Adaptive filtering in wavelet frames: application to echo (multiple) suppression in geophysics

S. Ventosa, S. Le Roy, I. Huard, A. Pica, H. Rabeson, P. Ricarte, *L. Duval*, M.-Q. Pham, C. Chaux, J.-C. Pesquet

IFPEN
laurent.duval [ad] ifpen.fr

SIERRA 2014, Saint-Étienne

2014/03/25

In just one slide: on echoes and morphing

Wavelet frame coefficients: data and model

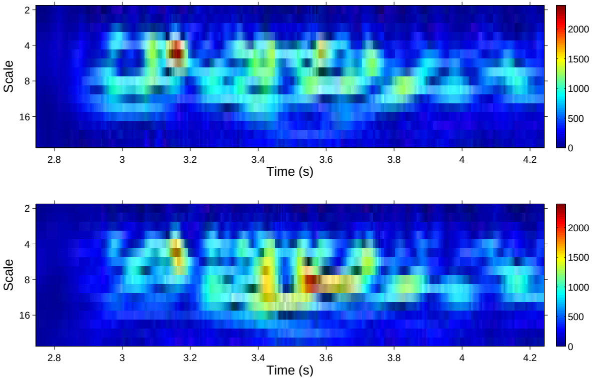


Figure 1: Morphing and adaptive subtraction required

Agenda

1. Issues in geophysical signal processing
2. Problem: multiple reflections (echoes)
   - adaptive filtering with approximate templates
3. Continuous, complex wavelet frames
   - how they (may) simplify adaptive filtering
   - and how they are discretized (back to the discrete world)
4. Adaptive filtering (morphing)
   - without constraint: unary filters (on analytic signals)
   - with constraints: proximal tools
5. Conclusions

Issues in geophysical signal processing

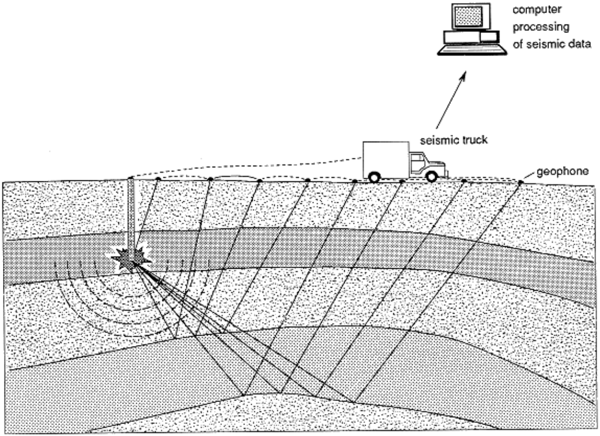


Figure 2: Seismic data acquisition and wave fields

Issues in geophysical signal processing

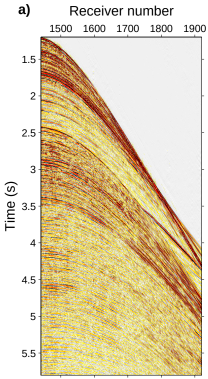


Figure 3: Seismic data: aspect & dimensions (time, offset)

Issues in geophysical signal processing

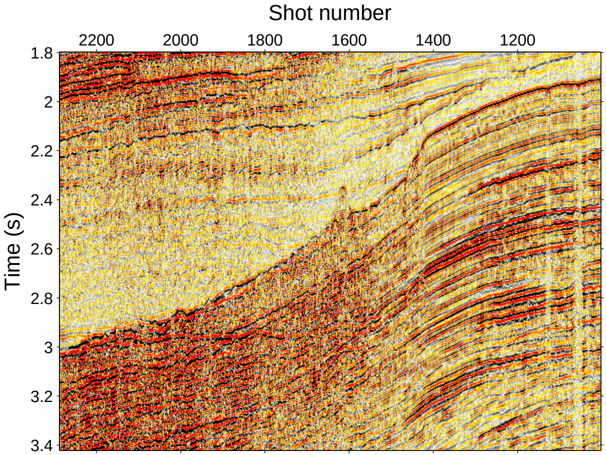


Figure 4: Seismic data: aspect & dimensions (time, offset)

Issues in geophysical signal processing

Reflection seismology:

- seismic waves propagate through the subsurface medium
- seismic traces: seismic wave fields recorded at the surface
  - primary reflections: geological interfaces
  - many types of distortions/disturbances
- processing goal: extract relevant information for seismic data
- led to important signal processing tools:
  - ℓ_1-promoted deconvolution (Claerbout, 1973)
  - wavelets (Morlet, 1975)
- exabytes (10^6 gigabytes) of incoming data
  - need for fast, scalable (and robust) algorithms

Multiple reflections and templates

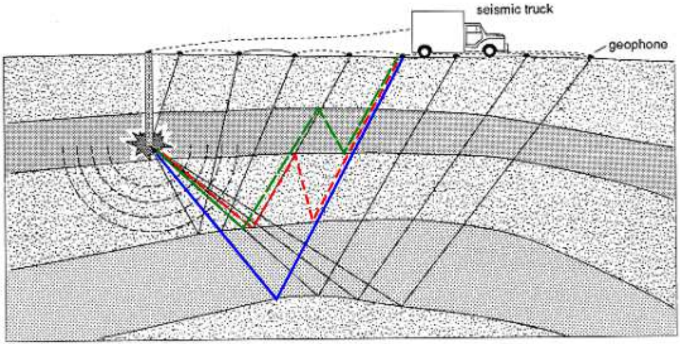


Figure 5: Seismic data acquisition: focus on multiple reflections

Multiple reflections and templates

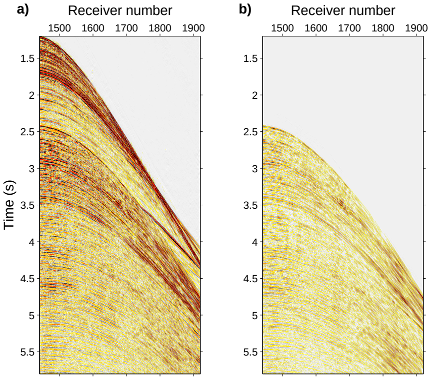


Figure 5: Reflection data: shot gather and template

Multiple reflections and templates

Multiple reflections:

- seismic waves bouncing between layers
- one of the most severe types of interferences
- obscure deep reflection layers
- high cross-correlation between primaries (p) and multiples (m)
- additional incoherent noise (n)
- $d(t) = p(t)+m(t)+n(t)$
  - with approximate templates: $r_1(t)$, $r_2(t)$,... $r_J(t)$
- Issue: how to adapt and subtract approximate templates?

Multiple reflections and templates

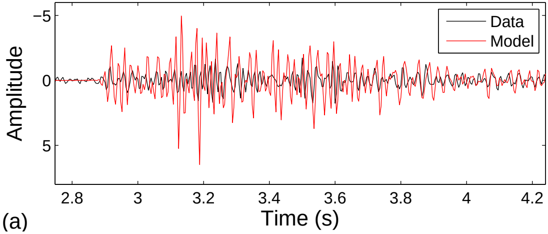


Figure 6: Multiple reflections: data trace d and template r_1

Multiple reflections and templates

Multiple filtering:

- multiple prediction (correlation, wave equation) has limitations
- templates are not accurate
  - $m(t) \approx \sum_j h_j \ast r_j$?
  - standard: identify, apply a matching filer, subtract

$$\mathbf{h}_{\text{opt}} = \underset{\mathbf{h} \in \mathbb{R}^l}{\arg\min} \left\| d - \mathbf{h} \ast \mathbf{r} \right\|^2$$

- primaries and multiples are not (fully) uncorrelated
  - same (seismic) source
  - similarities/dissimilarities in time/frequency
- variations in amplitude, waveform, delay
- issues in matching filter length:
  - short filters and windows: local details
  - long filters and windows: large scale effects

Multiple reflections and templates

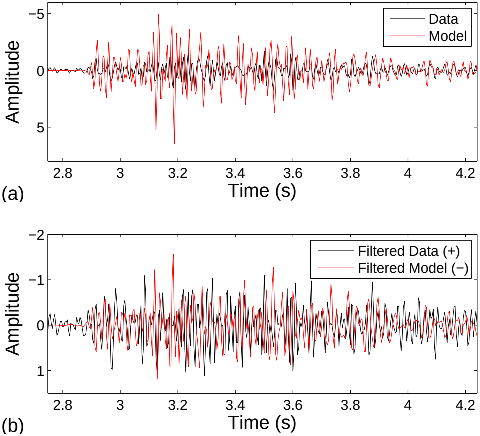


Figure 7: Multiple reflections: data trace, template and adaptation

Multiple reflections and templates

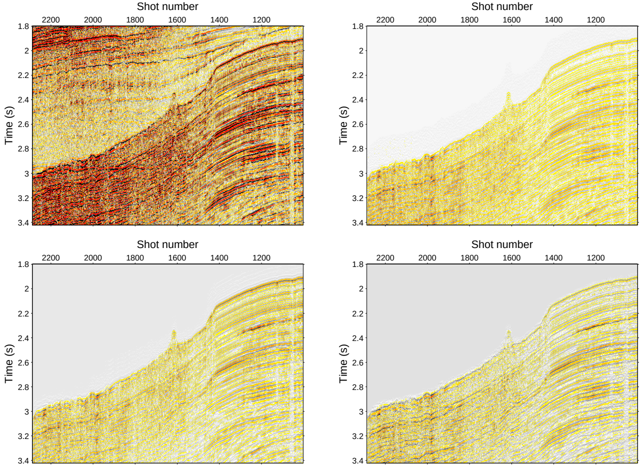


Figure 8: Multiple reflections: data trace and templates, 2D version

Multiple reflections and templates

- A long history of multiple filtering methods
  - general idea: combine adaptive filtering and transforms
    - data transforms: Fourier, Radon
    - enhance the differences between primaries, multiples and noise
    - reinforce the adaptive filtering capacity
  - intrication with adaptive filtering?
    - might be complicated (think about inverse transform)
- First simple approach:
  - exploit the non-stationary in the data
  - naturally allow both large scale & local detail matching

$\Rightarrow$ Redundant wavelet frames

- intermediate complexity in the transform
- simplicity in the (unary/FIR) adaptive filtering

Hilbert transform and pairs

Reminders [Gabor-1946][Ville-1948]

$$\widehat{\mathcal{H}\{f\}}(\omega) = -\imath\,\mathrm{sign}(\omega)\,\widehat{f}(\omega)$$

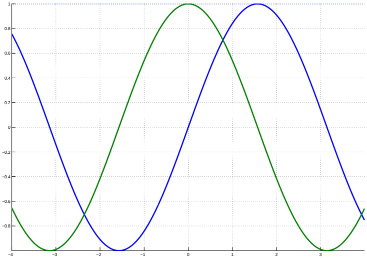


Figure 9: Hilbert pair 1

Hilbert transform and pairs

Reminders [Gabor-1946][Ville-1948]

$$\widehat{\mathcal{H}\{f\}}(\omega) = -\imath\,\mathrm{sign}(\omega)\,\widehat{f}(\omega)$$

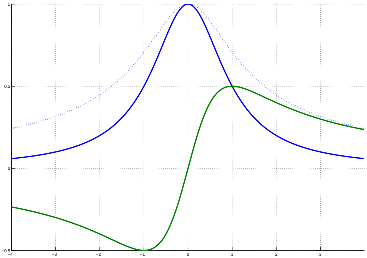


Figure 9: Hilbert pair 2

Hilbert transform and pairs

Reminders [Gabor-1946][Ville-1948]

$$\widehat{\mathcal{H}\{f\}}(\omega) = -\imath\,\mathrm{sign}(\omega)\,\widehat{f}(\omega)$$

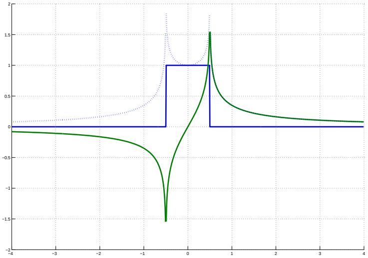


Figure 9: Hilbert pair 3

Hilbert transform and pairs

Reminders [Gabor-1946][Ville-1948]

$$\widehat{\mathcal{H}\{f\}}(\omega) = -\imath\,\mathrm{sign}(\omega)\hat{f}(\omega)$$

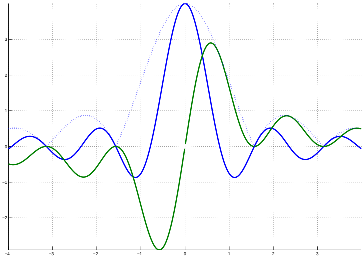


Figure 9: Hilbert pair 4

Continuous & complex wavelets

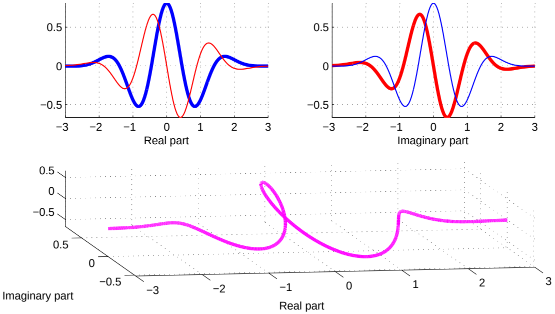


Figure 10: Complex wavelets at two different scales — 1

Continuous & complex wavelets

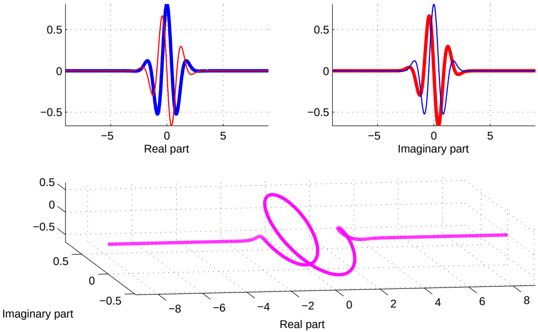


Figure 11: Complex wavelets at two different scales — 2

Continuous wavelets

- Transformation group:

$$\text{affine} = \text{translation } (\tau) + \text{dilation } (a)$$

- Basis functions:

$$\psi_{\tau,a}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-\tau}{a}\right)$$

  - $a > 1$: dilation
  - $a < 1$: contraction
  - $1/\sqrt{a}$: energy normalization
  - multiresolution (vs monoresolution in STFT/Gabor)

$$\psi_{\tau,a}(t) \xrightarrow{\text{FT}} \sqrt{a}\Psi(af)e^{-\imath 2\pi f\tau}$$

Continuous wavelets

- Definition

$$C_s(\tau, a) = \int s(t)\psi^*_{\tau,a}(t)dt$$

- Vector interpretation

$$C_s(\tau, a) = \langle s(t), \psi_{\tau,a}(t)\rangle$$

  projection onto time-scale atoms (vs STFT time-frequency)
- Redundant transform: $\tau \rightarrow \tau \times a$ “samples”
- Parseval-like formula

$$C_s(\tau, a) = \langle S(f), \Psi_{\tau,a}(f)\rangle$$

  $\Rightarrow$ sounder time-scale domain operations! (cf. Fourier)

Continuous wavelets

Introductory example

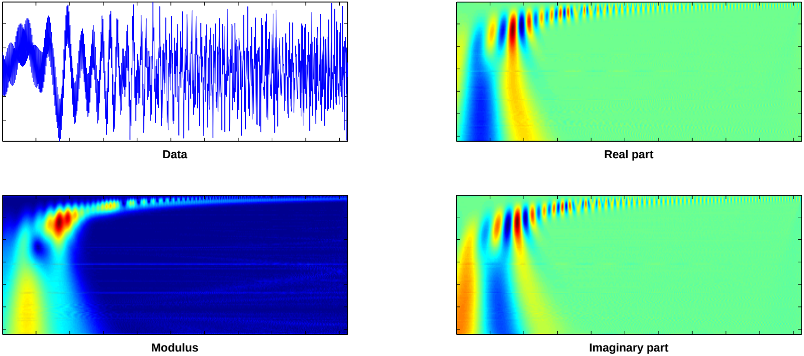


Figure 12: Noisy chirp mixture in time-scale & sampling

Continuous wavelets

Noise spread & feature simplification (signal vs wiggle)

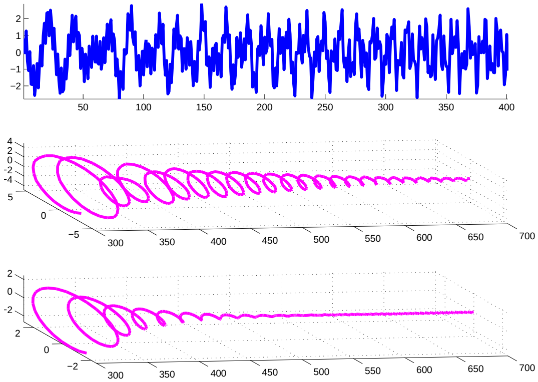


Figure 13: Noisy chirp mixture in time-scale: zoomed scaled wiggles

Continuous wavelets

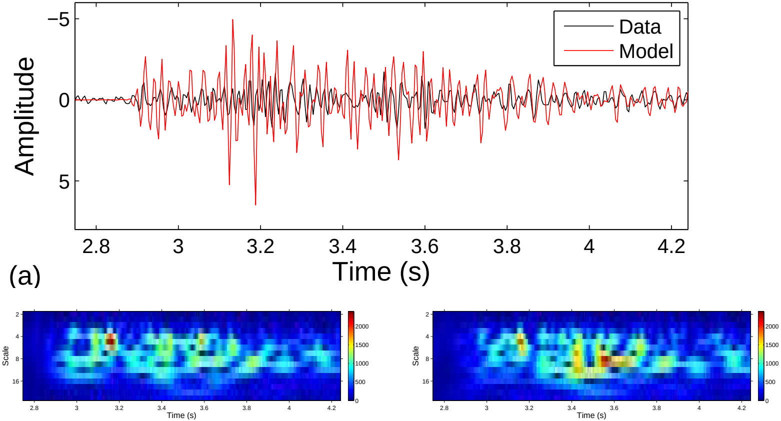


Figure 14: Which morphing is easier: time or time-scale?

Continuous wavelets

- Inversion with another wavelet ϕ

$$s(t) = \iint C_s(u,a)\phi_{u,a}(t)\frac{duda}{a^2}$$

$\Rightarrow$ time-scale domain processing! (back to the trace signal)

- Scalogram

$$|C_s(t,a)|^2$$

- Energy conversation

$$E = \iint |C_s(t,a)|^2 \frac{dtda}{a^2}$$

- Parseval-like formula

$$\langle s_1, s_2 \rangle = \iint C_{s_1}(t,a)C^*_{s_2}(t,a)\frac{dtda}{a^2}$$

Continuous wavelets

- Wavelet existence: admissibility criterion

$$0 < A_h = \int_0^{+\infty} \frac{\widehat{\Phi}^*(\nu)\Psi(\nu)}{\nu} d\nu = \int_{-\infty}^{0} \frac{\widehat{\Phi}^*(\nu)\Psi(\nu)}{\nu} d\nu < \infty$$

  generally normalized to 1
- Easy to satisfy (common freq. support midway 0 & ∞)
- With $\psi = \phi$, induces band-pass property:
  - necessary condition: $|\Phi(0)| = 0$, or zero-average shape
  - amplitude spectrum neglectable w.r.t. $|\nu|$ at infinity
- Example: Morlet-Gabor (not truly admissible)

$$\psi(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{t^2}{2\sigma^2}} e^{-\imath 2\pi f_0 t}$$

Discretization and redundancy

Being practical again: dealing with discrete signals

- Can one sample in time-scale (CWT) domain:

$$C_s(\tau, a) = \int s(t)\psi^*_{\tau,a}(t)dt, \quad \psi_{\tau,a}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-\tau}{a}\right)$$

  with $c_{j,k} = C_s(kb_0a_0^j, a_0^j),\ (j,k) \in \mathbb{Z}$ and still be able to recover $s(t)$?
- Result 1 (Daubechies, 1984): there exists a wavelet frame if $a_0b_0 < C$, (depending on ψ). A frame is generally redundant
- Result 2 (Meyer, 1985): there exist an orthonormal basis for a specific ψ (non trivial, Meyer wavelet) and $a_0 = 2$ $b_0 = 1$

Now: how to choose the practical level of redundancy?

Discretization and redundancy

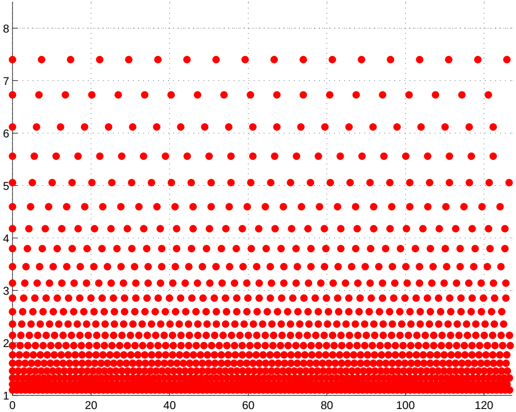


Figure 15: Wavelet frame sampling: $J = 21$, $b_0 = 1$, $a_0 = 1.1$

Discretization and redundancy

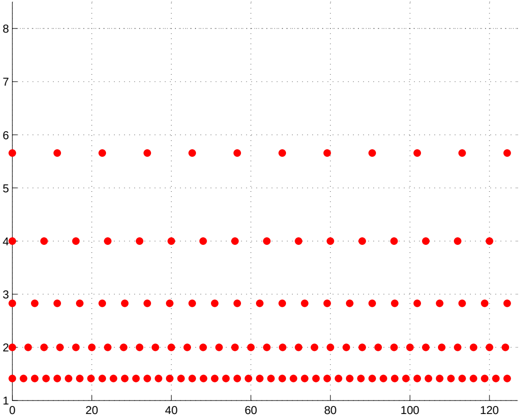


Figure 15: Wavelet frame sampling: $J = 5$, $b_0 = 2$, $a_0 = \sqrt{2}$

Discretization and redundancy

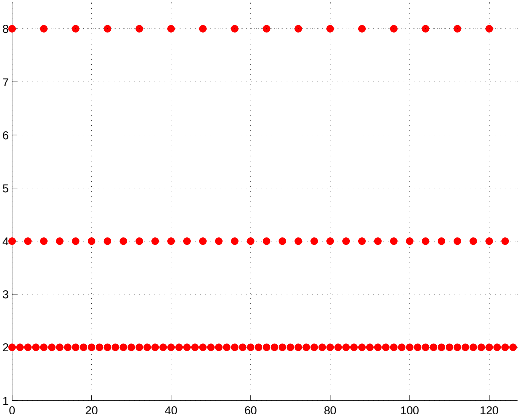


Figure 15: Wavelet frame sampling: $J = 3$, $b_0 = 1$, $a_0 = 2$

Discretization and redundancy

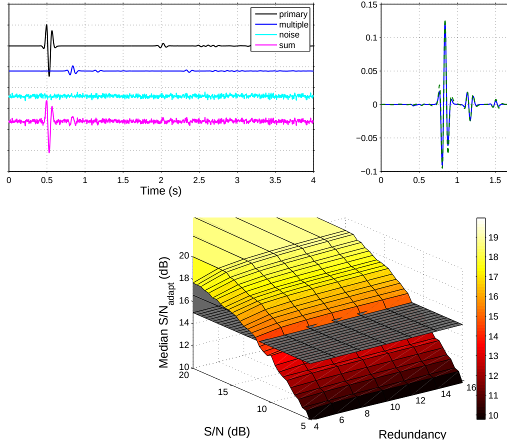


Figure 16: Redundancy selection with variable noise experiments

Discretization and redundancy

- Complex Morlet wavelet:

$$\psi(t) = \pi^{-1/4} e^{-i\omega_0 t} e^{-t^2/2},\ \omega_0\text{: central frequency}$$

- Discretized time r, octave j, voice v:

$$\psi^v_{r,j}[n] = \frac{1}{\sqrt{2^{j+v/V}}} \psi\left(\frac{nT - r2^j b_0}{2^{j+v/V}}\right),\ b_0\text{: sampling at scale zero}$$

- Time-scale analysis:

$$\mathbf{d} = d^v_{r,j} = \left\langle d[n], \psi^v_{r,j}[n] \right\rangle = \sum_n d[n] \overline{\psi^v_{r,j}[n]}$$

Discretization and redundancy

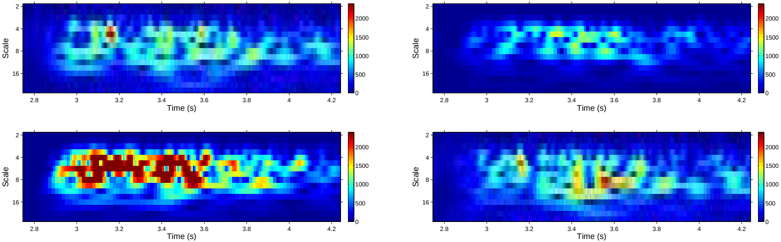


Figure 17: Morlet wavelet scalograms, data and templates

Take advantage from the closest similarity/dissimilarity:

- remember wiggles: on sliding windows, at each scale, a single complex coefficient compensates amplitude and phase

Unary filters

- Windowed unary adaptation: complex unary filter $\mathbf{h}$ ($\mathbf{a}_{\text{opt}}$) compensates delay/amplitude mismatches:

$$\mathbf{a}_{\text{opt}} = \underset{\{a_j\}(j\in J)}{\arg\min} \left\| \mathbf{d} - \sum_j a_j \mathbf{r}_k \right\|^2$$

- Vector Wiener equations for complex signals:

$$\langle \mathbf{d}, \mathbf{r}_m \rangle = \sum_j a_j \langle \mathbf{r}_j, \mathbf{r}_m \rangle$$

- Time-scale synthesis:

$$\hat{d}[n] = \sum_r \sum_{j,v} \hat{d}^v_{r,j} \widetilde{\psi}^v_{r,j}[n]$$

Results

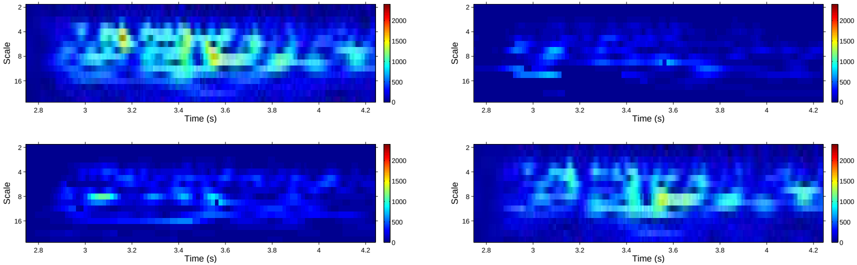


Figure 18: Wavelet scalograms, data and templates, after unary adaptation

Results (reminders)

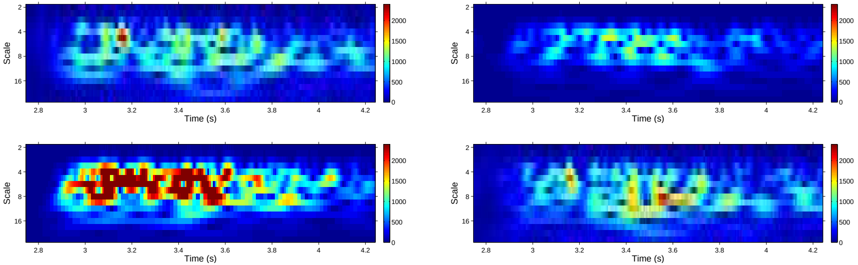


Figure 19: Wavelet scalograms, data and templates

Results

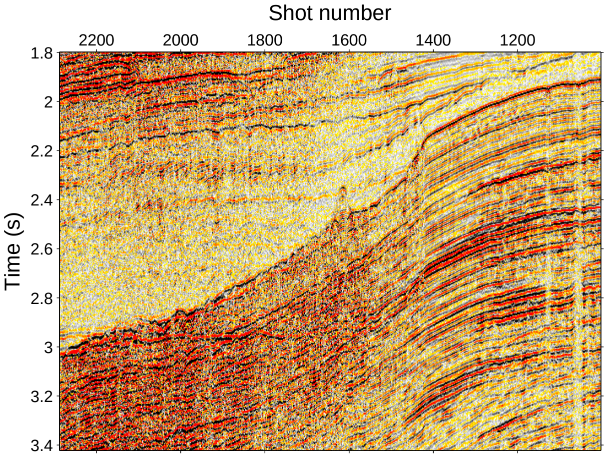


Figure 20: Original data

Results

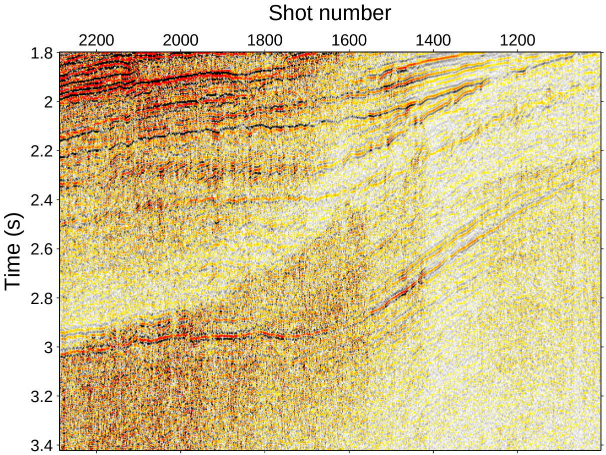


Figure 21: Filtered data, "best" template

Results

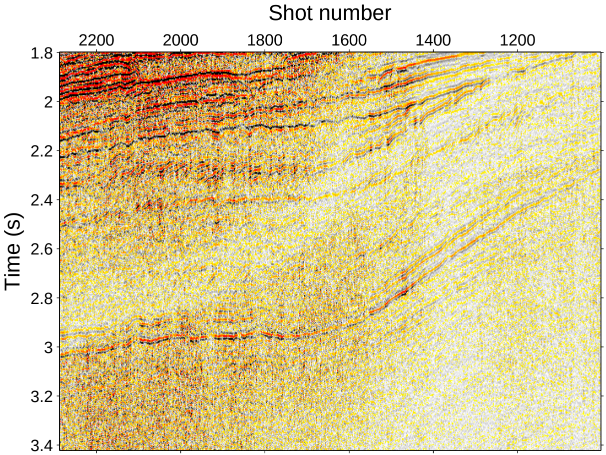


Figure 22: Filtered data, three templates

Going a little further

Impose geophysical data related assumptions: e.g. sparsity

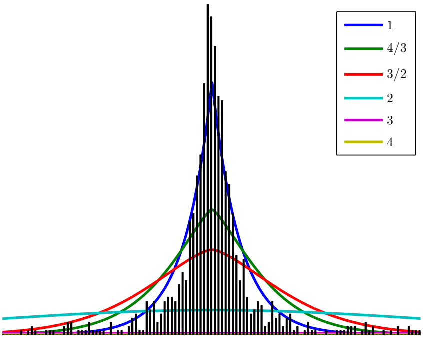


Figure 23: Generalized Gaussian modeling of seismic data wavelet frame decomposition with different power laws.

Variational approach

$$\boxed{\underset{x\in\mathcal{H}}{\text{minimize}} \quad \sum_{j=1}^{J} f_j(L_j x)}$$

with lower-semicontinuous proper convex functions f_j and bounded linear operators L_j.

Variational approach

$$\boxed{\underset{x\in\mathcal{H}}{\text{minimize}} \quad \sum_{j=1}^{J} f_j(L_j x)}$$

with lower-semicontinuous proper convex functions f_j and bounded linear operators L_j.

- f_j can be related to noise (e.g. a quadratic term when the noise is Gaussian),

Variational approach

$$\boxed{\underset{x\in\mathcal{H}}{\text{minimize}} \quad \sum_{j=1}^{J} f_j(L_j x)}$$

with lower-semicontinuous proper convex functions f_j and bounded linear operators L_j.

- f_j can be related to noise (e.g. a quadratic term when the noise is Gaussian),
- f_j can be related to some a priori on the target solution (e.g. an a priori on the wavelet coefficient distribution),

Variational approach

$$\boxed{\underset{x\in\mathcal{H}}{\text{minimize}} \quad \sum_{j=1}^{J} f_j(L_j x)}$$

with lower-semicontinuous proper convex functions f_j and bounded linear operators L_j.

- f_j can be related to noise (e.g. a quadratic term when the noise is Gaussian),
- f_j can be related to some a priori on the target solution (e.g. an a priori on the wavelet coefficient distribution),
- f_j can be related to a constraint (e.g. a support constraint),

Variational approach

$$\boxed{\underset{x\in\mathcal{H}}{\text{minimize}} \quad \sum_{j=1}^{J} f_j(L_j x)}$$

with lower-semicontinuous proper convex functions f_j and bounded linear operators L_j.

- f_j can be related to noise (e.g. a quadratic term when the noise is Gaussian),
- f_j can be related to some a priori on the target solution (e.g. an a priori on the wavelet coefficient distribution),
- f_j can be related to a constraint (e.g. a support constraint),
- L_j can model a blur operator,

Variational approach

$$\boxed{\underset{x\in\mathcal{H}}{\text{minimize}} \quad \sum_{j=1}^{J} f_j(L_j x)}$$

with lower-semicontinuous proper convex functions f_j and bounded linear operators L_j.

- f_j can be related to noise (e.g. a quadratic term when the noise is Gaussian),
- f_j can be related to some a priori on the target solution (e.g. an a priori on the wavelet coefficient distribution),
- f_j can be related to a constraint (e.g. a support constraint),
- L_j can model a blur operator,
- L_j can model a gradient operator (e.g. total variation),

Variational approach

$$\boxed{\underset{x\in\mathcal{H}}{\text{minimize}} \quad \sum_{j=1}^{J} f_j(L_j x)}$$

with lower-semicontinuous proper convex functions f_j and bounded linear operators L_j.

- f_j can be related to noise (e.g. a quadratic term when the noise is Gaussian),
- f_j can be related to some a priori on the target solution (e.g. an a priori on the wavelet coefficient distribution),
- f_j can be related to a constraint (e.g. a support constraint),
- L_j can model a blur operator,
- L_j can model a gradient operator (e.g. total variation),
- L_j can model a frame operator.

Problem re-formulation

$$\underbrace{d^{(k)}}_{\text{observed signal}} = \underbrace{\bar{p}^{(k)}}_{\text{primary}} + \underbrace{\bar{m}^{(k)}}_{\text{multiple}} + \underbrace{n^{(k)}}_{\text{noise}}$$

Assumption: templates linked to $\bar{m}^{(k)}$ throughout time-varying (FIR) filters:

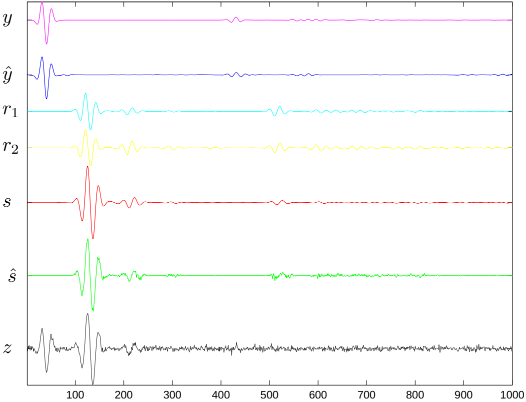
$$\boxed{\bar{m}^{(k)} = \sum_{j=0}^{J-1} \sum_{p} \bar{h}_j^{(p)}(k) r_j^{(k-p)}}$$

where

- $\bar{h}_j^{(k)}$: unknown impulse response of the filter corresponding to template j and time k, then:

$$\boxed{\underbrace{d}_{\text{observed signal}} = \underbrace{\bar{p}}_{\text{primary}} + \mathbf{R} \underbrace{\bar{\mathbf{h}}}_{\text{filter}} + \underbrace{n}_{\text{noise}}}$$

Results: synthetics (noise: $\sigma = 0.08$)



Original signal y

Estimated signal $\widehat{y}$

Model r_1

Model r_2

Original multiple s

Estimated multiple $\widehat{s}$

Observed signal z

Assumptions

- F is a frame, $\bar{p}$ is a realization of a random vector P:

$$f_P(p) \propto \exp(-\varphi(Fp)),$$

- $\bar{\mathbf{h}}$ is a realization of a random vector H:

$$f_H(\mathbf{h}) \propto \exp(-\rho(\mathbf{h})),$$

- n is a realization of a random vector N, of probability density:

$$f_N(n) \propto \exp(-\psi(n)),$$

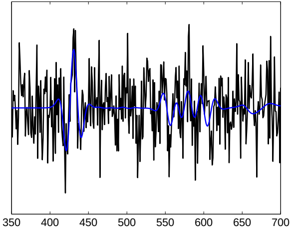
- slow variations along time and concentration of the filters

$$|h_j^{(n+1)}(p) - h_j^{(n)}(p)| \leqslant \varepsilon_{j,p} \; ; \qquad \sum_{j=0}^{J-1} \widetilde{\rho}_j(h_j) \leqslant \tau$$

Results: synthetics

$$\underset{y\in\mathbb{R}^N,\mathbf{h}\in\mathbb{R}^{NP}}{\text{minimize}} \quad \underbrace{\psi\big(z-\mathbf{Rh}-y\big)}_{\text{fidelity: noise-realted}} + \underbrace{\varphi(Fy)}_{\textit{a priori}\text{ on signal}} + \underbrace{\rho(\mathbf{h})}_{\textit{a priori}\text{ on filters}}$$

- $\varphi_k = \kappa_k|\cdot|$ (ℓ_1-norm) where $\kappa_k > 0$
- $\widetilde{\rho}_j(h_j)$: $\|h_j\|_{\ell_1}$, $\|h_j\|^2_{\ell_2}$ or $\|h_j\|_{\ell_{1,2}}$
- $\psi\big(z-\mathbf{Rh}-y\big)$: quadratic (Gaussian noise)



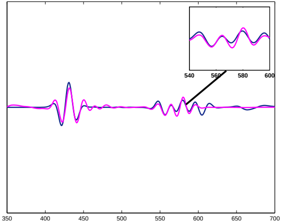


Figure 24: Simulated results with heavy noise.

Results: synthetics

	SNR_y			SNR_s		
$\sigma \setminus \widetilde{\rho}$	ℓ_1	ℓ_2	$\ell_{1,2}$	ℓ_1	ℓ_2	$\ell_{1,2}$
0.01	20.90	21.23	**23.57**	24.36	24.68	**26.74**
0.02	20.89	21.16	**23.51**	22.53	23.02	**23.76**
0.04	19.00	19.90	**20.67**	**20.15**	20.14	19.84
0.08	**17.55**	16.81	17.34	**16.96**	16.56	15.96

Signal-to-noise ratios (SNR, averaged over 100 noise realisations)

Results: potential on real data

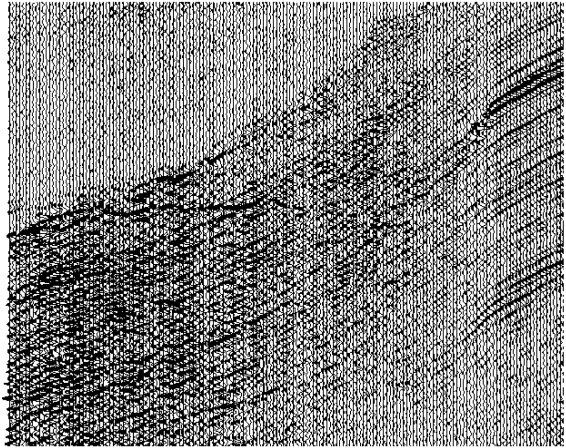


Figure 25: Portion of a receiver gather: recorded data.

Results: potential on real data

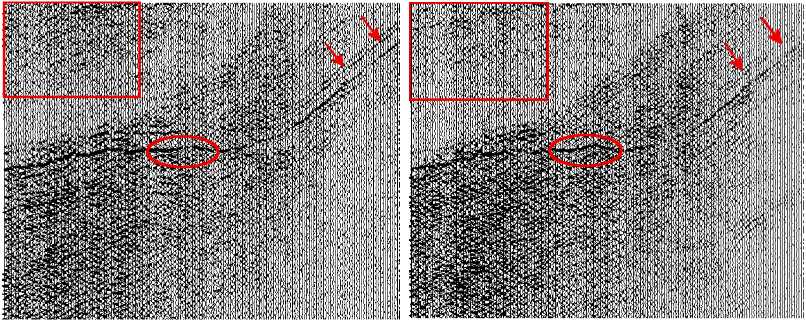


Figure 25: (a) Unary filters (b) Proximal FIR filters.

Conclusions

Take-away messages:

- Practical side
  - Competitive with more standard 2D processing
  - Very fast (unary part): industrial integration
- Technical side
  - Lots of choices, insights from 1D or 1.5D
  - Non-stationary, wavelet-based, adaptive multiple filtering
  - Take good care in cascaded processing
- Present work
  - Other applications: pattern matching, (voice) echo cancellation, ultrasonic/acoustic emissions with home-made templates
  - Going 2D: crucial choices on redundancy, directionality

Conclusions

Now what's next: curvelets, shearlets, dual-tree complex wavelets?

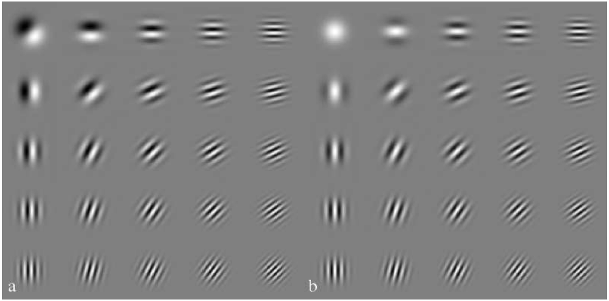


Figure 26: From T. Lee (TPAMI-1996): 2D Gabor filters (odd and even) or Weyl-Heisenberg coherent states

References

Ventosa, S., S. Le Roy, I. Huard, A. Pica, H. Rabeson, P. Ricarte, and L. Duval, 2012, Adaptive multiple subtraction with wavelet-based complex unary Wiener filters: Geophysics, **77**, V183–V192; http://arxiv.org/abs/1108.4674

Pham, M. Q., C. Chaux, L. Duval, L. and J.-C. Pesquet, 2014, A Primal-Dual Proximal Algorithm for Sparse Template-Based Adaptive Filtering: Application to Seismic Multiple Removal: IEEE Trans. Signal Process., accepted; http://tinyurl.com/proximal-multiple

Jacques, L., L. Duval, C. Chaux, and G. Peyré, 2011, A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity: Signal Process., **91**, 2699–2730; http://arxiv.org/abs/1101.5320