A NEW SPARSITY-ENABLED SIGNAL SEPARATION METHOD BASED ON SIGNAL RESONANCE

Ivan Selesnick
Polytechnic Institute of New York University
Brooklyn, New York
Introduction

Problem: Decomposition of a signal into the sum of two components:

1. Oscillatory (rhythmic, tonal) component
2. Transient (non-oscillatory) component

Outline

1. Signal resonance and Q-factors
2. Morphological component analysis (MCA)
3. Rational-dilation wavelet transform (RADWT)
4. Split augmented Lagrangian shrinkage algorithm (SALSA)
5. Examples

References (MDCT, etc)
Oscillatory (rhythmic) and Transient Components in EEG

Many measured signals have both an oscillatory and a non-oscillatory component.

Rhythms of the EEG:

Delta 0 - 3 Hz
Theta 4 - 7 Hz
Alpha 8 - 12 Hz
Beta 12 - 30 Hz
Gamma 26 - 100 Hz

Transients in EEG due to:
1) unwanted measurement artifacts
2) non-rhythmic brain activity (spikes, spindles, and vertex waves)
Figure 1: The resonance of an isolated pulse can be quantified by its Q-factor, defined as the ratio of its center frequency to its bandwidth. Pulses 1 and 3, essentially a single cycle in duration, are low-resonance pulses. Pulses 2 and 4, whose oscillations are more sustained, are high-resonance pulses.
Resonance-based signal decomposition

Figure 2: Resonance- and frequency-based filtering. (a) Decomposition of a test signal into high- and low-resonance components. The high-resonance signal component is sparsely represented using a high Q-factor RADWT. Similarly, the low-resonance signal component is sparsely represented using a low Q-factor RADWT. (b) Decomposition of a test signal into low, mid, and high frequency components using LTI discrete-time filters.
Resonance-based signal decomposition must be nonlinear

Figure 3: Resonance-based signal decomposition must be nonlinear: The signal in the bottom left panel is the sum of the signals above it; however, the low-resonance component of a sum is not the sum of the low-resonance components. The same is true for the high-resonance component. Neither the low- nor high-resonance components satisfy the superposition property.
Rational-dilation wavelet transform (RADWT)

Prior work on rational-dilation wavelet transforms addresses the critically-sampled case.

2. P. Auscher (1992)

Rational-dilation wavelet transform (RADWT)

Figure 4: Analysis and synthesis filter banks for the implementation of the rational-dilation wavelet transform (RADWT). The dilation factor is $q/p$ and the redundancy is $(s(1 - p/q))^{-1}$ assuming iteration of the filter bank on its low-pass (upper) branch ad infinitum.

\[
Y(\omega) = \sum_{k=0}^{q-1} L_k(\omega) X\left(\omega + p \frac{k}{q} \frac{2\pi}{p}\right) + \sum_{k=0}^{s-1} M_k(\omega) X\left(\omega + k \frac{2\pi}{s}\right),
\]

where

\[
L_k(\omega) = \frac{1}{pq} \sum_{n=0}^{p-1} H\left(\frac{\omega}{p} + k \frac{2\pi}{q} + n \frac{2\pi}{p}\right) H^*\left(\frac{\omega}{p} + n \frac{2\pi}{p}\right),
\]

\[
M_k(\omega) = \frac{1}{s} \left[ G\left(\omega + k \frac{2\pi}{s}\right) G^*(\omega) \right].
\]

There are no perfect reconstruction filters with rational-transfer functions unless the filter bank is orthonormal ($p + 1 = q = s$).

For overcomplete case, use filters with non-rational transfer functions for PR.
Perfect reconstruction can be attained by

\[ H(\omega) = \begin{cases} \sqrt{pq} & |\omega| \leq (1 - \frac{1}{s}) \frac{\pi}{p} \\ 0 & |\omega| \in \left[ \frac{\pi}{q}, \pi \right] \end{cases} \]

\[ G(\omega) = \begin{cases} 0 & |\omega| \leq (1 - \frac{1}{s}) \pi \\ \sqrt{s} & |\omega| \in \left[ \frac{p\pi}{q}, \pi \right] \end{cases} \]

where transition-bands are chosen so as to satisfy

\[ \frac{1}{pq} \left| H\left(\frac{\omega}{p}\right) \right|^2 + \frac{1}{s} \left| G(\omega) \right|^2 = 1. \]

With \( q = 3, p = 2, s = 2 \):
Figure 5: Low Q-factor rational-dilation wavelet transforms (RADWT) with $p = 2$, $q = 3$, $s = 1$. The wavelet is approximately the Mexican hat function. The RADWT is 3-times overcomplete.
Figure 6: High Q-factor rational-dilation wavelet transforms (RADWT) with $p = 5$, $q = 6$, $s = 2$. The dilation factor is 1.2, much closer to 1 than the dyadic wavelet transform. The RADWT is 3-times overcomplete.
Low Q-factor vs High Q-factor RADWT

SUBBANDS (LOW−Q RADWT)

SUBBANDS (HIGH−Q RADWT)

P = 2, Q = 3, S = 1, Levels = 10
Dilation = 1.50, Redundancy = 3.00

P = 5, Q = 6, S = 2, Levels = 20
Dilation = 1.20, Redundancy = 3.00
Low Q-factor vs High Q-factor RADWT after sparsification

SUBBANDS (LOW-Q RADWT)

SUBBANDS (HIGH-Q RADWT)

P = 2, Q = 3, S = 1, Levels = 10  
Dilation = 1.50, Redundancy = 3.00

P = 5, Q = 6, S = 2, Levels = 20  
Dilation = 1.20, Redundancy = 3.00
Low Q-factor vs High Q-factor RADWT

**SUBBANDS (LOW-Q RADWT)**

- SUBBAND 1: 2.58%
- SUBBAND 2: 6.31%
- SUBBAND 3: 10.70%
- SUBBAND 4: 16.11%
- SUBBAND 5: 21.44%
- SUBBAND 6: 22.50%
- SUBBAND 7: 13.51%
- SUBBAND 8: 4.87%
- SUBBAND 9: 1.42%

**SUBBANDS (HIGH-Q RADWT)**

- SUBBAND 6: 1.82%
- SUBBAND 7: 2.75%
- SUBBAND 8: 3.02%
- SUBBAND 9: 3.82%
- SUBBAND 10: 5.65%
- SUBBAND 11: 6.82%
- SUBBAND 12: 6.95%
- SUBBAND 13: 8.72%
- SUBBAND 14: 12.37%
- SUBBAND 15: 13.98%
- SUBBAND 16: 12.04%
- SUBBAND 17: 8.44%
- SUBBAND 18: 5.41%
- SUBBAND 19: 3.07%

- P = 2, Q = 3, S = 1, Levels = 10, Dilation = 1.50, Redundancy = 3.00
- P = 5, Q = 6, S = 2, Levels = 20, Dilation = 1.20, Redundancy = 3.00
Low Q-factor vs High Q-factor RADWT after sparsification

### SUBBANDS (LOW-Q RADWT)

- **P = 2, Q = 3, S = 1, Levels = 10**
- **Dilation = 1.50, Redundancy = 3.00**

### SUBBANDS (HIGH-Q RADWT)

- **P = 5, Q = 6, S = 2, Levels = 20**
- **Dilation = 1.20, Redundancy = 3.00**
Constant-Q vs Constant-BW

**Constant-Q**

- Frequency responses
- Analysis function
- Fixed 'resonance'
- Frequency-dependent temporal duration

**Constant-BW**

- Frequency responses
- Analysis function
- Frequency-dependent 'resonance'
- Fixed temporal duration
**Rational-dilation wavelet transform (RADWT)**

Summary:

1. Fully-discrete, modestly overcomplete

2. Exact perfect reconstruction (‘self-inverting’)

3. Adjustable Q-factor:
   Can attain higher Q-factors than (or same low Q-factor of) the dyadic WT.
   \[ \Rightarrow \] Can achieve higher-frequency resolution needed for oscillatory signals.

4. Samples the time-frequency plane more densely in both time and frequency.
   \[ \Rightarrow \] Exactly invertible, fully-discrete approximation of the continuous WT.

5. FFT-based implementation

Reference:

Morphological Component Analysis (MCA)

Given an observed signal

\[ x = x_1 + x_2, \text{ with } x, x_1, x_2 \in \mathbb{R}^N, \]

the goal of MCA is to estimate/determine \( x_1 \) and \( x_2 \) individually. Assuming that \( x_1 \) and \( x_2 \) can be sparsely represented in bases (or frames) \( S_1 \) and \( S_2 \) respectively, they can be estimated by minimizing the objective function,

\[
J(w_1, w_2) = \|x - S_1 w_1 - S_2 w_2\|_2^2 + \lambda_1 ||w_1||_1 + \lambda_2 ||w_2||_1
\]

with respect to \( w_1 \) and \( w_2 \). Then MCA provides the estimates

\[ \hat{x}_1 = S_1 w_1 \]

and

\[ \hat{x}_2 = S_2 w_2. \]

Reference:

Why not an $\ell_2$-norm penalty?

If the $\ell_2$-norm is used for the penalty term,

$$J(w_1, w_2) = \|x - S_1 w_1 - S_2 w_2\|^2_2 + \lambda_1 \|w_1\|^2_2 + \lambda_2 \|w_2\|^2_2,$$

then, using $S_1 S_1^t = S_2 S_2^t = I$, the minimizing $w_1$ and $w_2$ can be found in closed form:

$$w_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_1 \lambda_2} S_1^t x$$

$$w_2 = \frac{\lambda_w}{\lambda_1 + \lambda_2 + \lambda_1 \lambda_2} S_2^t x$$

and the estimated components $\hat{x}_1 = S_1 w_1$ and $\hat{x}_2 = S_2 w_2$ are given by

$$\hat{x}_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2 + \lambda_1 \lambda_2} x$$

$$\hat{x}_2 = \frac{\lambda_2}{\lambda_1 + \lambda_2 + \lambda_1 \lambda_2} x$$

Both $\hat{x}_1$ and $\hat{x}_2$ are just scaled versions of $x$.

$$\implies \text{No separation at all!}$$
MCA as a linear inverse problem

The objective function

\[ J(w_1, w_2) = \| x - S_1 w_1 - S_2 w_2 \|_2^2 + \lambda_1 \| w_1 \|_1 + \lambda_2 \| w_2 \|_1 \]

can be written as

\[ J(w) = \| x - Hw \|_2^2 + \| \lambda^t w \|_1 \]

where

\[ H = \begin{bmatrix} S_1 & S_2 \end{bmatrix}, \quad w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}. \]

An \( \ell_1 \)-regularized linear inverse problem . . .

• Non-differentiable

• Convex

\[ \implies \text{Use } \textit{Iterative Soft Thresholding Algorithm (ISTA)} \text{ or another algorithm to minimize } J(w). \]

Other algorithms include FISTA, SALSA, TwIST, etc.
Split augmented Lagrangian shrinkage algorithm (SALSA)

SALSA is an algorithm for minimizing

\[ J(w) = \|x - Hw\|_2^2 + \lambda \|w\|_1 \]

SALSA is based on the minimization of

\[
\min_u f_1(u) + f_2(u)
\]

by the alternating split augmented Lagrangian algorithm:

\[
u^{(k+1)} = \operatorname{arg\,min}_u f_1(u) + \mu \|u - v^{(k)} - d^{(k)}\|_2^2 \quad (2)
\]

\[
v^{(k+1)} = \operatorname{arg\,min}_v f_2(v) + \mu \|u^{(k+1)} - v - d^{(k)}\|_2^2 \quad (3)
\]

\[
d^{(k+1)} = d^{(k)} - u^{(k+1)} + v^{(k+1)} \quad (4)
\]

Reference:

Applying SALSA to the MCA problem yields the iterative algorithm:

\[ b_1^{(k)} = S_1^t x + \mu (w_1^{(k)} + d_1^{(k)}) \]
\[ b_2^{(k)} = S_2^t x + \mu (w_2^{(k)} + d_2^{(k)}) \]
\[ c^{(k)} = S_1 b_1^{(k)} + S_2 b_2^{(k)} \]
\[ u_1^{(k+1)} = \frac{1}{\mu} b_1^{(k)} - \frac{1}{\mu(\mu + 2)} S_1^t c^{(k)} \]
\[ u_2^{(k+1)} = \frac{1}{\mu} b_2^{(k)} - \frac{1}{\mu(\mu + 2)} S_2^t c^{(k)} \]
\[ w_1^{(k+1)} = \text{soft}\left( u_1^{(k+1)} - d_1^{(k)}, \frac{\lambda_1}{2\mu} \right) \]
\[ w_2^{(k+1)} = \text{soft}\left( u_2^{(k+1)} - d_2^{(k)}, \frac{\lambda_2}{2\mu} \right) \]

where \( \text{soft}(x, T) \) is the soft-threshold rule with threshold \( T \),

\[ \text{soft}(x, T) = x \max(0, 1 - T/|x|). \]

Note: no matrix inverses; only forward and inverse transforms.
Figure 7: Reduction of objective function during the first 100 iterations. SALSA converges faster than ISTA.
A constant bandwidth and a constant Q-factor decomposition can have high coherence due to some analysis functions, from each decomposition, having similar frequency support. This can degrade the results of MCA in principle.
Two constant Q-factor decompositions with markedly different Q-factors will have low coherence because no analysis functions from the two decompositions will have similar frequency support. This is beneficial for the operation of MCA.
Small coherence between low and high Q-factor RADWTs

Figure 8: For reliable resonance-based decomposition, the inner product between the low-Q and high-Q wavelets should be small for all dilations and translations. The computation of the maximum inner product is simplified by assuming the wavelets are ideal band-pass functions and expressing the inner product in the frequency domain.

The inner products can be defined in the frequency domain,

\[ \rho(f_1, f_2) := \int \Psi_1(f) \Psi_2(f) \, df, \]

as a function of their center frequencies (equivalently, dilation).

The maximum value of the inner product, \( \rho(f_1, f_2) \), occurs when \( f_2 = f_1 \left(2 + 1/Q_1\right)/(2 + 1/Q_2) \) and is given by

\[ \rho_{\text{max}} = \sqrt{\frac{Q_1 + 1/2}{Q_2 + 1/2}}, \quad Q_2 > Q_1. \] (12)
Two constant bandwidth decompositions with markedly different bandwidths will also have low coherence and are therefore also suitable transform for MCA-based signal decomposition. This gives a bandwidth-based decomposition, rather than a resonance-based decomposition.
Example: Resonance-selective nonlinear band-pass filtering

Figure 9: LTI band-pass filtering. The test signal (a) consists of a sinusoidal pulse of frequency 0.1 cycles/sample and a transient. Band-pass filters 1 and 2 in (b) are tuned to the frequencies 0.07 and 0.10 cycles/second respectively. The output signals, obtained by filtering the test signal with each of the two band-pass filters, are shown in (c) and (d). The output of band-pass filter 1, illustrated in (c), contains oscillations due to the transient in the test signal. Moreover, the transient oscillations in (c) have a frequency of 0.07 Hz even though the test signal (a) contains no sustained oscillatory behavior at this frequency.
Figure 10: Resonance-based decomposition and band-pass filtering. When resonance-based analysis method is applied to the test signal in Fig. 9a, it yields the high- and low-resonance components illustrated in (a) and (b). The output signals, obtained by filtering the high-resonance component (a) with each of the two band-pass filters shown in Fig. 9b, are illustrated in (c) and (d). The transient oscillations in (c) are substantially reduced compared to Fig. 9c.
Example: Resonance-based decomposition of speech

Figure 11: Decomposition of a speech signal ("I’m") into high- and low-resonance components. The high-resonance component (b) contains the sustained oscillations present in the speech signal, while the low-resonance component (c) contains non-oscillatory transients. (The residual is not shown.)
Figure 12: Frequency spectra of the speech signal in Fig. 11 and of the extracted high- and low-resonance components. The spectra are computed using the 50 msec segment from 0.05 to 0.10 seconds. The energy of each resonance component is widely distributed in frequency and their frequency-spectra overlap.
Figure 13: Frequency decomposition of high-resonance component in Fig. 11. Reconstructing the high-resonance component from a few subbands of the high Q-factor RADWT at a time, yields an efficient AM/FM decomposition.
RECONSTRUCTION FROM ONE SINGLE SUBBAND (LOW Q-FACTOR RADWT)
Conclusion: Resonance-based signal decomposition

Low Q-factor RADWT used for sparse representation of the transient component.

High Q-factor RADWT used for sparse representation of the oscillatory (rhythmic) component.

Morphological component analysis (MCA) used to separate the two signal components.

- Oscillatory component not necessarily high-pass — may contain both low and high frequencies.
- Transient component not necessarily a low-pass signal — may contains sharp bumps and jumps.