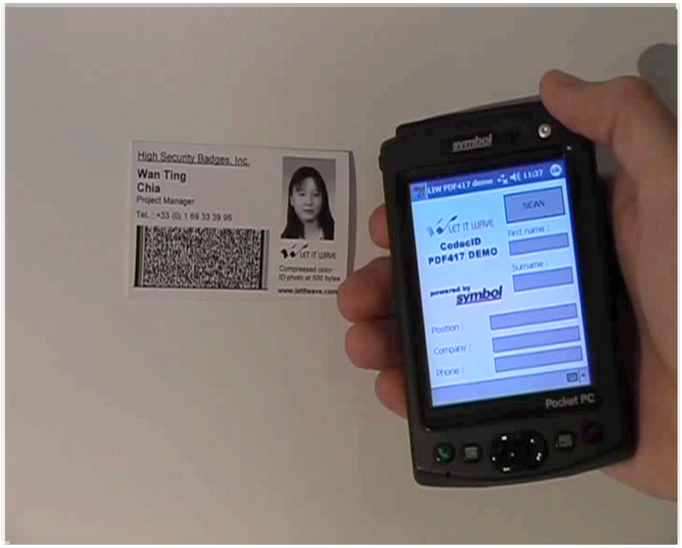
> **Secure ID cards** by storing digital photo on low memory

500 bytes

JPEG-2000
Wavelets

500 bytes

LET IT WAVE
Bandlets



Secure identification
Example : 2D barcodes

Discovery of Marketing

> **Storage of a digital ID photo on low memory chips or 2D barcodes**

- National markets : ID cards, visas, driving licenses, social insurance cards
- Transport : passes
- Corporate : site access
- Events : trade shows, congresses, sport events

Grand European IST Prize 2005

> **Transmission over low bandwidth professional secure networks**

> **NICE to have or MUST have ?** **just nice...**

Xlets Beyond Wavelets

- Xlets take advantage of the image geometric regularity: *bandlets,* curvelets, contourlets, edglets, wedglets...
- "Failure" to improve wavelet approximations for natural images.

- Can wavelet approximations be asymptotically improved ?
  I don't think so for static natural images.

Denoising in Redundant Dictionaries

- Measure a signal plus a Gaussian white noise:

$$X[n] = f[n] + W[n] \quad \text{for} \quad 0 \le n < N \ .$$

- Orthogonal projection estimator in $\mathbf{V}_\Lambda = \mathrm{Vect}\{\phi_p\}_{p\in\Lambda}$ selected in a dictionary:

$$X_\Lambda = \sum_{p\in\Lambda} a[p]\,\phi_p \ .$$

- Risk: $E\{\|f - X_\Lambda\|^2\} = \|f - f_\Lambda\|^2 + E\{\|W_\Lambda\|^2\}$.
- If Λ is fixed then $E\{\|W_\Lambda\|^2\} = \sigma^2\,|\Lambda|$.
- Oracle choice: find Λ which minimizes

$$E\{\|f - X_\Lambda\|^2\} = \|f - f_\Lambda\|^2 + \sigma^2\,|\Lambda|$$

Penalized Estimation

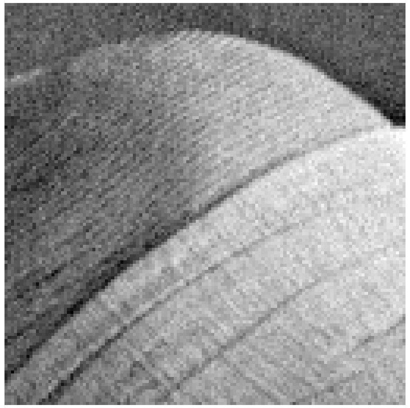
- Choosing Λ is a model selection problem.

- **Theorem:** For $T > \sigma\sqrt{2\log_e P}$

$$\tilde{\Lambda} = \arg\min_{\Lambda\subset\Gamma}\left(\|X - X_\Lambda\|^2 + T^2\,|\Lambda|\right)$$

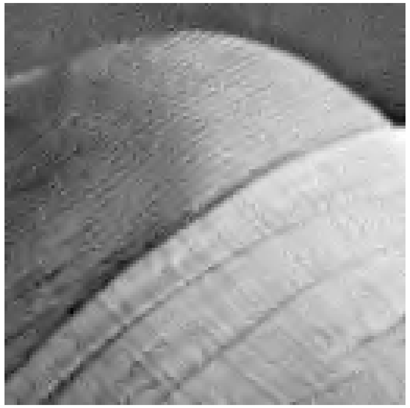
satisfies $$E\{\|f - X_{\tilde{\Lambda}}\|^2\} = 4\min_{\Lambda\subset\Gamma}\left(\|f - f_\Lambda\|^2 + T^2\,|\Lambda| + \sigma^2\right)$$

Noisy image



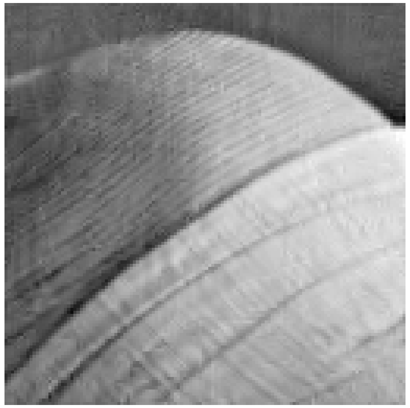
PSNR = 22db

Wavelet thresholding



PSNR = 25.3db

Bandlet thresholding



PSNR = 26.4db

Greedy Matching Pursuits

- Let $\mathcal{D} = \{\phi_p\}_{p\in\Gamma}$ be a dictionary of P > N vectors.
- A best M-term approximation minimizes the Lagrangian:

$$\mathcal{L}_0 = \|f - f_\Lambda\|^2 + T^2\,|\Lambda|$$

- Finding Λ is in general an NP complete problem.

- **Greedy choice** of the approximation vectors.
- The approximation of f over $\phi_{p_0} \in \mathcal{D}$ yields

$$f = \langle f, \phi_{p_0}\rangle + Rf$$

$$\|f\|^2 = |\langle f, \phi_{p_0}\rangle|^2 + \|Rf\|^2$$

- To minimize the residual error we choose

$$\phi_{p_0} = \arg\max_{\phi_p \in \mathcal{D}} |\langle f, \phi_p\rangle|$$

Matching Pursuit Iterations

- Initialize $R^0 f = f$
- For each $m > 0$

$$\phi_{p_m} = \arg \max_{\phi_p \in \mathcal{D}} |\langle R^m f, \phi_p \rangle|$$

$$R^m f = \langle R^m f, \phi_{p_m} \rangle + R^{m+1} f$$

$$\|R^m f\|^2 = |\langle R^m f, \phi_{p_m} \rangle|^2 + \|R^{m+1} f\|^2$$

- It results:

$$f = \sum_{m=0}^{M-1} \langle R^m f, \phi_{p_m} \rangle + R^M f$$

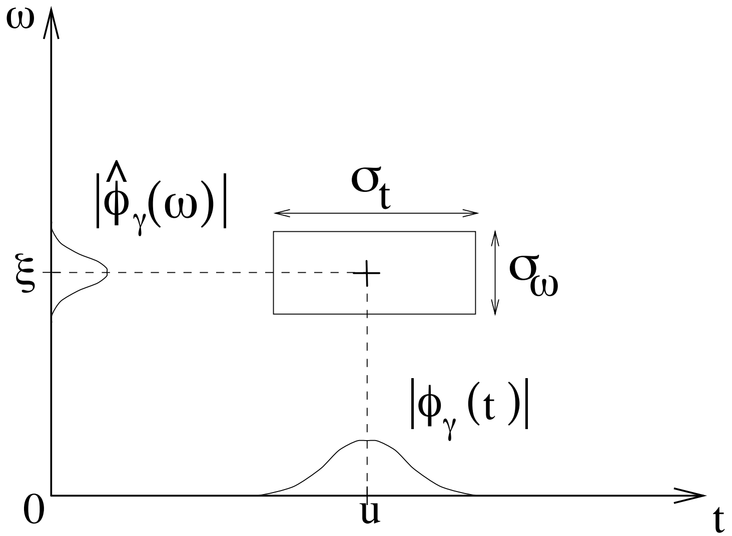
$$\|f\|^2 = \sum_{m=0}^{M-1} |\langle R^m f, \phi_{p_m} \rangle|^2 + \|R^M f\|^2$$

Time-Frequency Decompositions

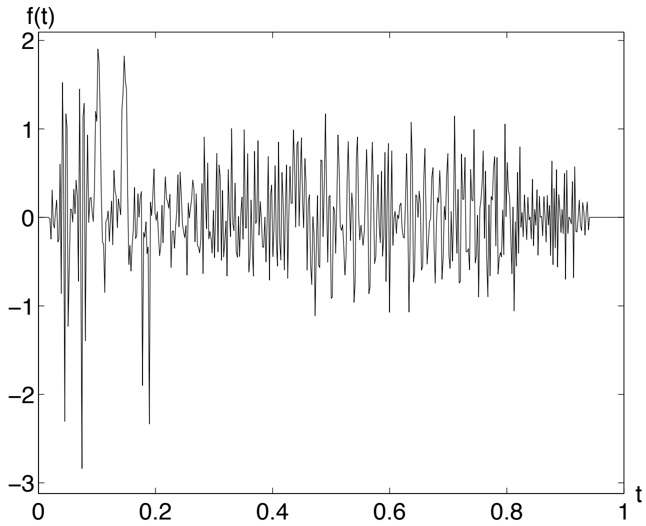
- Dictionary of time-frequency atoms:

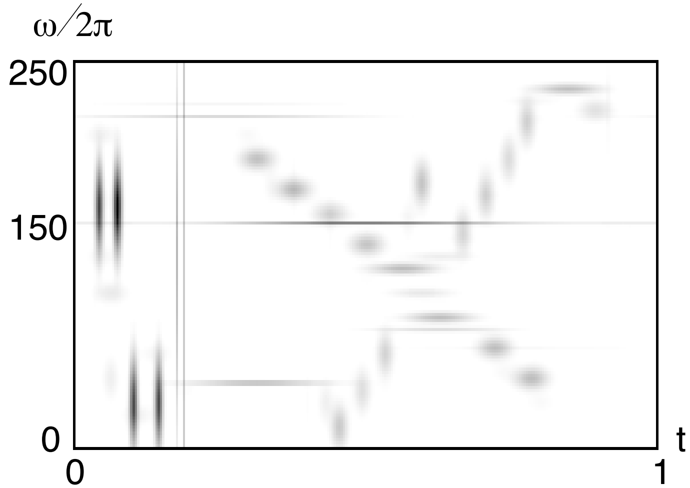
$$\mathcal{D} = \left\{ \phi_p(t) = \frac{1}{\sqrt{2^j}} w(\frac{t-u}{2^j})\, e^{i\xi t} \right\}_{(j,u,\xi)\in\Gamma}$$

- It includes Dirac and Fourier bases, wavelets and window Fourier atoms.

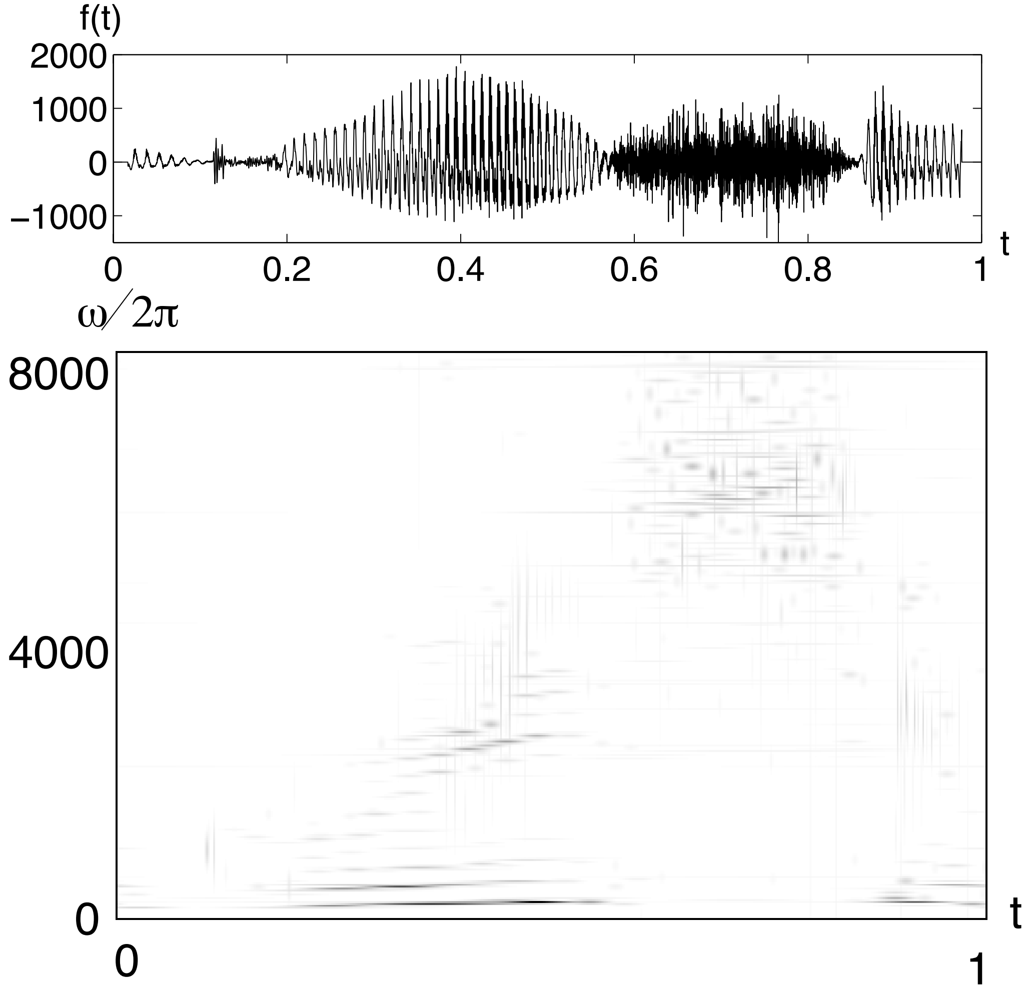


Time Frequency Matching Pursuits





Speech Signals



Applications: denoising, compression (video),
pattern recognition, but instabilities.

Orthogonal Matching Pursuit

- Initialize $R^0 f = f$
- For each $m > 0$ orthogonalize the projections:

$$\phi_{p_m} = \arg \max_{\phi_p \in \mathcal{D}} |\langle R^m f, \phi_p \rangle|$$

$$u_m = \phi_{p_m} - \sum_{l=0}^{m-1} \frac{\langle \phi_{p_m}, u_l \rangle}{\|u_l\|^2} u_l \ .$$

$$R^{m+1} f = \frac{\langle R^m f, u_m \rangle}{\|u_m\|^2} + R^{m+1} f$$

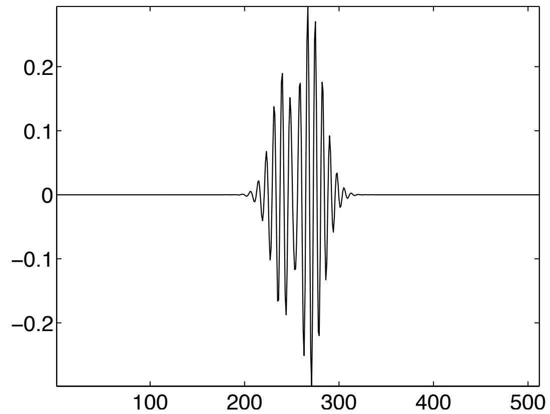
- After N iterations, it gives decomposition in an orthogonal basis:

$$f = \sum_{m=0}^{N-1} \frac{\langle R^m f, u_m \rangle}{\|u_m\|^2}.$$

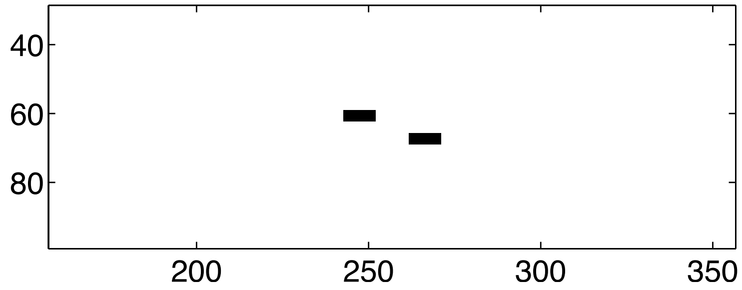
Matching versus Basis Pursuit

- Greediness is not optimal.

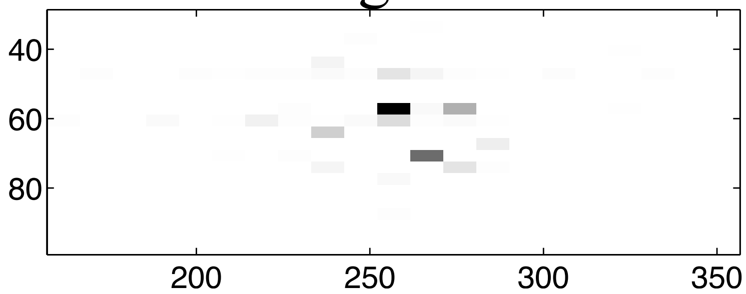
Signal f



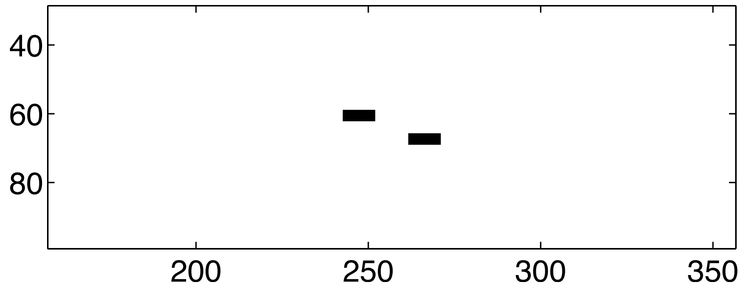
Time-Frequency Coefficients



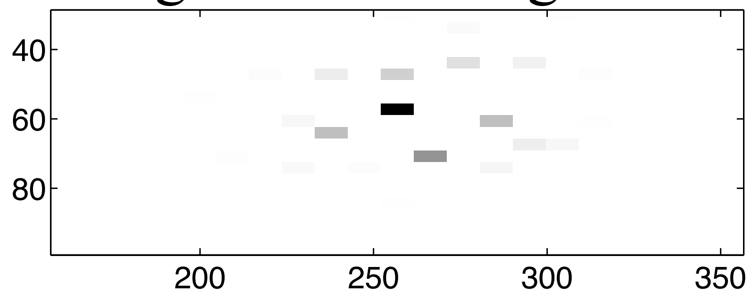
Matching Pursuit



Basis Pursuit



Orthogonal Matching Pursuit



Basis Pursuit

Find a sparse representation $f = \sum_{p \in \Gamma} a[p]\, \phi_p$

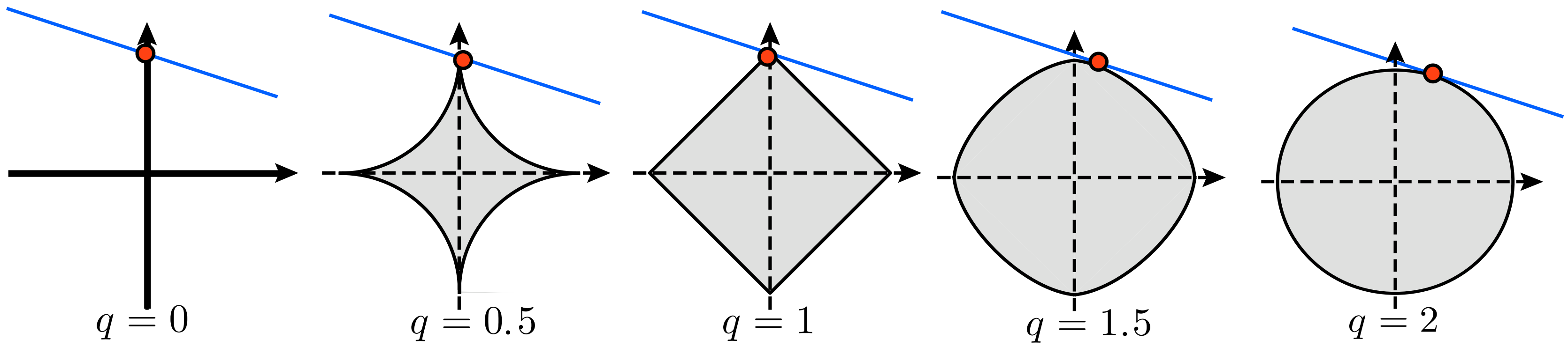
by minimizing an $\mathbf{l^q}$ norm : $\|a\|_q = \left(\sum_{n \in \Gamma} |a[n]|^q \right)^{1/q}$.

$\mathbf{l^0}$ norm : number of non-zero $a[n]$

Not convex for $q < 1$.

$\mathbf{l^1}$ norm : $\|a\|_1 = \sum_{n \in \Gamma} |a[n]|$.

$\mathbf{l^q}$ balls



Basis Pursuit Approximation

- A basis pursuit computes a sparse approximation

$$\tilde{f} = \sum_{p \in \Gamma} \tilde{a}[p]\, \phi_p$$

solution of a convex optimization problem:

$$\tilde{a} = \arg \min_{a \in \mathbf{R}^P} \|a\|_1 \quad \text{subject to} \quad \|f - \sum_{p \in \Gamma} a[p] \phi_p\| \leq \epsilon.$$

- Lagrangian formulation: find $\tilde{a}$ which minimizes

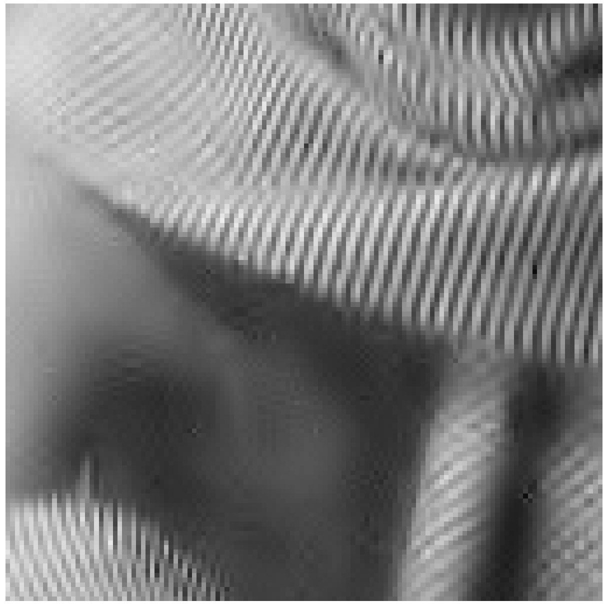
$$\mathcal{L}_1(a) = \frac{1}{2} \|f - \sum_{p \in \Gamma} a[p]\, \phi_p\|^2 + T\, \|a\|_1$$

Basis Pursuit Denoising

Original Image

Noisy Image

Wavelet Thresh.



Basis Pursuit with Wavelet Cosine Diction.

Computation of l1 Minimization

- Many iterative algorithms to minimize

$$\mathcal{L}_1(a) = \frac{1}{2}\|f - \sum_{p\in\Gamma} a[p]\,\phi_p\|^2 + T\,\|a\|_1$$

- Simplest one: iterative thresholding.

Initialisation $b = \Phi f = \{\langle f, \phi_p\rangle\}_{p\in\Gamma}$ and $a_0 = 0$.

For $k > 0$

Gradient step: $\tilde{a}_k = a_k + \gamma(b - \Phi\,\Phi^* a_k)$

Soft thresholding: $a_{k+1} = \tilde{a}_k \max\Big(1 - \frac{\gamma T}{|\tilde{a}_k|}, 0\Big).$

Exact Recovery

- Suppose that the signal is sparse

$$f = \sum_{p \in \Lambda} a[p]\, \phi_p = \sum_{p \in \Lambda} \langle f, \phi_p \rangle\, \tilde{\phi}_p \in \mathbf{V}_\Lambda .$$

can we recover Λ with pursuit algorithms ?

- With matching pursuits, need that

$$C(R^m f, \Lambda^c) = \frac{\max_{q \in \Lambda^c} |\langle R^m f, \phi_q \rangle|}{\max_{p \in \Lambda} |\langle R^m f, \phi_p \rangle|} < 1 \ .$$

- Exact Recovery Criteria

$$ERC(\Lambda) = \sup_{h \in \mathbf{V}_\Lambda} C(h, \Lambda^c) = \max_{q \in \Lambda^c} \sum_{p \in \Lambda} |\langle \tilde{\phi}_p, \phi_q \rangle| \ .$$

ERC Recovery

- **Theorem:** The approximation support Λ of $f \in \mathbf{V}_\Lambda$ is exactly recovered by an orthogonal matching pursuit or a basis pursuit if

$$ERC(\Lambda) < 1.$$

- **Theorem (stability):** If f is not exactly sparse, an orthogonal matching pursuit $\tilde{f}_M$ with $M = |\Lambda|$ iterations satisfies:

$$\|f - \tilde{f}_M\| \leq \left(1 + \frac{|\Lambda|}{A_\Lambda\,(1 - ERC(\Lambda))^2}\right) \|f - f_\Lambda\| \,.$$

- Similar result for a basis pursuit.

Dictionary Coherence

- Dictionary mutual coherence

$$\mu(\mathcal{D}) = \sup_{(p,q)\in\Gamma^2} |\langle \phi_p, \phi_q \rangle|$$

- Dirac-Fourier dictionary: $\mu(\mathcal{D}) = N^{-1/2}$.

- **Theorem:**

$$ERC(\Lambda) < 1 \quad \text{if} \quad |\Lambda| < \frac{1}{2}\left(1 + \frac{1}{\mu(\mathcal{D})}\right)$$

- Exact recovery is possible for sufficiently sparse signals in incoherent dictionaries.

2nd. Conclusion

- Redundant dictionaries can improve approximation, compression, denoising.
- Finding optimal approximation is NP complete but can be approximated with matching or basis pursuits.
- May be used for pattern recognition but problems of instabilities.
- The stability depends upon the dictionary coherence.

- Major applications to inverse problems, super-resolution and compress sensing.