

Variational Mode Decomposition with Missing Data

Guebin Choi^a · Hee-Seok Oh^{a,1} · Youngjo Lee^a · Donghoh Kim^b · Kyungsang Yu^c

^aDepartment of Statistics, Seoul National University

^bDepartment of Applied Mathematics, Sejong University

^cDepartment of Clinical Pharmacology and Therapeutics, Seoul National University

(Received March 2, 2015; Revised March 24, 2015; Accepted March 26, 2015)

Abstract

Dragomiretskiy and Zosso (2014) developed a new decomposition method, termed variational mode decomposition (VMD), which is efficient for handling the tone detection and separation of signals. However, VMD may be inefficient in the presence of missing data since it is based on a fast Fourier transform (FFT) algorithm. To overcome this problem, we propose a new approach based on a novel combination of VMD and hierarchical (or h)-likelihood method. The h -likelihood provides an effective imputation methodology for missing data when VMD decomposes the signal into several meaningful modes. A simulation study and real data analysis demonstrates that the proposed method can produce substantially effective results.

Keywords: Empirical mode decomposition, FFT, h -likelihood, missing data, variational mode decomposition.

1. 서론

신호(signal)란 시간 또는 공간에 따라서 값이 변화하는 물리적인 변수이다. 신호의 종류에는 온도, 전류, 속도, 음성, 이미지 등이 있으며 이러한 신호에 포함된 다양한 주파수를 분석하여 주파수별로 신호를 분리하고자 하는 연구가 지속적으로 진행되었다. Huang 등 (1998)은 특정 신호를 제거과정들을 통하여 고주파 함수부터 저주파 함수까지 순차적으로 추출하는 경험적모드분해법(empirical mode decomposition)을 제안하였다. 최근에 Dragomiretskiy와 Zosso (2014)는 경험적모드분해법의 단점을 보완한 변동모드분해법(variational mode decomposition; VMD)을 제시하였다.

변동모드분해법은 신호를 여러 개의 모드의 합으로 이루어져 있다고 가정하고, 각각의 모드를 분해하는 방법이다. 변동모드분해법은 경험적모드분해법에 비하여 주파수 탐색 및 분리(tone detection and tone separation)에 탁월한 성능을 보인다. 또한 고속푸리에변환(fast fourier transform)을 기반으로 한 알고리즘을 사용하여 상대적으로 경험적모드분해법보다 잡음에 민감하지 않다. 하지만 고속푸리에 변환은 기본적으로 동일한 시간간격(혹은 공간적 간격)으로 측정된 등간격 신호자료를 분석하기 위한 목적으로 설계된 변환이기 때문에, 특정 시간 혹은 공간에 결측이 있을 경우 자료를 적절히 분석하지

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No.20110030037).

¹Corresponding author: Department of Statistics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-747, Korea. E-mail: heeseok@stats.snu.ac.kr

못하는 단점이 있다. 본 논문에서는 자료에 결측치가 존재할 때 Lee와 Nelder (1996)의 다단계우도와 변동모드분해법을 결합하여 새로운 변동모드분해법을 제시하고자 한다. 참고로 Kim 등 (2006)은 결측치가 있을 경우 웨이블릿 임계화 기법을 적용하기 위하여, 다단계우도함수를 이용하여 결측치를 대체(imputation)하는 방법을 제시하였다.

본 논문의 전개내용은 다음과 같다. 2장에서 변동모드분해법에 대해 간략하게 소개하고, 3장에서는 다단계우도를 이용한 수정된 변동모드분해법을 제안한다. 4장에서는 모의실험을 통하여 제안된 방법에 대한 우수성을 평가하고, 5장에서는 실제관측자료를 분석하여 기존의 변동모드분해법과 수정된 변동모드분해법의 성능을 비교한다.

2. 변동모드분해법

변동모드분해법의 목적은 신호자료 f 를 여러 개의 내재모드함수(intrinsic mode function) u_k 로 분해하는 것이다. 여기에서 내재모드함수는 아래와 같이 정의한다.

정의 2.1 내재모드함수란

$$u_k(t) = A_k(t) \cos(\phi_k(t))$$

의 형태를 가지는 진폭 및 주파수변조신호들(AMPLITUDE-MODULATED-FREQUENCY-MODULATED(AM-FM) SIGNALS)이다. 여기에서 $\phi_k(t)$ 는 비감소함수(NON-DECREASING FUNCTION)이며, $A_k(t)$ 는 비음함수(NON-NEGATIVE FUNCTION)이다.

내재모드함수를 알기 위해서는 순간주파수함수(instantaneous frequency function)인 $\omega_k(t) := \phi'_k(t)$ 와 그에 해당하는 진폭함수 $A_k(t)$ 를 알아내는 것이 중요하다. 여기에서 $\phi'_k(t)$ 는 $\phi_k(t)$ 의 도함수를 의미한다. 따라서 변동모드분해법은 적절한 K 개의 주파수함수 $\omega_1, \omega_2, \dots, \omega_K$ 와 그에 대응하는 내재모드함수 $u_1(t), u_2(t), \dots, u_K(t)$ 를 추정하는 과정이라고 요약할 수 있다. 여기에서 신호를 구성하는 내재모드함수의 개수 K 는 알려져 있는 것으로 가정하자. 예를 들어

$$y = t + \cos(40\pi t) + \cos(80\pi t)$$

와 같은 신호를 관측했다면, 변동모드분해법은 각 내재모드함수의 주파수함수 $\omega_1 = 0, \omega_2 = 40\pi, \omega_3 = 80\pi$ 와 내재모드함수 $u_1(t) = t, u_2(t) = \cos(40\pi t), u_3(t) = \cos(80\pi t)$ 를 적절하게 찾아내는 것이 목적이다. Dragomireskiy와 Zosso (2014)는 아래와 같은 최적화과정

$$\min_{\{u_k\}, \{\omega_k\}} \left\{ \sum_{k=1}^K \left\| \partial t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \quad \text{s.t.} \quad \sum_{k=1}^K u_k(t) = f \quad (2.1)$$

을 통해 내재모드함수를 도출하는 방법을 제시하였다. 여기서 δ 는 디랙함수(Dirac function), $j^2 = -1$, $\|\cdot\|_2$ 은 L_2 거리, ∂t 는 t 에 대한 미분을 의미한다. $*$ 는 컨볼루션(convolution)을 의미한다. 즉, 두 함수 f, g 와 임의의 실수 τ 에 대하여 $(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau$ 를 의미한다. 또한 기술을 간단히 하기 위하여 $\{u_1(t), \dots, u_K(t)\}$ 를 $\{u_k\}$ 로, $\{\omega_1, \dots, \omega_K\}$ 를 $\{\omega_k\}$ 로 표기하였다. 식 (2.1)을 살펴보면 내재모드함수 $u_k(t)$ 에 단계적으로 3가지 변환이 적용된 것을 알 수 있다.

1. $u_k(t) \rightarrow \left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t)$.

이 과정은 $u_k(t)$ 를 양의 주파수만 가지는 신호로 변환하는 과정이다. $(\delta(t) + j/(\pi t)) * u_k(t)$ 를 푸리에 변환하면, $(1 + j \operatorname{sgn}(\omega))\hat{u}_k(\omega)$ 이 되고 이를 계산하면 쉽게 음의 주파수부분이 소거됨을 알 수 있

다. 여기에서 sgn 는 부호함수이다. 그리고 이러한 과정을 거쳐서 양의 주파수만 남겨진 신호를 해석적 신호라고 한다. 즉 $(\delta(t) + j/(\pi t)) * u_k(t)$ 는 $u_k(t)$ 의 해석적 신호이다.

일반적으로 실수신호(real-valued signal)는 양의 주파수와 음의 주파수를 모두 가지며, 주파수영역에서의 진폭이 $\omega = 0$ 에 대하여 대칭이다. 즉, 임의의 실수신호 $u_k(t)$ 에 대해서 $\|\hat{u}(\omega)\|_2^2 = \|\hat{u}(-\omega)\|_2^2$ 이 성립한다. 따라서 실수신호의 경우 양의 주파수성분에 해당하는 스펙트럼만을 고려하여 분석하여도 충분하다 (음의 주파수성분은 양의 주파수성분을 대칭시켜 얻어낼 수 있다.). 따라서 $u_k(t)$ 를 분석하는 대신에 $u_k(t)$ 의 해석적 신호를 분석하여도 분석에 아무런 지장이 없으며, 데이터를 반으로 줄임으로써 처리속도 역시 향상시킬 수 있다.

$$2. (\delta(t) + \frac{j}{\pi t}) * u_k(t) \rightarrow [(\delta(t) + \frac{j}{\pi t}) * u_k(t)] e^{-j\omega_k t}.$$

이 과정은 $u_k(t)$ 의 해석적 신호에 대한 주파수변조 과정이다. 만약에 이 과정에서 내재모드함수에 대한 주파수함수 ω_k 를 적절하게 선택하였다면, 주파수가 변조된 내재모드함수의 해석적 신호는 주파수가 0에 근접하게 된다. 바꾸어 말하면 주파수가 0에 근접하도록 내재모드함수의 해석적 신호를 변조시키는 ω_k 를 선택한다면 각 내재모드함수의 주파수함수를 올바르게 선택할 수 있다. 따라서 주파수가 변조된 내재모드함수의 해석적 신호의 주파수가 얼마나 0에 가까운지를 측정하고, 그것을 최소화 하는 ω_k 를 선택함으로써 내재모드함수의 주파수함수를 추정할 수 있다. 이러한 주파수의 측정은 아래에서 제시된다.

$$3. [(\delta(t) + \frac{j}{\pi t}) * u_k(t)] e^{-j\omega_k t} \rightarrow \|\partial t [(\delta(t) + \frac{j}{\pi t}) * u_k(t)] e^{-j\omega_k t}\|_2^2.$$

이 과정에서 $\|\partial t [(\delta(t) + j/(\pi t)) * u_k(t)] e^{-j\omega_k t}\|_2^2$ 이 주파수가 변조된 내재모드함수의 해석적 신호의 주파수가 얼마나 0에 가까운지를 측정하는 측도이다. 결국 주파수가 변조된 내재모드함수의 해석적 신호 $[(\delta(t) + j/(\pi t)) * u_k(t)] e^{-j\omega_k t}$ 의 시간에 대한 변화량을 측정하는 측도인데 여기에서 적절한 ω_k 를 선택하였다면, 이 측도의 값은 0이 된다.

따라서 $\sum_{k=1}^K u_k(t) = f$ 라는 제약조건하에서

$$\sum_{k=1}^K \left\| \partial t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2$$

를 최소화하는 $\omega_1, \omega_2, \dots, \omega_K$ 와 $u_1(t), u_2(t), \dots, u_K(t)$ 를 구하는 것은 각 내재모드함수의 적절한 중심 주파수 $\{\omega_k\}$ 를 구하고 $\{u_k\}$ 를 $\sum_k u_k(t) = f$ 라는 제약조건하에서 구한다는 의미가 된다.

여기에서 augmented Lagrangian method를 사용하면 수식 (2.1)을 최소화하는 문제는 다음의 안장점(saddle point)를 구하는 문제와 동치가 됨을 알 수 있다 (Bertsekas, 1976; Nocedal과 Wright, 2006).

$$\begin{aligned} \mathcal{L}(\{u_k\}, \{\omega_k\}, \lambda) := & \alpha \sum_{k=1}^K \left\| \partial t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\ & + \left\| f(t) - \sum_{k=1}^K u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^K u_k(t) \right\rangle. \end{aligned} \quad (2.2)$$

여기에서 $\langle a, b \rangle$ 는 a 와 b 의 내적을 의미한다. Alternate direction method of multiplier(ADMM) (Hestenes, 1969; Rockafellar, 1973; Bertsekas, 1982)을 이용하면 아래와 같은 알고리즘으로 수식 (2.2)의 안장점을 구할 수 있다. 알고리즘상에서는 표현을 간단하기 위해서 $u_k(t)$ 를 u_k 로 표기하였다.

알고리즘 1 : ADMM 최적화 기법을 적용한 변동모드분해법 알고리즘

Initialize $\{u_k^1\}, \{\omega_k^1\}, \lambda^1$. $n \leftarrow 0$.

repeat

$n \leftarrow n + 1$

for $k = 1 : K$ **do**

Update u_k :

$u_k^{n+1} \leftarrow \operatorname{argmin}_{u_k} \mathcal{L} \left(\{u_{i < k}^{n+1}\}, \{u_{i \geq k}^{n+1}\}, \{\omega_i^n\}, \lambda^n \right)$

end for

for $k = 1 : K$ **do**

Update ω_k :

$\omega_k^{n+1} \leftarrow \operatorname{argmin}_{\omega_k} \mathcal{L} \left(\{u_i^{n+1}\}, \{\omega_{i < k}^{n+1}\}, \{\omega_{i \geq k}^n\}, \lambda^n \right)$

end for

Dual ascent : $\lambda^{n+1} \leftarrow \lambda^n + \tau \left(f - \sum_k u_k^{n+1} \right)$

until convergence : $\sum_k \frac{\|u_k^{n+1} - u_k^n\|_2^2}{\|u_k^n\|_2^2} < \epsilon$.

Dragomiretskiy와 Zosso (2014)에 따르면 위의 알고리즘은 아래의 알고리즘으로 간략화 될 수 있다.

알고리즘 2 : 변동모드분해 알고리즘

Initialize $\{\hat{u}_k^1\}, \{\omega_k^1\}, \hat{\lambda}^1$. $n \leftarrow 0$.

repeat

$n \leftarrow n + 1$

for $k = 1 : K$ **do**

Update \hat{u}_k for all $\omega \geq 0$:

$u_k^{n+1} \leftarrow \frac{\hat{f}(\omega) - \sum_{i < k} \hat{u}_i^{n+1}(\omega) - \sum_{i \geq k} \hat{u}_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2}$

end for

for $k = 1 : K$ **do**

Update ω_k :

$\omega_k^{n+1} \leftarrow \frac{\int_0^\infty \omega |\hat{u}_k^{n+1}(\omega)|_2^2 d\omega}{\int_0^\infty |\hat{u}_k^{n+1}(\omega)|_2^2 d\omega}$

end for

Dual ascent : $\hat{\lambda}^{n+1} \leftarrow \hat{\lambda}^n + \tau \left(\hat{f} - \sum_k \hat{u}_k^{n+1} \right)$

until convergence : $\sum_k \frac{\|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2}{\|\hat{u}_k^n\|_2^2} < \epsilon$.

여기에서 \hat{f} 는 f 의 푸리에변환을 의미한다. 비록 원래의 신호가 연속적이라고 하더라도 실제로 관측하는 신호는 이산적인 신호이다. 따라서 변동모드분해법의 실질적인 알고리즘은 등간격으로 관측된 신호 f 에 대한 고속푸리에변환을 기반으로 한다. 하지만 경우에 따라 분석하고자 하는 신호 f 를 결측치

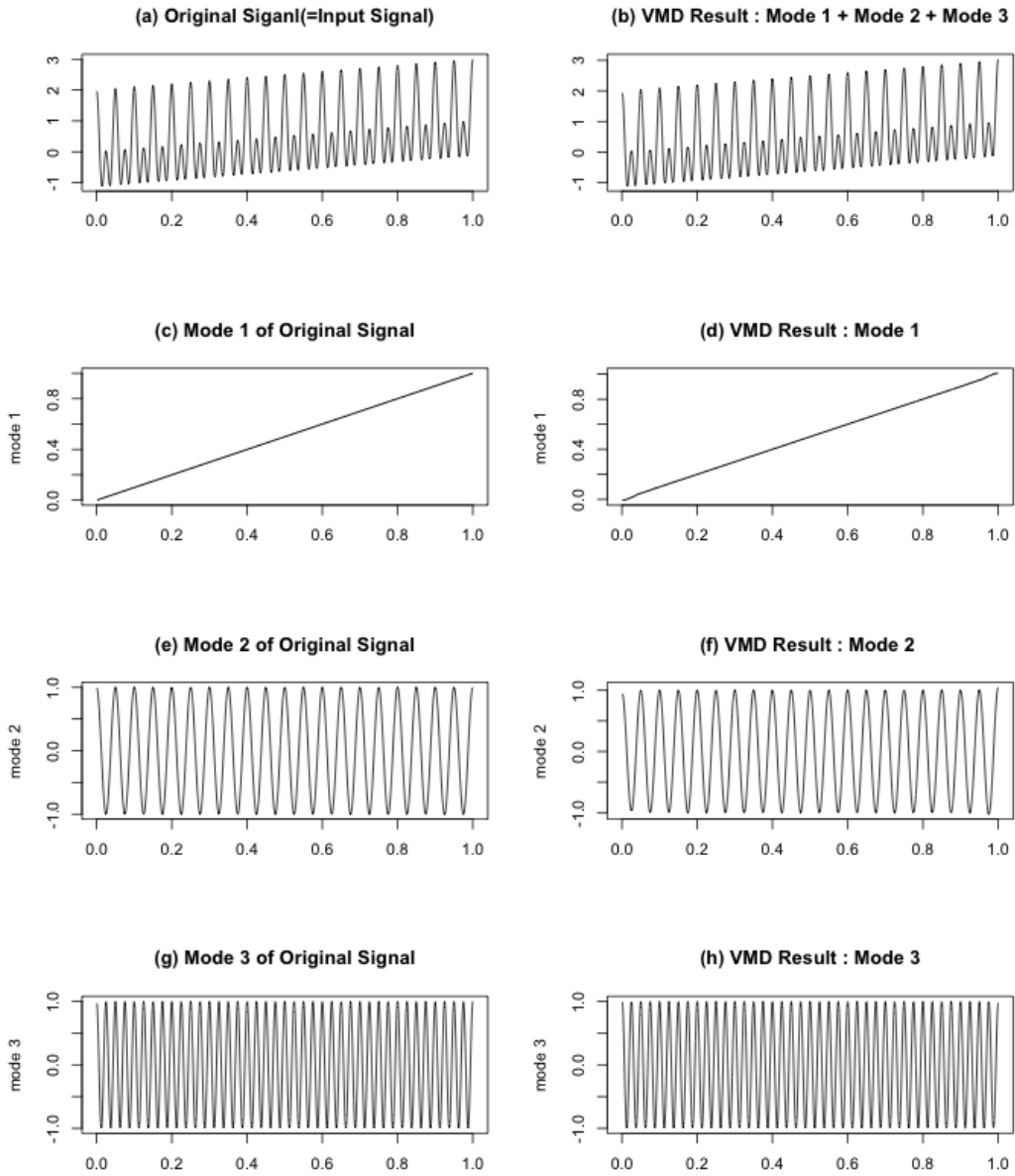


Figure 2.1. (a) Original signal *i.e.*, $y = t + \cos(40\pi t) + \cos(80\pi t)$. (c), (e), (g) Modes of original signal. (b) VMD reconstruction. (d), (f), (h) VMD modes.

등으로 인해 등간적으로 측정하지 못할 경우도 존재한다. 이러한 경우 고속푸리에변환과정을 기반으로 한 변동모드분해법 알고리즘에 문제가 생길 수 있다. Figure 2.1과 Figure 2.2은 이와 같은 문제를 보여주는 예이다. Figure 2.1은 정상적으로 관측된 신호에 대한 변동모드분해법의 분해결과이고, Figure

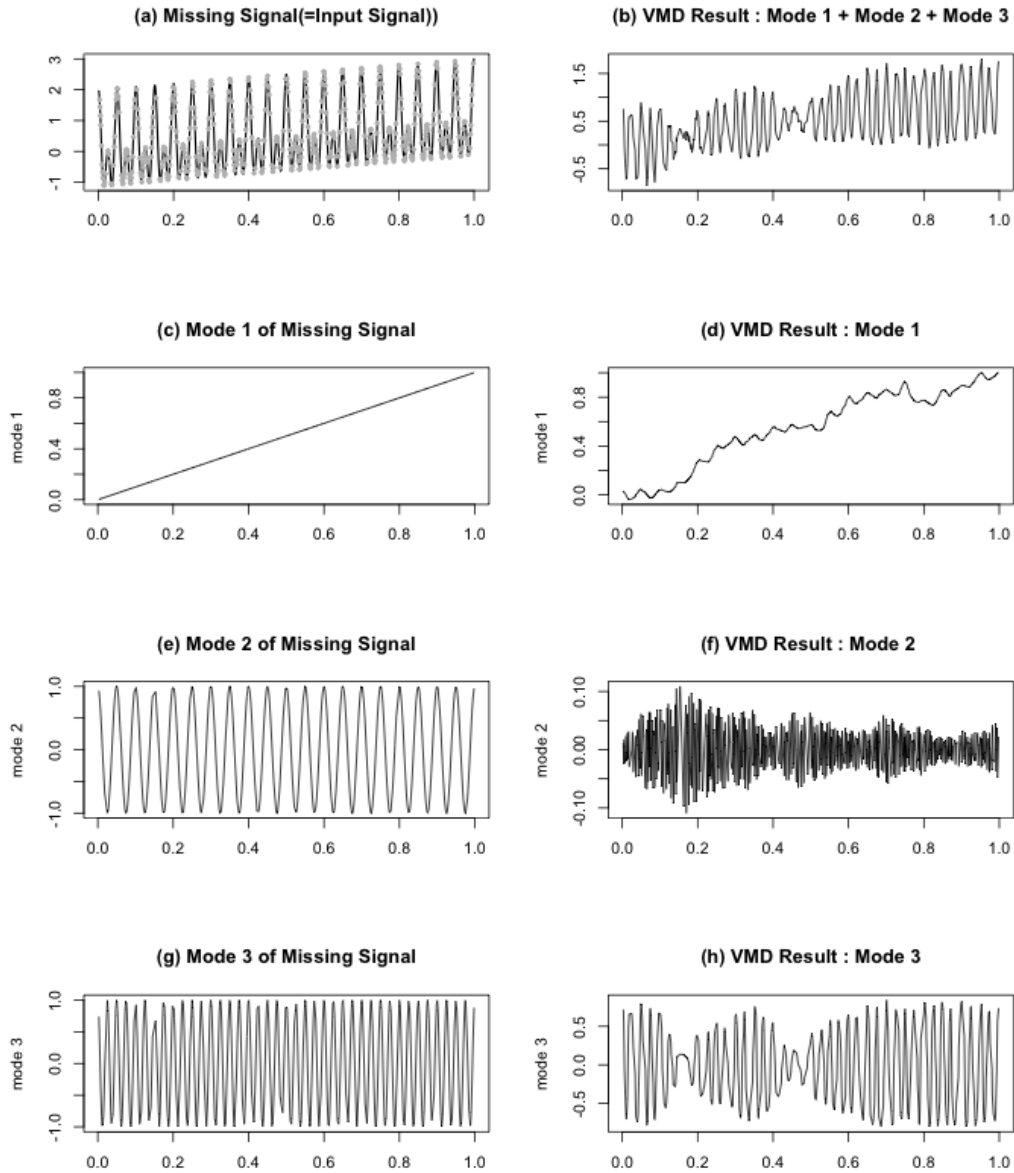


Figure 2.2. (a) Missing signal of $y = t + \cos(40\pi t) + \cos(80\pi t)$ with missing rate 40%. Dots denote missing value. (c), (e), (g) Modes of the missing signal. (b) VMD reconstruction. (d), (f), (h) VMD modes

2.2은 40%의 자료가 결측된 신호에 대한 변동모드분해법의 결과이다. 보는 것 처럼 결측치가 존재할 경우 기존의 변동모드분해법은 내재모드함수를 제대로 분리해 내지 못하는 것을 알 수 있다. 여기에서 결측치는 전체 관측치에서 비복원 단순추출법을 이용하여 추출하였으며, 향후 논문에 언급된 모든 결측치 추출법 역시 이와 동일하게 수행하였다.

3. 수정된 변동모드분해법

제안하고자 하는 수정된 변동모드분해법은 결측치를 대체(imputation)하는 과정과 기존의 변동모드분해법을 적절하게 결합하여 이루어진다.

원자료 y_1, \dots, y_n 를 \mathbf{y}_{com} 으로, m 개의 관측된 자료 y_1, \dots, y_m 를 \mathbf{y}_{obs} 로 그리고 나머지 $(n - m)$ 개의 결측된 자료를 \mathbf{y}_{mis} 로 각각 표기하자. 원자료 \mathbf{y}_{com} 은 정규분포 $N(\boldsymbol{\mu}, \sigma^2 I)$ 을 따른다고 가정하자. 여기에서 평균 μ_i 를 푸리에 기저 \mathbf{F} 를 이용하여 $\mu_i = (\mathbf{F}\boldsymbol{\theta})_i$ 로 표현할 수 있다. $\boldsymbol{\theta}$ 는 푸리에 계수를 나타낸다. Lee와 Nelder (1996)은 결측자료의 효과적인 대체를 위해 다음의 로그 다단계우도함수를 최대화할 것을 제안하였다.

$$h = h(\boldsymbol{\theta}, \sigma^2, \mathbf{y}_{mis}; \mathbf{y}_{obs}) = \ell_{com}(\boldsymbol{\theta}, \sigma^2; \mathbf{y}_{com}) = \ell_{obs}(\boldsymbol{\theta}, \sigma^2; \mathbf{y}_{obs}) + \ell_{mis}(\boldsymbol{\theta}, \sigma^2; \mathbf{y}_{mis}), \quad (3.1)$$

여기에서 ℓ_{obs} 는 관측된 자료 \mathbf{y}_{obs} 에 대한 로그우도함수이고, ℓ_{mis} 는 결측된 자료 \mathbf{y}_{mis} 에 대한 로그우도함수이다. 정규분포하에서 위의 로그 다단계우도함수를 다음과 같다.

$$h = -\frac{1}{2\sigma^2} \sum_{i=1}^m (y_i - \mu_i)^2 - \frac{m}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=m+1}^n (y_i - \mu_i)^2 - \frac{n-m}{2} \log \sigma^2.$$

따라서 \mathbf{y}_{mis} 를 대체하기 위해서는

$$\frac{\partial h}{\partial y_i} = -(y_i - \mu_i) = 0, \quad i = m+1, \dots, n$$

에 의해 $\hat{y}_{mis,i} = \mu_i$, $i = m+1, \dots, n$ 가 된다. \mathbf{y}_{mis} 를 대체하여 등간격으로 관측된 자료형태를 이루면, 별점로그우도함수 \mathcal{P}_{com} 를 최대화 함으로서 평균 μ_i 에 대한 적절한 푸리에추정량을 구할 수 있다.

$$\mathcal{P}_{com} \Big|_{\mathbf{y}_{mis}=\hat{\mathbf{y}}_{mis}} = (\mathcal{H}_{obs} + \mathcal{H}_{mis}) \Big|_{\mathbf{y}_{mis}=\hat{\mathbf{y}}_{mis}} - \lambda q(\boldsymbol{\theta}), \quad (3.2)$$

여기에서 $\mathcal{H}_{obs} = -(1/2\sigma^2) \sum_{i=1}^m (y_i - \mu_i)^2$, $\mathcal{H}_{mis} = -(1/2\sigma^2) \sum_{i=m+1}^n (y_i - \mu_i)^2$ 이며, $\lambda \geq 0$ 는 조율모수(tuning parameter), $q(\boldsymbol{\theta})$ 는 벌점함수이다. 따라서 식 (3.2)를 최대화하는 것은 아래 식을 최소화하는 것과 동치가 된다.

$$S(\boldsymbol{\theta}) = \sum_{i=1}^n [y_i - \mu_i]^2 \Big|_{\mathbf{y}_{mis}=\hat{\mathbf{y}}_{mis}} + \lambda q(\boldsymbol{\theta}). \quad (3.3)$$

위의 과정과 변동모드분해법을 결합하여 수정된 알고리즘을 다음과 같이 제안한다.

알고리즘 3 : 수정된 변동모드분해법 알고리즘

1. $\hat{\mathbf{y}}_{mis}^0$ 에 초기값을 대입한다.
2. 수렴할 때까지 아래의 과정을 반복한다.
 - (a) 결측치를 대체한 자료 $\mathbf{y}_{new}^{L-1} = (\mathbf{y}_{obs}, \hat{\mathbf{y}}_{mis}^{L-1})$ 에 대해 식 (3.3)을 최소화하여 푸리에추정량 $\hat{\boldsymbol{\mu}}^{L-1}$ 를 구한다.
 - (b) $\hat{\boldsymbol{\mu}}^{L-1}$ 에 대해 식 (3.1)를 최대화하여 $\hat{\mathbf{y}}_{mis}^L$ 를 대체한다.
 - (c) $\mathbf{y}_{new}^L = (\mathbf{y}_{obs}, \hat{\mathbf{y}}_{mis}^L)$ 에 변동모드분해법을 적용하여 내재모드함수 $\{u_k^L\}$ 추정한다.
3. 수렴된 $\{u_k^L\}$ 를 최종 내재모드함수의 추정량으로 사용한다.

4. 모의실험

4.1. 수정된 변동모드분해법과 기존 변동모드분해법의 비교

우선 모의자료 $y = t + \cos(40\pi t) + \cos(80\pi t)$ 를 고려하여 보자. 이 신호에 40% 결측치가 존재할 경우 기존 변동모드분해법의 분해결과는 Figure 2.2에 제시하였다. Figure 4.1은 3장에서 제안한 알고리즘에 따라 결측치가 있는 신호자료를 분해한 결과이다. Figure 4.1에서 보는 것처럼 효과적인 분해결과를 얻을 수 있다. 참고로 벌점함수는 $q(\theta) = \|\theta\|_2^2$ 와 같이 하였으며, λ 는 교차분석법(cross-validation)에 의해 선택되었다. 또한 이후의 모든 모의실험과 자료분석에서 역시 동일한 벌점함수와 조율모수선택법을 적용하였다. 수치적인 비교를 위해서 $MSE (= (1/n)\sum_{i=1}^n (u(x_i) - \hat{u}(x_i))^2)$ 를 계산한 결과는 첫 번째 내재모드함수에서 0.023, 두 번째 내재모드함수에서 0.021, 세 번째 내재모드함수에서 0.030이다. 반면 기존 방법에 의한 MSE값인 0.0086, 0.467, 그리고 0.495으로 2, 3번째 내재모드함수의 분해과정이 현저하게 개선되었음을 알 수 있다.

4.2. 다양한 결측률에 따른 비교

결측률에 따라 수정된 방법의 효과를 파악하고자 한다. 모의자료 $y = t + \cos(40\pi t) + \cos(80\pi t)$ 에 결측률을 10%, 20%, 30%, 40%, 50%, 60%, 70%로 변화시켜 기존 방법과 수정된 방법의 우수성을 비교하고자 한다. 결측률에 따라 100번씩 모의실험을 수행하였고, Figure 4.2는 내재모드함수에 대한 MSE값의 상자그림 결과이다. 그림에서 좌측의 (a), (c), (e)는 기존 변동모드분해법을 사용하여 얻은 결과이고, 우측의 (b), (d), (f)는 수정된 변동모드분해법을 이용하여 얻은 결과이다. 분석결과는 다음과 같이 정리할 수 있다.

1. 전체적으로 수정된 변동모드분해법에 의해 산출된 내재모드함수가 기존의 변동모드분해법에 의해 산출된 내재모드함수보다 더 적은 MSE를 가진다.
2. 기존 변동모드분해법은 고주파로 갈수록 결측률에 따라 MSE의 값이 더욱 민감하게 반응한다. 즉 고주파에 해당하는 내재모드함수의 경우 결측치가 많아질수록 내재모드함수의 추정이 매우 어렵게 된다. 이는 고주파의 경우 결측률이 높아지면 제대로 된 파형을 알기 어렵고, 마치 저주파처럼 보이게 되는 현상(에일리어싱(aliasing)) 때문이다. 특히 세 번째 내재모드함수의 경우 10% 결측치만 있을 경우에도 원래의 내재모드함수를 거의 추정하지 못한다. 이는 자료가 등간격으로 측정되지 않은 문제와 디지털 신호를 아날로그 신호로 복원하는데 필요한 최소한의 샘플수가 확보되지 않은 문제가 혼합되어 나타나는 현상이다.
3. 수정된 변동모드분해법은 결측률이 50% 한도 내에서 기존 방법보다 고주파에 해당하는 내재모드함수를 더욱 효과적으로 분해한다. 기존 변동모드분해법은 고주파로 갈수록 결측률에 따라 내재모드함수의 추출이 어려운데 반해, 수정된 방법은 고주파일 경우에도 내재모드함수를 적절히 추출하기 때문에 해석할 수 있다.

4.3. 잡음이 포함된 신호를 이용한 비교

본 절에서는 신호가 잡음에 의해 오염된 경우 수정된 변동모드분해법과 기존 변동모드분해법을 비교분석하고자 한다. 모의 신호는 다음과 같다.

$$y = \cos(10\pi t + 10\pi t^2) + \begin{cases} \cos(60\pi t), & t \leq 0.5, \\ \cos(80\pi t), & t > 0.5 \end{cases} + \epsilon, \quad \epsilon \sim N(0, \sigma^2)$$

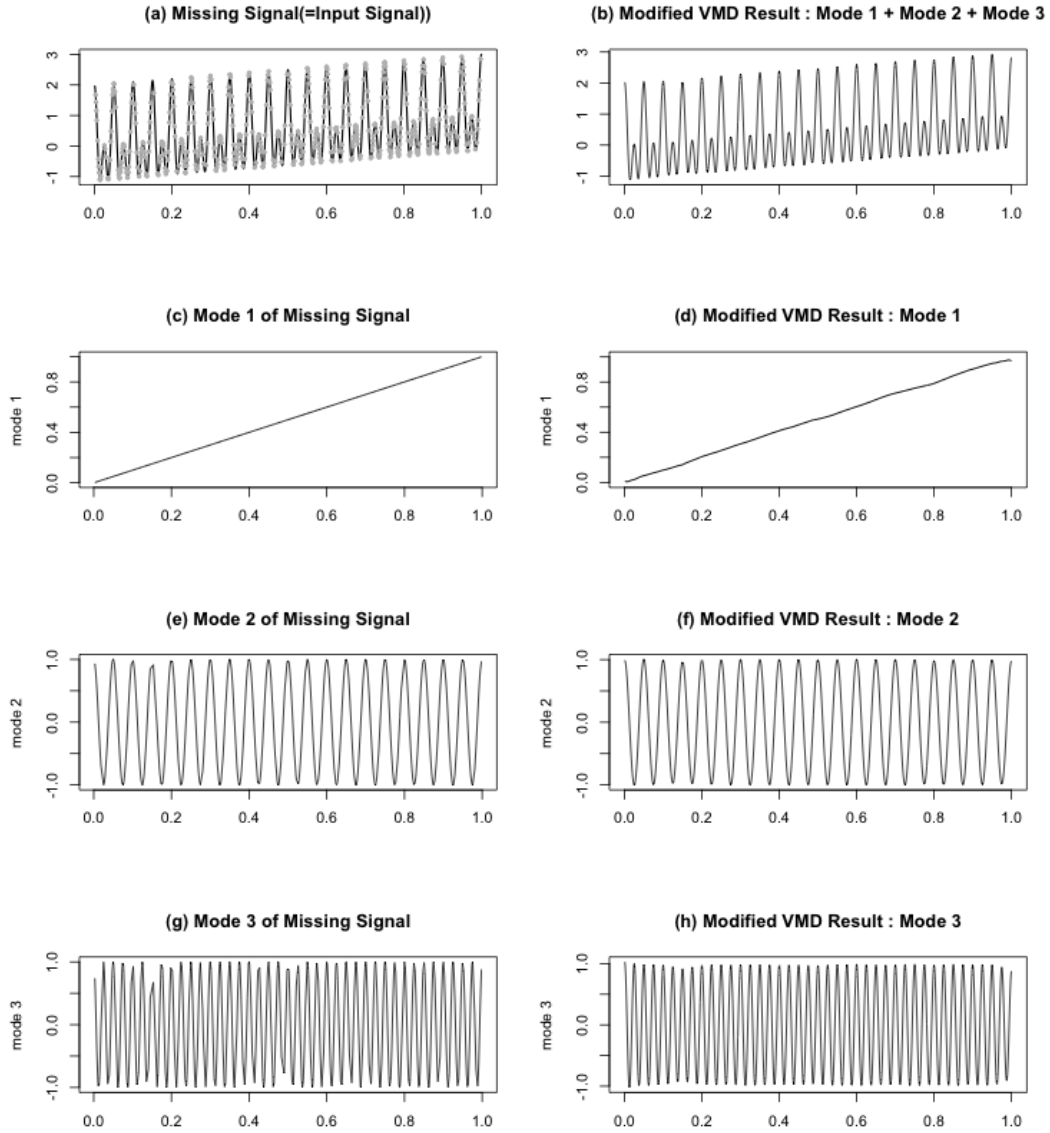


Figure 4.1. (a) Missing signal of $y = t + \cos(40\pi t) + \cos(80\pi t)$ with missing rate 40%. Dots denote missing value. (c), (e), (g) Modes of missing signal. (b) Modified VMD reconstruction. (d), (f), (h) Modified VMD modes.

위의 모의신호는 두 가지 내재모드함수 $u_1(t), u_2(t)$

$$u_1(t) = \cos(10\pi t + 10\pi t^2),$$

$$u_2(t) = \begin{cases} \cos(60\pi t), & t \leq 0.5, \\ \cos(80\pi t), & t > 0.5 \end{cases}$$

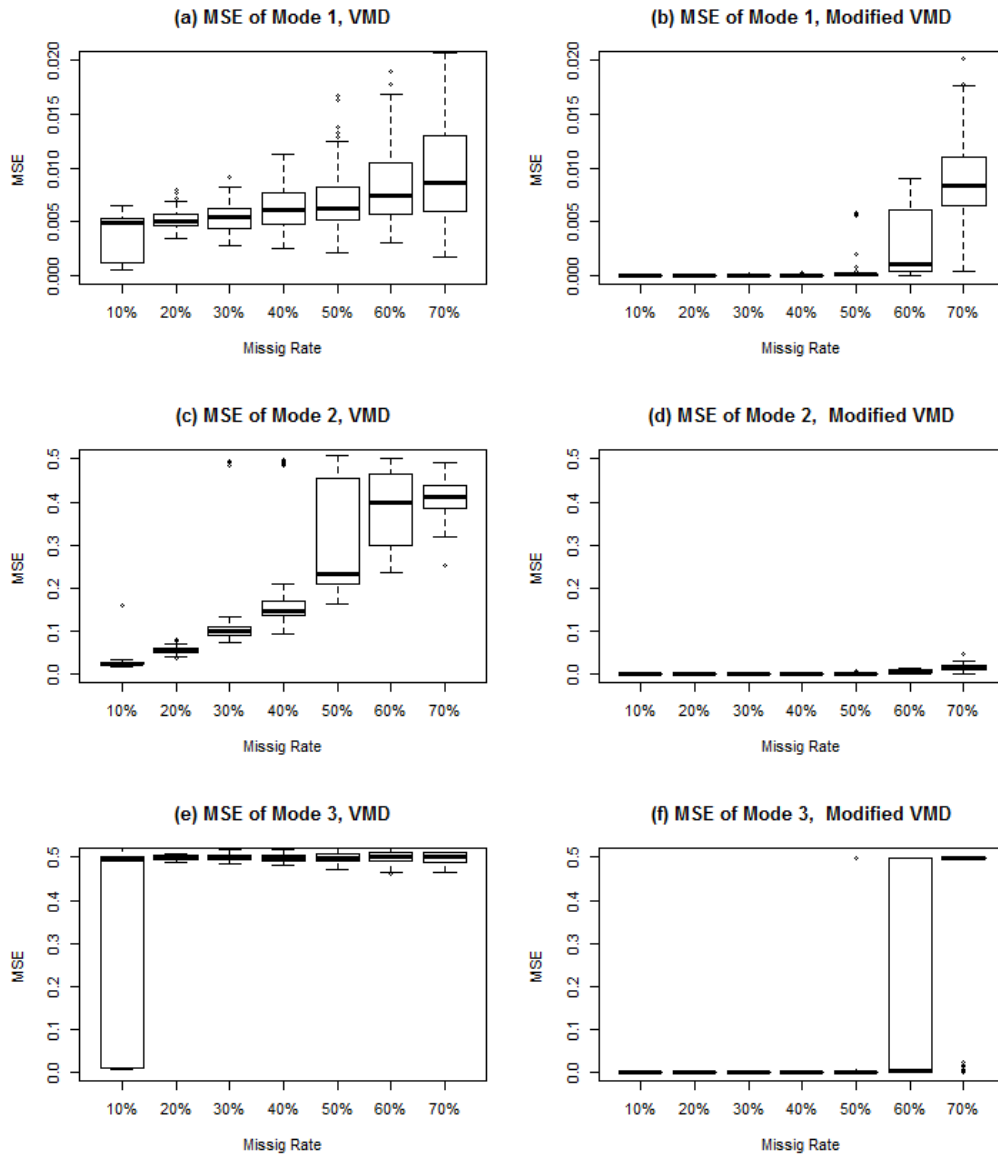


Figure 4.2. (a), (c), (d) Boxplots of MSE values by VMD reconstructions. (b), (d), (f) Boxplots of MSE values by modified VMD reconstructions

와 잡음의 합으로 구성되어 있다. Figure 4.3는 SNR = 5이고 결측률이 50%인 경우 변동모드분해법과 수정된 변동모드분해법으로 각각 분해한 결과이다. 위의 모의신호의 경우 실제 내재모드함수는 2개이지만 모의실험에서는 잡음을 반영하기 위해 내재모드함수를 3개로 설정하고 변동모드분석법을 적용하였다. Figure 4.3에서, (a)는 원래 신호에 SNR = 5의 잡음을 반영하고, 50% 결측률을 적용한 신호이다. (d)와 (g)는 각각 내재모드함수 $u_1(t)$ 와 $u_2(t)$ 를 의미한다. 중간열의 (b), (e), (h)는 기존의 변동

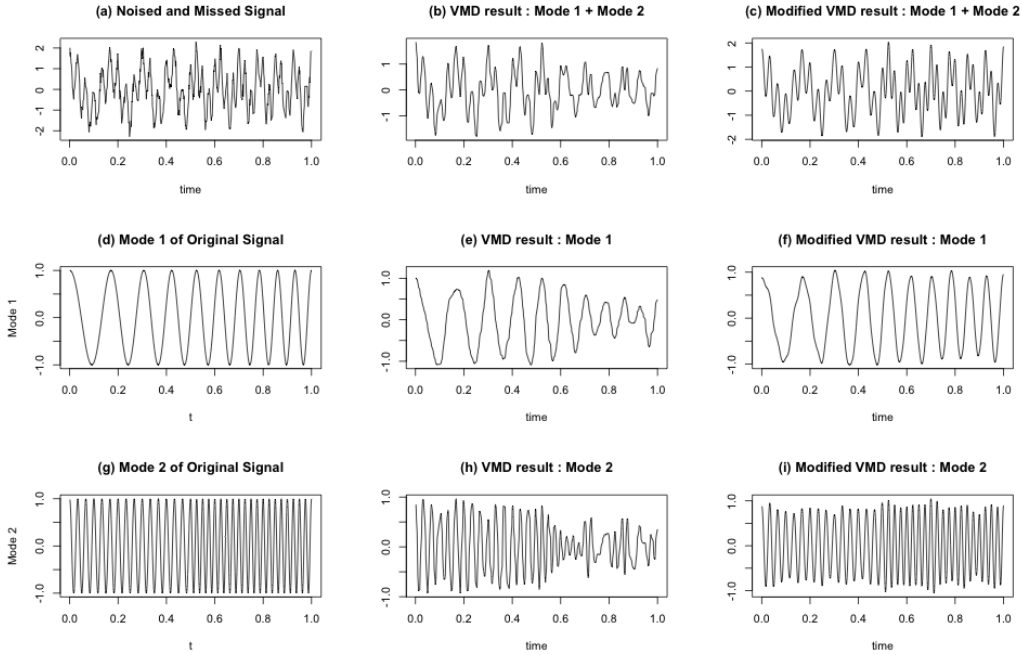


Figure 4.3. (a) Noisy signal (missing rate = 50%, SNR = 5). (d), (g) Modes of the noisy signal. (b) VMD reconstruction. (e), (h) VMD modes. (c) Modified-VMD reconstruction. (f), (i) Modified VMD modes.

모드분해법을 이용하여 얻은 결과이고, 우측열에 있는 (c), (f), (i)는 수정된 변동모드분해법을 이용하여 얻은 결과이다. 여기에서 신호를 복원하기 위한 (b)와 (c)에서는 잡음이 포함된 3번째 내재모드함수는 제외하고 나머지 내재모드함수들의 합으로 구성하였다. 분석결과 수정된 변동모드분해법이 기존의 변동모드분해법보다 잡음에 대해서 강건함을 알 수 있다. 하지만 기존의 변동모드분해법으로는 $t > 0.5$ 인 부분에서 복원된 신호가 상당히 왜곡되고 있다는 것을 알 수 있다. 이는 4.2절에서 살펴보았듯이 고주파로 갈수록 결측이 있는 자료의 복원이 어렵기 때문이라고 해석할 수 있다. 특히 두 번째 모드

$$u_2 = \begin{cases} \cos(60\pi t), & t \leq 0.5, \\ \cos(80\pi t), & t > 0.5 \end{cases}$$

의 경우 시점 $t = 0.5$ 을 기점으로 갑자기 고주파의 신호로 바뀌는데, 이로 인하여 $t > 0.5$ 인 부분에서는 신호의 분해가 제대로 이루어지지 않고 있다는 것을 Figure (h)에서 확인할 수 있다. 또한 u_1 의 경우도 (e)에서 볼 수 있듯이 고주파로 갈수록 신호의 분해가 왜곡되어 나타나고 있다는 것을 확인할 수 있었다. 반면에 수정된 변동모드분해법은 잡음이 포함된 신호에 대해서도 기본적으로 기존의 변동모드분해법에 비해 우수한 성능을 보여주며, 특히 고주파에 해당하는 내재모드함수의 경우 그 성능이 크게 개선되었음을 알 수 있다.

다양한 신호대잡음비(signal-to-noise ratio; SNR)를 이용하여 수정된 변동모드분해법과 기존 변동모드분해법을 비교분석하고자 잡음의 비율을 SNR = 9, 8, 7, 6, 5 순으로 다양화하며 MSE를 비교하는 모의 실험을 수행하였다. 앞의 절과 마찬가지로 100번의 모의실험 후 MSE값을 Figure 4.4의 상자그림으로 나타내었다. 전체적으로 수정된 변동모드분해법이 기존의 변동모드분해법보다 더 좋은 효율을 보이는 것을 확인할 수 있다.

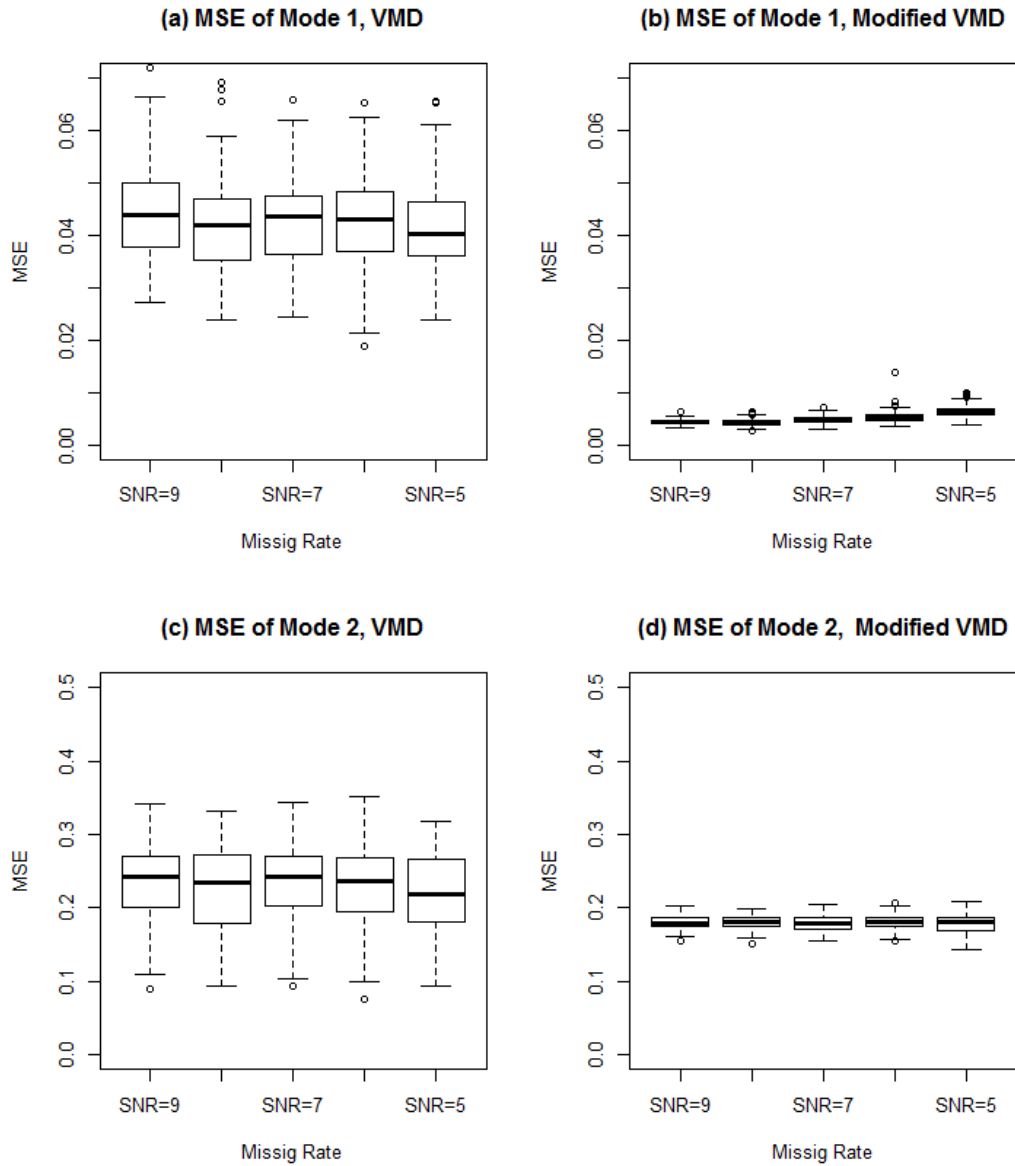


Figure 4.4. (a), (c) Boxplots of MSE values by VMD reconstructions. (b), (d) Boxplots of MSE values by modified VMD reconstructions

5. 실제 자료의 분석

본 장에서는 수정된 변동모드분해법에 따라 대왕고래(blue whale)의 울음소리를 적절한 내재모드함수로 분해하고 그 결과를 해석하고자 한다. 대왕고래는 두 가지 타입의 소리를 내는 것으로 알려져 있다. 하나는 “trill”이라는 A유형의 울음소리이고, 다른 하나는 “moan”이라는 B유형의 울음소리이다. A유

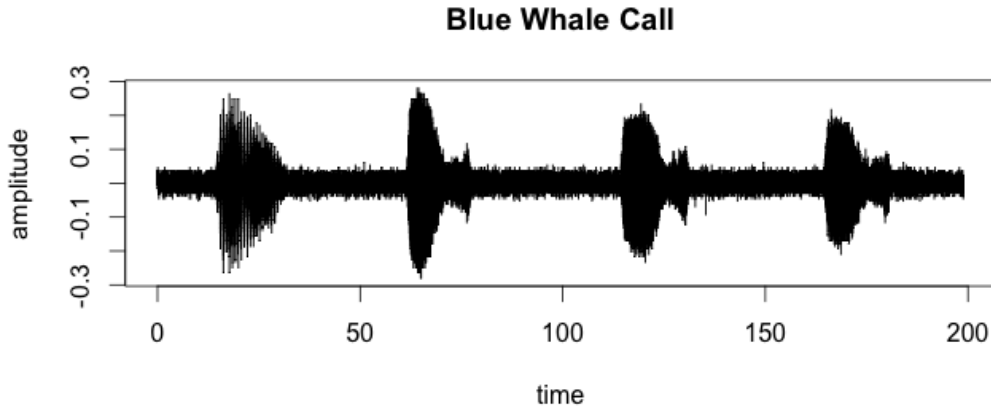


Figure 5.1. Blue whale call.

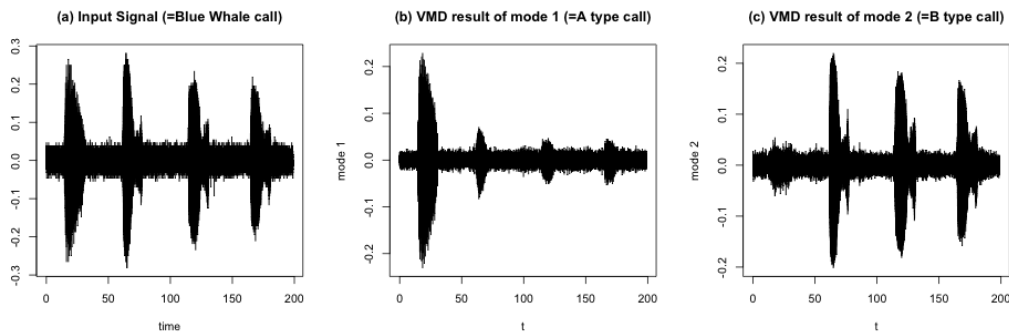


Figure 5.2. VMD results of blue whale call. (a) Input signal. (b) First VMD mode (A type call). (c) Second VMD mode (B type call).

형의 울음소리는 높은 주파수를 포함하며 마치 새의 지저귐 혹은 마우스의 클릭소리를 연상케 하는 울음소리이다. 반면에 B유형의 울음소리는 상대적으로 낮은 주파수를 포함하며 뱃고동 소리를 연상하게 한다.

Figure 5.1은 대왕고래의 울음소리에 대한 신호를 나타낸 자료이다. 자료는 매트랩 공식홈페이지에 있는 고속푸리에변환 예제에 포함된 자료를 사용하였다 (<http://kr.mathworks.com/help/matlab/math/fast-fourier-transform-fft.html>). Figure 5.1의 신호는 크게 네 가지의 소리로 구성되어 있다. 첫 번째가 A유형의 소리이고 두 번째부터 네 번째가 B유형의 소리이다. 따라서 Figure 5.1의 패턴은 “A”, “B”, “B”, “B” 유형의 소리로 구성된 패턴을 보인다. 변동모드분해법을 사용하여 위의 신호에 대한 내재모드함수를 추출해보자. A유형의 소리와 B유형의 소리로 분해하기 위해 모드의 수를 $K = 2$ 로 설정한다.

Figure 5.2는 실제자료를 변동모드분해법으로 분해한 결과이다. 첫 번째 내재모드함수에 A유형의 울음소리가 추출되었고 두 번째 내재모드함수에 B유형의 울음소리가 추출되어 대왕고래의 울음소리에 대한 신호를 유형별로 적절히 분류하였다고 보인다.

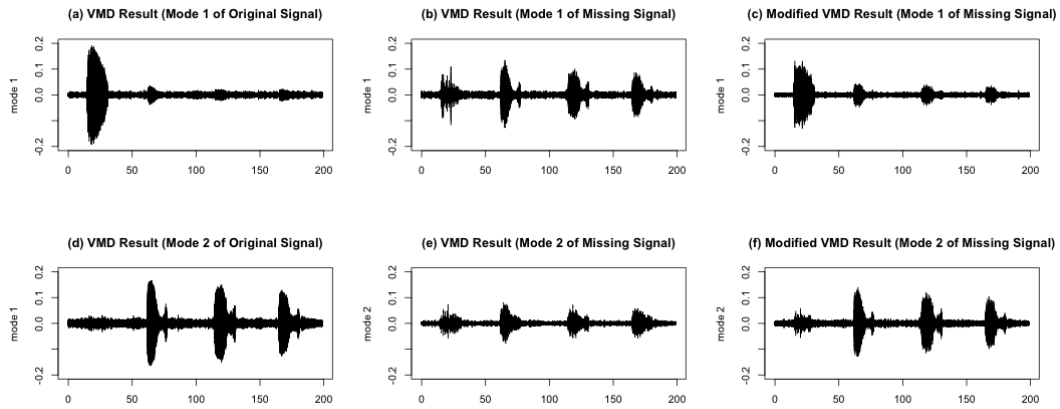


Figure 5.3. (a), (d) VMD modes of the original signal. (b), (e) VMD modes of the missing signal. (c), (f) Modified VMD modes of the missing signal.

다음에는 결측이 존재하는 실제데이터에 대하여 변동모드분해법과 수정된 변동모드분해법의 성능을 비교하기 위해 고래신호에서 40%의 결측치를 발생시켜 분석하여 보았다. Figure 5.3의 좌측열은 Figure 5.2의 결측이 없는 원래신호에 대하여 기존의 변동모드분해법을 산출한 내재모드함수이다. 결측이 없는 경우 앞서 살펴본 바와 같이 변동모드분해법이 고래의 울음소리를 두 가지 유형으로 잘 추출하고 있음을 확인할 수 있다. 가운데 열은 40%의 결측치를 임의로 발생시킨 신호자료에 대하여 기존의 변동모드분해법을 적용한 결과이다. 추출결과 두 내재모드함수에 모두 비슷한 패턴의 신호가 포함된 것으로 확인되었다. 이는 결측치로 인하여 기존의 변동모드분해법이 두 가지 유형의 고래울음소리를 제대로 분류하지 못한다고 해석할 수 있다. 마지막 우측열은 동일한 결측 신호자료에 대하여 수정된 변동모드분해법을 이용하여 분석한 결과이다. 기존의 변동모드분해법과 달리 결측치가 있는 경우에도 두 가지 유형의 고래울음소리를 적절하게 분류하고 있음을 확인할 수 있다.

6. 토의 및 결론

기존의 변동모드분해법은 결측치가 있는 신호자료에 적합하지 않다는 단점을 가지고 있다. 본 연구에서는 이러한 단점을 보완하여 기존의 변동모드분해법에 다단계우도함수를 적절히 결합한 수정된 변동모드분해법을 제시하였다. 수정된 변동모드분해법은 기존의 방법에 비하여 결측된 자료구조에 대하여서도 적절한 내재모드함수를 추출하는 것을 모의실험을 통하여 확인하였다. 특히 고주파 성분에 해당하는 내재모드함수의 추출부분에서는 기존의 변동모드분해법에 비해 월등한 성능을 보였다. 또한 실제자료분석을 통하여서도 의미 있는 결과를 도출하였다. 결측된 자료구조에서 고래의 두 가지 형태의 울음소리를 기존의 변동모드분해법에 비하여 효과적으로 분리해 그 효율성을 입증하였다.

References

- Bertsekas, D. P. (1976). Multiplier methods: A survey, *Automatica*, **12**, 133–145.
- Bertsekas, D. P. (1982). Constrained optimization and Lagrange multiplier methods, *Computer Science and Applied Mathematics*.
- Dragomiretskiy, K. and Zosso, D. (2014). Variational mode decomposition, *IEEE Transactions on Signal Processing*, **62**, 531–544.

- Hestenes, M. R. (1969). Multiplier and gradient methods, *Journal of Optimization Theory and Applications*, **4**, 303–320.
- Huang, N. E., Shen, Z., Long, S. R., Wu, M. L., Shih, H. H., Zheng, Q., Yen, N. C., Tung, C. C. and Liu, H. H. (1998). The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis, *Proceeding of the Royal Society of London A*, **454**, 903–995.
- Kim, D., Lee, Y. and Oh, H.-S. (2006). Hierarchical-likelihood-based wavelet method for denoising signals with missing data, *IEEE Signal Processing Letters*, **13**, 361–364.
- Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion), *Journal of the Royal Statistical Society, Series B (Methodological)*, **58**, 619–678.
- Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*, 2nd ed Springer.
- Rockafellar, R. T. (1973). A dual approach to solving nonlinear programming problems by unconstrained optimization, *Mathematical Programming*, **5**, 303–320.

결측치가 있는 자료에서의 변동모드분해법

최규빈^a · 오희석^{a,1} · 이영조^a · 김동호^b · 유경상^c

^a서울대학교 통계학과, ^b세종대학교 수학과통계학부, ^c서울대학교 의과대학 임상약리학교실

(2015년 3월 2일 접수, 2015년 3월 24일 수정, 2015년 3월 26일 채택)

요약

최근에 Dragomiretskiy와 Zosso (2014)는 경험적모드분해의 단점을 보완하여 새로운 신호 분해방법인 변동모드분해법(Variational Mode Decomposition)을 고안하였다. 기본적으로 변동모드분해법은 경험적모드분해법에 비하여 주파수 탐색 및 분리(tone detection and tone separation)에 탁월한 성능을 보인다. 또한 고속푸리에변환을 기반으로 한 알고리즘을 사용하여 경험적모드분해법보다 잡음에 강건하다는 장점이 있다. 하지만 변동모드분해법은 결측 등으로 신호가 동일한 시간간격 혹은 공간적 간격으로 측정되지 않은 경우 제대로 동작하지 않는 단점이 있다. 이를 보완하기 위해서 본 논문에서는 변동모드분해법에 다단계우도함수를 조합하는 새로운 방법을 제안한다. 여기에서 다단계우도함수는 변동모드분해법이 신호를 적절한 내재모드함수로 분해하기 전에 결측치를 대체하는 효율적인 방법을 제시한다. 모의실험과 실제 자료의 사례연구를 통하여 변동모드분해법이 기존의 방법보다 더 효율적으로 신호를 분해한다는 것을 보일 것이다.

주요용어: 경험적모드분해법, 고속푸리에변환, 다단계우도함수, 결측자료, 변동모드분해법.

이 논문은 2014년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 20110030037).

¹교신저자: (151-747) 서울시 관악구 관악로 1, 서울대 통계학과. E-mail: heeseok@stats.snu.ac.kr