Null Space Pursuit: An Operator-based Approach to Adaptive Signal Separation

Peng Silong

(co-author: Hwang Wen-Liang)

silong.peng@ia.ac.cn

Institute of Automation, CAS

submitted to IEEE T. Signal Processing (AQ)

2009.12.23

Agenda

1. Motivation: Basic of signal separation
2. Null Space Pursuit
3. Experiments and discussion

Signal separation:

$$S = \sum S_k + R$$

S: given signal to be analyzed: complex and information mixed;

S_k: **SIMPLE** signals and/or physical meaningful;

R: reasonable residual.

What is SIMPLE?

Different methods give different answers, OR different answers give different methods.

Simpler problem: $S = U + V$, U is a simple signal, V is the residual.

Existing separation methods:

1. MAP (Maximal a posterior probability);

2. Sparsity based method: based on sparse representation;

3. Norm based method: based on new norm;

4. Empirical method: EMD etc.

5. Other methods: basis, frame, ICA ..., .

MAP

$$\{\hat{U}, \hat{V}\} = \arg\min_{U,V} - \log P(U, V|S) \tag{1}$$

under the constrain: $S = U + V$
By using Bayesian role and Lagrange method:

$$\{\hat{U}, \hat{V}\} = \arg\min_{U,V}\{-\log P_u(u, \theta_u) - \log P_v(v, \theta_v) + \lambda\varphi(S - U - V)\} \tag{2}$$

where U and V have probability of given form $P_u(u, \theta_u)$ and $P_v(v, \theta_v)$ respectively, U and V are independent each other.

SIMPLE means simple probability form.
Application: denoising ...

Sparsity based method

Basic Idea: one basis only good for arbitrary subset of all.

Two basis or dictionaries: Φ_u and Φ_v, such that U can be represented by Φ_u in sparser form than by Φ_v, V can be represented by Φ_v in sparser form than by Φ_u.

$$\{\hat{\alpha}_u, \hat{\alpha}_v\} = \arg\min_{\alpha_u, \alpha_v} \{||\alpha_u||_p + ||\alpha_v||_p + F(\alpha_u, \alpha_v)\} \qquad (3)$$

subjecting $U = \Phi_u \alpha_u$ and $V = \Phi_v \alpha_v$, $F(\alpha_u, \alpha_v)$ is a Lagrange term.
The norm can be chosen as $p = 0, 1, 2$ or some other arbitrary p.

Sparser: if $U = \Phi_u \alpha_1 = \Phi_v \alpha_2$, then $||\alpha_1||_0 < ||\alpha_2||_0$, where $|| \cdot ||_0$ denotes the number of nonzero element.

SIMPLE: sparse in known basis.

Methods: Donoho's approach: Basis pursuit and image decomposition: image = cartoon + texture + noise.

For cartoon image, dictionaries: wavelet family
For texture image: DCT family.

Norm based method

Idea: good signal in good space, bad signal in bad space.

Two norms: $||\cdot||_u$ and $||\cdot||_v$, $||U||_u$ is very small and $||U||_v$ is big, at the same way, $||V||_v$ is very small and $||V||_u$ is big.

$$\{\hat{U}, \hat{V}\} = \arg\min_{U,V}\{||U||_u + ||V||_v + F(U, V)\} \tag{4}$$

where $F(U, V)$ is also a Lagrange term.
Norm associated space: E F G proposed by Vese et al.

SIMPLE: different space or norm.
Application: image decomposition.

Empirical method

EMD (Empirical mode decomposition): $S = \sum S_k + R$, where S_k is the so called IMF (intrinsic mode function) which is defined as:

1. the numbers of extrema and zero-crossing are different at most by 1.
2. at any point, the local mean of signal is zero, (local symmetry).

Separation method: sifting scheme.

SIMPLE: IMF, simple? maybe!
Application: HHT in signal analysis ...

Shortcoming of above methods:

decomposition by a basis or frame is too detail, can not reflect the coherent information.

other methods except EMD are very rough separation which can not separate signals with similar property, for example, both signals in bad space or both are textures.

IMF: hard to understand.

Operator based signal separation

Suppose that there is an operators: $\mathcal{T}_u$, such that $\mathcal{T}_u U = 0$ and $||\mathcal{T}_u(S - U)|| > 0$, then the task of signal separation can be done by solving the following optimization problem.

$$\hat{V} = \arg\min_{V}\{||\mathcal{T}_u(S - V)||^2 + \lambda||V||^2\} \tag{5}$$

SIMPLE: null space of operator. In discrete case, null space is finite dimension.

Signal separation based on local narrow band signals

local narrow band signal: If there exist a singular local operator $\mathcal{T}$ such that $\mathcal{T}S = 0$.
Here $\mathcal{T}$ is chosen as either the first type operator, which is similar to local integral or mean envelope, or the second type operator, which is defined as:

$$\mathcal{T} = \frac{d^2}{dt^2} + \varpi(t)^2, \tag{6}$$

$$\mathcal{T}^2(at + b)cos(\varpi(t)t + c) = 0$$

The separation can be done by solving the following optimization:

$$\hat{V}(t) = \arg\min_{V}\{||\mathcal{T}(S - V)||^2 + \lambda||\mathcal{D}V||^2\} \tag{7}$$

see: S.L. Peng and W.L. Hwang, "Adaptive Signal Decomposition Based on Local Narrow Band Signals," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 2669-2676, Jul. 2008.

Problem:

1. estimation of $\varpi(t)$, which is assumed to be the instantanuous frequency of the unknown U;
previous method: estimating $\varpi(t)$ from the extrema of S which essentially assumes that U can be estimated from the extrema of S.

shortcoming:

a). the extrema could be affected by noise;
b). the extrema could be affected by V.

2. the choice of λ, bad λ gives bad result.

Matching Pursuit

he MP algorithm decomposes a signal into a linear expansion of the bases, g_j, in an over-complete dictionary D by a succession of greedy steps. The signal S is first decomposed into

$$S = < S, g_{j_0} > g_{j_0} + R^1 S,$$

where $g_{j_0} = \arg_{g_j \in D} \max\{| < f, g_j > |\}$, $R^1 S$ is the residual signal after approximating S in the direction of g_{j_0}, and $< R^1 S, g_{j_0} > = 0$. After M iterations, we have

$$S = \sum_{k=0}^{M-1} < R^k S, g_{j_k} > g_{j_k} + R^M S, \tag{8}$$

where S is approximated by the number of M atoms and the residual $R^M S$.

We associate each $g_j \in D$ with an operator T_{g_j} such that

$$T_{g_j} : \cdot \rightarrow \cdot - < \cdot, g_j > g_j. \tag{9}$$

Since g_j is a local basis, T_{g_j} is a local operator. In addition, T_{g_j} is a singular local operator and g_j is in the null space of T_{g_j} because

$$T_{g_j}(S - < S, g_j > g_j) = S - < S, g_j > g_j \text{ and } T_{g_j}(g_j) = 0. \tag{10}$$

Thus, the MP algorithm can be regarded as using the dictionary of operators $\{T_{g_j} | g_j \in D\}$ to decompose a signal.

It applies each operator in the operator dictionary to a signal, and selects the operator T_g that satisfies

$$\min_{T \in \{T_{g_j} | g_j \in D\}} \| TS \|^2. \tag{11}$$

Let T_g be the above solution. Then, based on the definition of T_g in Equation (9), we have

$$T_g S = S - < S, g > g. \tag{12}$$

Because $T_g S$ has the minimum norm, $< S, g > g$ has the maximum norm. Thus, the MP algorithm greedily selects the operator that can remove the most components from the null space of the optimal operator in the signal.

Another view of EMD

Assume that the positions of a signal's local extrema are invariant during the sifting process.
In such cases, calculation of the mean value during the process can be represented as a linear operator, A, for all iterations. Let us assume that the sifting process converges after k iterations. Since the mean envelope derived from the extrema of S is AS, we have $IMF = (I - A)^k S$. According to the definition of the IMF, $A(IMF) = 0$; thus, S is in the null space of the operator $A(I - A)^k$ and IMF is in the null space of A.

Related to nonlinear vibration equation

Given a nonlinear vibration equation

$$u'' = f(t, u, u')$$

under arbitrary initial or boundary condition.
Commonly they will write the solution as

$$u = \sum_{k=0}^{\infty} u_k$$

where u_k satisfies

$$u_k'' + \lambda_k(t)^2 u_k = f_k$$

here $\lambda_k(t)$ can be seen as the instantaneous frequency of u_k.
Refer to Adomian Decomposition method.

Null Space Pursuit

To solve $S = U + V$, consider the following optimization

$$\min_{\alpha(t), V, \lambda_1, \gamma, \lambda_2} \left\{ || \left(\frac{d^2}{dt^2} + \alpha(t) \right) (S - V)||^2 + \lambda_1(||V||^2 + \gamma||S - V||^2) + \lambda_2||D\alpha(t)||^2 \right\}. \tag{13}$$

where $\mathcal{D}$ is a regularized operator on α. In discrete case, let

$$\mathcal{F} = || (D + P_\alpha) (S - V)||^2 + \lambda_1(||V||^2 + \gamma||S - V||^2) + \lambda_2||D\alpha||^2 \tag{14}$$

Then $\dfrac{\partial F}{\partial \alpha} = 0$ leads to

$$\alpha = (A^T A + \lambda_2 D^T D)^{-1} A^T D(S - V)) \tag{15}$$

Similarly, $\dfrac{\partial F}{\partial V} = 0$ leads to

$$V = \left(Q^T Q + (1+\gamma)\lambda_1 I\right)^{-1} \left(Q^T QS + \gamma\lambda_1 S\right) \tag{16}$$

where $Q = D + P_\alpha$.

Signal model

Suppose that $S = \tilde{U} + \tilde{V}$, where $\tilde{U}$ and $\tilde{V}$ are orthogonal to each other. Assume that $\hat{V}$ and α are the solution of (14), such that

$$\hat{V} = \beta_1 \tilde{U} + \beta_2 \tilde{V},$$

and

$$(D + P_\alpha)\, \tilde{U} = 0,$$

where β_1 and β_2 are numbers to be determined (assume that $\gamma = 2$).

$$\begin{aligned} \mathcal{F}(\alpha, \hat{V}) \;&=\; (1-\beta_2)^2 ||\,(D + P_\alpha)\,\tilde{V}||^2 + \lambda_1 \left(\beta_1^2 + (1-\beta_1)^2\right) ||\tilde{U}||^2 \\ &\quad + \lambda_1 \left(\beta_2^2 + (1-\beta_2)^2\right) ||\tilde{V}||^2 + \lambda_2 ||D\alpha||^2 \end{aligned} \tag{17}$$

where the orthogonality of $\tilde{U}$ and $\tilde{V}$ are used.

In the cost function $\mathcal{F}(\alpha, \hat{V})$, β_1 and β_2 are unknowns. It is easy to see that the two unknowns are separable, that is, the optimization of $\mathcal{F}(\alpha, \hat{V})$ can be divided into two components as

$$\mathcal{F}_1 = \lambda_1(\beta_1^2 + (1 - \beta_1)^2)||\tilde{U}||^2 \tag{18}$$

and $\mathcal{F} - \mathcal{F}_1$. Since λ_1 and $\tilde{U}$ are assumed to be known, then the optimization of $\mathcal{F}_1$ will lead to $\beta_1 = 0.5$. Similarly, the optimization of $\mathcal{F} - \mathcal{F}_1$ leads to

$$\beta_2 = \frac{||\,(D + P_\alpha)\,\tilde{V}||^2 + \lambda_1||\tilde{V}||^2}{||\,(D + P_\alpha)\,\tilde{V}||^2 + 2\lambda_1||\tilde{V}||^2}.$$

Under assumption $||\,(D + P_\alpha)\,\tilde{V}|| > 0$, and λ_1 is very small, then $\beta_2 \simeq 1$.

This result shows that the final optimal solution of (14) is like $\hat{V} \simeq 0.5\tilde{U} + \tilde{V}$, that means $S - \hat{V} \simeq 0.5\tilde{U}$, by using the orthogonality,

$$\frac{(S-\hat{V})^T S}{||S-\hat{V}||^2} \simeq 2.$$

On the other hand, from (16),
$\hat{V} = \left(Q^T Q + 2\lambda_1 I\right)^{-1}\left(Q^T QS + \lambda_1 S\right)$, then
$S - \hat{V} = \lambda_1 (Q^T Q + 2\lambda_1 I)^{-1} S$ and then

$$\frac{(S-\hat{V})^T S}{||S-\hat{V}||^2} = \frac{1}{\lambda_1}\frac{S^T MS}{S^T M^T MS} \tag{19}$$

where $M = (Q^T Q + 2\lambda_1 I)^{-1}$. By using the fact that $\dfrac{(S-\hat{V})^T S}{||S-\hat{V}||^2} \simeq 2$, then

$$\frac{S^T MS}{2S^T M^T MS} \simeq \lambda_1. \tag{20}$$

Null Space Pursuit

Step 1 Input: the signal S, the parameter $\hat{\lambda}_2$, the stopping threshold ϵ, and the initial values of λ_1^0 and γ^0.

Step 2 Let $j = 0$, $\hat{V}_j = 0$, $\lambda_1^j = \lambda_1^0$, and $\gamma^j = \gamma^0$.

Step 3 Compute $\hat{\alpha}_j$ as follows:

$$\hat{\alpha}_j = (A_j^T A_j + \hat{\lambda}_2 D^T D)^{-1} A_j^T D(S - \hat{V}_j)), \quad (21)$$

where A_j is a diagonal matrix whose diagonal elements are equal to $(S - \hat{V}_j)$.

Step 4 Compute λ_1^{j+1} as follows:

$$\lambda_1^{j+1} = \frac{S^T M(Q_j, \lambda_1^j, \gamma^j)^T S}{(1+\gamma^0) S^T M(Q_j, \lambda_1^j, \gamma^j)^T M(Q_j, \lambda_1^j, \gamma^j) S}, \quad (22)$$

where $M(Q_j, \lambda_1^j, \gamma^j) = (Q_j^T Q_j + (1+\gamma^j)\lambda_1^j I)^{-1}$ and $Q_j = D + P_{\hat{\alpha}_j}$.

Step 5 Compute $\hat{V}_{j+1}$ according to the Equation (**??**) as follows:

$$\hat{V}_{j+1} = \left(Q_j^T Q_j + (1+\gamma^j)\lambda_1^{j+1} I\right)^{-1} \left(Q_j^T Q_j S + \gamma^j \lambda_1^{j+1} S\right). \quad (23)$$

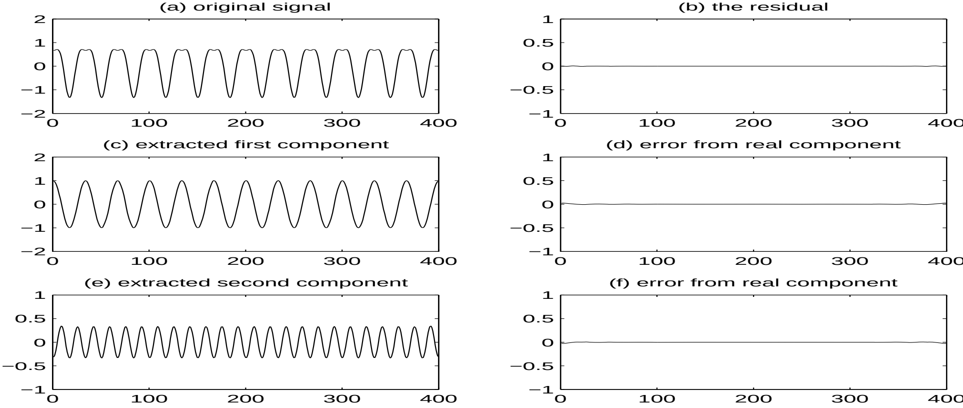
Step 6 Compute γ^{j+1} as follows:

$$\gamma^{j+1} = \frac{(S - \hat{V}_{j+1})^T S}{||S - \hat{V}_{j+1}||^2} - 1. \quad (24)$$

Step 7 If $||\hat{V}_{j+1} - \hat{V}_j|| > \epsilon ||S||$, then set $j = j+1$ and go to **Step 3**.

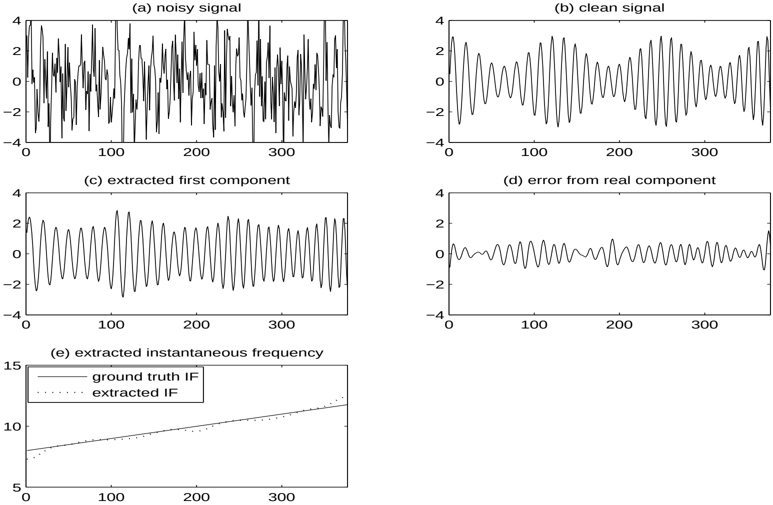
Step 8 Output: the optimal residual signal $\hat{V} = \hat{V}_{j+1}$, the parameter $\hat{\lambda}_1 = \lambda_1^{j+1}$, the leakage parameter $\hat{\gamma} = \gamma^{j+1}$, and the operator parameters $\hat{\alpha} = \alpha_j$.

Example 1. Harmonic signal extraction



Comments: Using the NSP algorithm to decompose $\cos(4t) - \frac{1}{3}\cos(8t)$. Top left: The input signal. Top right: the residual signal after the first and second subcomponents are extracted from the input signal. Middle left: the first extracted component. Middle right: the error signal obtained by subtracting the first extracted component from $\cos(4t)$. Bottom left: the second extracted component. Bottom right: the error signal obtained by subtracting the second extracted component from $-\frac{1}{3}\cos(8t)$. We use $\lambda_1^0 = 1e - 6$ as the initial value to extract both subcomponents.

example 2. Denoising for chirp signal



Comments: Removing the noise from a noisy chirp signal. Top left: the noisy chirp signal, which has an SNR of -0.12 dB. Top right: the clean chirp signal. Middle left: the extracted subcomponent, which has an SNR of 9 dB. Middle right: the error signal obtained by subtracting the extracted component from the clean signal. Bottom left: the instantaneous frequency of the extracted component (extracted IF) and that of the clean chirp signal (ground truth IF) are superimposed. The initial value of λ_1^0 is set at 0.00005. Note that, in this case, the noise level is so high that the instantaneous frequency of the chirp signal cannot be estimated accurately from the extrema of the noisy signal.

The proposed algorithm can estimate the local frequency of extracted component. In above signal, we can verify this issue.

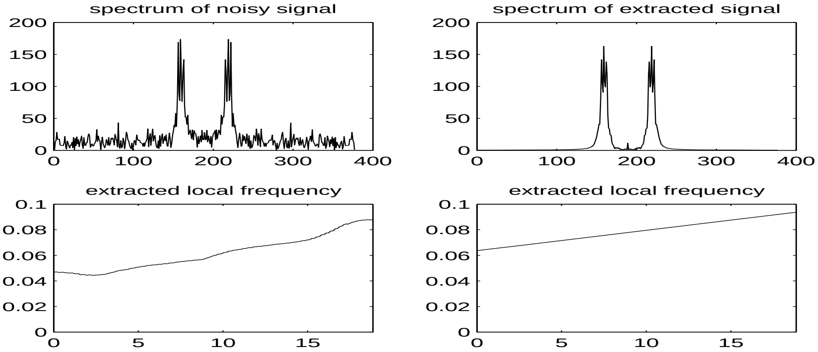


Figure: Top left:signal spectrum with noise. Top right: extracted signal spectrum. bottom left: instantanuous frequency of extracted first component. bottom right: instantanuous frequency of real component.

example 3. Denoising for piecewise smooth signal

The initial value of λ_1 is also 0.0001, stops at 0.00015.

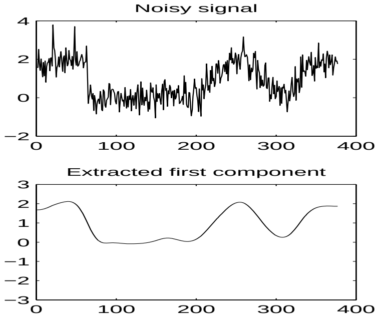


Figure: Top left:signal with noise. Top right: clean signal. bottom left: extracted first component. bottom right: error from real component.

example 4. Analysis of real signal 1

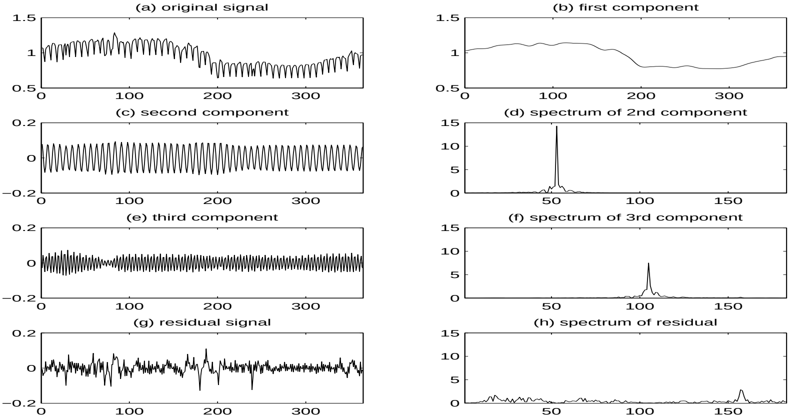


Figure: Top left:real signal. Other five figures are the first 5 components.

example 4. Analysis of real signal 2

Temperature in past 150 years.

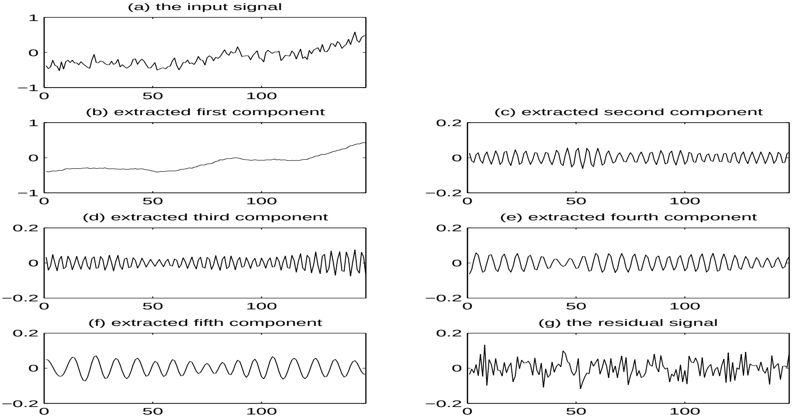


Figure: Top left:real signal. Other five figures are the first 5 components by NSP.

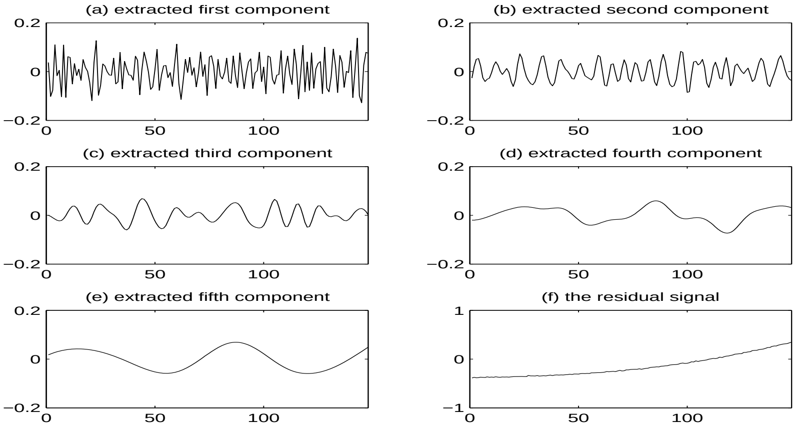


Figure: Separation results of the global surface temperature data (see subfigure (a) of Figure 6) derived by the EEMD algorithm. The code of the algorithm can be found in [**?**]. Subfigures (a),(b),(c),(d),(e), and (f) are the extracted first, second, third, fourth and fifth components and the residual signal respectively.

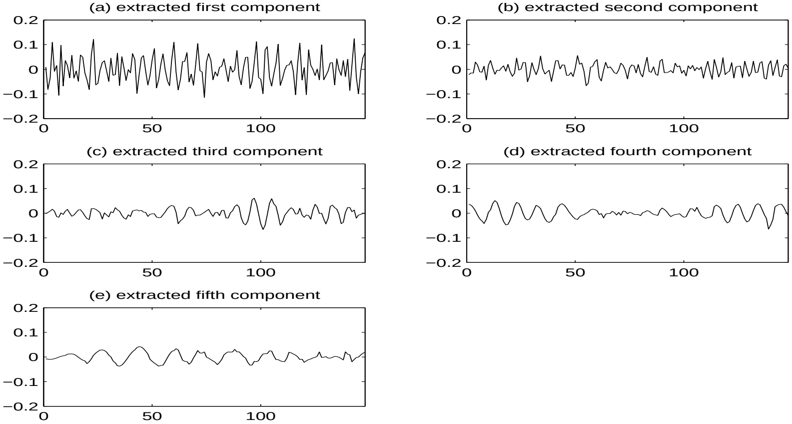


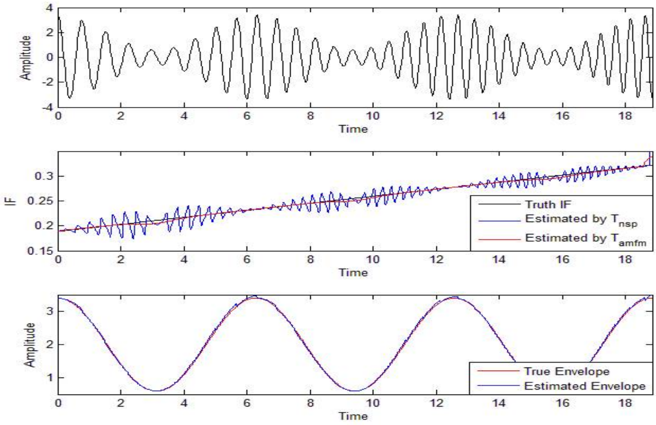
Figure: Separation results of the global surface temperature data (see subfigure (a) of Figure 6) derived by using the separation algorithm in [?].

extension

The operator can be extended to the following form

$$\frac{d^2}{dt^2} + p\frac{d}{dt} + q$$

which means the vibration has damping term.
Such operator will lead to better behavior in estimating instantaneous frequency and amplitude.



Amplitude
Time
IF
Truth IF
Estimated by T_{nsp}
Estimated by T_{amfm}
Time
Amplitude
True Envelope
Estimated Envelope
Time

Figure:

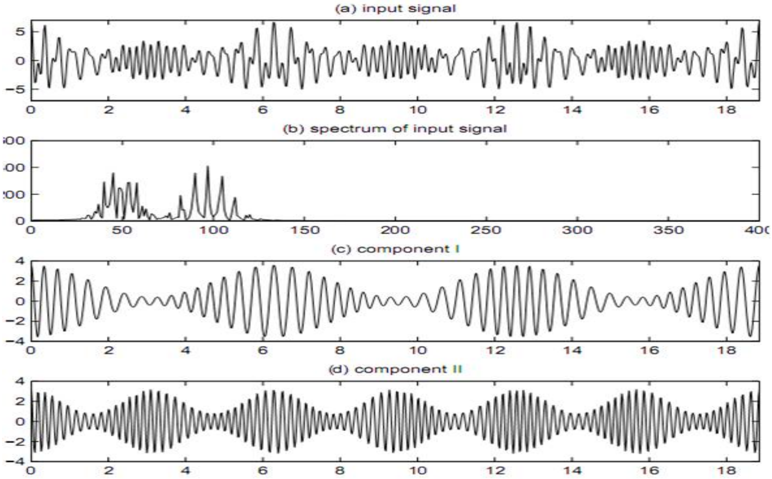

(a) input signal
(b) spectrum of input signal
(c) component I
(d) component II

Figure:

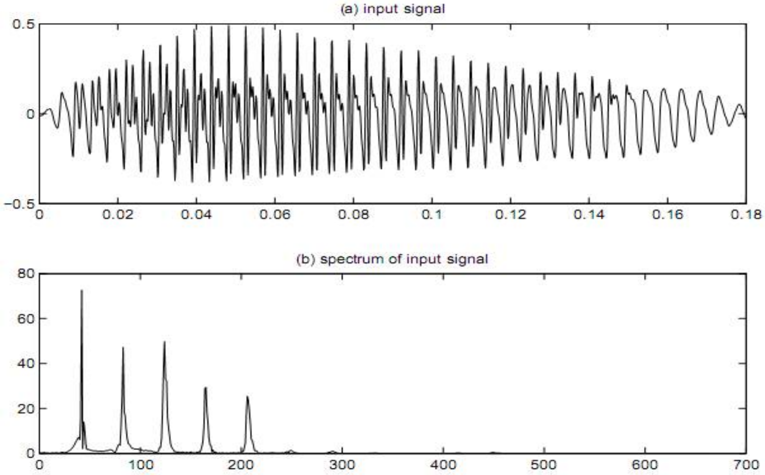

(a) input signal
0.5
0
−0.5
0
0.02
0.04
0.06
0.08
0.1
0.12
0.14
0.16
0.18
(b) spectrum of input signal
80
60
40
20
0
0
100
200
300
400
500
600
700

Figure:

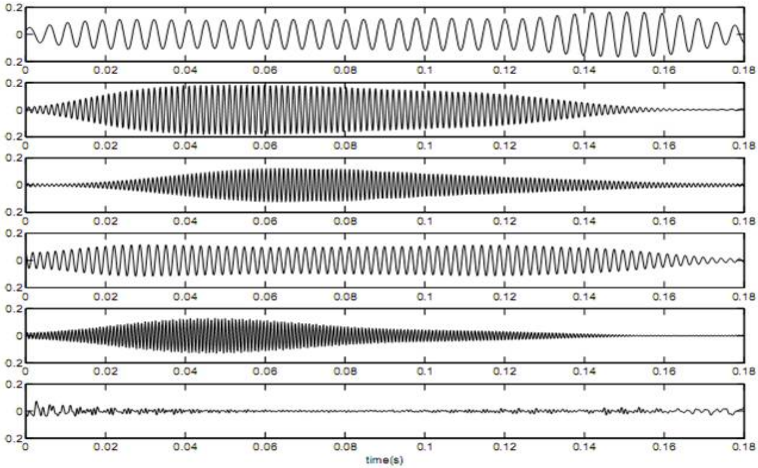


Figure:

Discussion and conclusion

1. operator based signal separation.
2. data driven, independent of extrema;
3. robust against noise;
4. automatic parameter determination;
5. instantanuous frequency estimation;
6. easily generalized to images.

Thanks and Questions