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

Silong Peng and Wen-Liang Hwang

***Abstract*—The operator-based signal separation approach uses an adaptive operator to separate a signal into additive subcomponents. The approach can be formulated as an optimization problem whose optimal solution can be derived analytically. However, the following issues must still be resolved: estimating the robustness of the operator's parameters and the Lagrangian multipliers, and determining how much of the information in the null space of the operator should be retained in the residual signal. To address these problems, we propose a novel optimization formula for operator-based signal separation and show that the parameters of the problem can be estimated adaptively. We demonstrate the effectiveness of the proposed method by processing several signals, including real-life signals.**

***Index Terms*—Adaptive signal separation, operator-based, optimal parameter estimation.**

I. Introduction

In recent years, the single-channel signal separation problem, which involves decomposing a signal into its coherent subcomponents, has attracted a great deal of attention because it affects many applications. Typical single-channel signal separation approaches decompose a signal into a mixture of several additive coherent subcomponents [1]–[3], [11], [14], [18], [19]. The methods used to separate signals vary because different subcomponents are used to construct the signals. For example, in the empirical mode decomposition (EMD) approach [11], [17], [20], an oscillatory signal is decomposed into a sum of intrinsic mode functions (IMFs); and in the matching pursuit (MP) approach [3], [6], a signal is decomposed into a sum of time-frequency atoms. The following three recently proposed approaches are of particular interest to us.

1) Sparsity-Based Approach: In a series of papers, the sparsity-based approach (also called morphological component analysis) was developed to separate textures from the piecewise smooth components [5], [18]. Assume that a signal S is decomposed into a sum of U and V. Two dictionaries, Φ_u and Φ_v, are associated with U and V respectively, such that U (respectively, V) can be represented by Φ_u (respectively, Φ_v) with sparse coefficients (the number of non-zero coefficients is small), but it cannot be sparsely represented by Φ_v (respectively, Φ_u). Given the two dictionaries Φ_u and Φ_v, the separation of S can be achieved by solving the following optimization problem:

$$\min_{\alpha_u,\alpha_v}\{\|\alpha_u\|_1+\|\alpha_v\|_1+F(\alpha_u,\alpha_v)\} \tag{1}$$

where $U=\Phi_u\alpha_u$, $V=\Phi_v\alpha_v$, and $F(\alpha_u,\alpha_v)$ is a Lagrange term. Starck *et al.* [18] use the approach to separate an image into a cartoon component and a texture component. The two dictionaries used in [18] are the wavelet family for cartoon images and the Gabor family for textures. In general, the dictionaries employed can be learned from a training process.

2) Norm-Based Approach: Vese *et al.* [19] assume that the components U and V lie in different metric spaces with norms $\|\cdot\|_u$ and $\|\cdot\|_v$, respectively. To separate the signal S into U and V, it is necessary that $\|U\|_u$ is small and $\|U\|_v$ is large, and $\|V\|_v$ is small and $\|V\|_u$ is large. To find U and V, the following optimization problem must be solved:

$$\min_{U,V}\{\|U\|_u+\|V\|_v+F(U,V)\} \tag{2}$$

where $F(U,V)$ is a Lagrange term. The authors show that the cartoon and texture components of an image can be separated by their approach. Following Rudin, Osher, and Fatemi [9], the bounded variation (BV) is chosen as the space for cartoon images; and following Y. Meyer [12], three spaces, $E=B^{-1}_{\infty,\infty}$, $F=\mathrm{div}(BMO)$ and $G=\mathrm{div}(L^\infty)$, are proposed for texture images. Details of the definition of the spaces and their corresponding norms can be found in [7], [12], and [19]. The approach can be extended to obtain a hierarchical decomposition of images for denoising, deblurring, and segmentation purposes [13].

Although the sparsity- and norm-based methods are based on an ingenious idea and derive solutions according to advanced mathematical techniques, they cannot separate many signals successfully. For example, if a signal is a sum of narrow band signals, it is very difficult to find sparse dictionaries or general spaces to separate the narrow band subcomponents effectively. In fact, any narrow band signal is a sparse representation of the dictionary of the DCT; therefore, all the narrow band components of a signal share the same sparse dictionary. To separate the fine-structured components of a signal, such as narrow band components, one needs a signal separation method that can derive the operations based on the structures of a signal, after which the operations are used to separate the signal into the

Manuscript received May 21, 2009; accepted January 08, 2010. Date of publication January 26, 2010; date of current version April 14, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lieven De Lathauwer. This work was supported by the National Natural Science Foundation of China (60972126), the Joint Funds of the National Natural Science Foundation of China (U0935002/L05), and the Beijing Natural Science Foundation (4102060).

S. Peng is with the Institute of Automation, The Chinese Academy of Sciences, Beijing 100190, China (e-mail: silong.peng@ia.ac.cn).

W.-L. Hwang is with the Institute of Information Science, Academia Sinica, Taipei 11529, Taiwan, R.O.C. He is also with the Department of Computer Science and Information Engineering, Kainan University, Taoyuan, Taiwan, R.O.C. (e-mail: whwang@iis.sinica.edu.tw).

Digital Object Identifier 10.1109/TSP.2010.2041606

desired components. This was one of the motivations for the development of the operator-based approach.

3) Operator-Based Approach: The operator-based approach separates S into U and V so that $V = S - U$ is in the null space of an operator $\mathcal{T}_s$. The subindex, s, of the operator indicates that the operator can be estimated from the signal S. The following optimization method is used to estimate the signal $\hat{U}$ that minimizes the problem:

$$\min_U \left\{ \|\mathcal{T}_s(S-U)\|^2 + \lambda \|D(U)\|^2 \right\} \tag{3}$$

where U is the residual signal, and D is a differential operator that regulates U. Minimizing the term $\|\mathcal{T}_s(S-U)\|^2$ indicates that $S-U$ is in the null space of the operator $\mathcal{T}_s$. The solution of (3) can be derived analytically:

$$\hat{U} = \left(\mathcal{T}_s^* \mathcal{T}_s + \lambda D^* D\right)^{-1} \mathcal{T}_s^* \mathcal{T}_s S. \tag{4}$$

In [14], we define a class of local narrow signals in the null space of the differential operator as follows:

$$\sum_{k \in Z} \alpha(k) \frac{d^k}{dt^k} \tag{5}$$

where $\{\alpha(k)\}$ is a square summable sequence belonging to $l_2(Z)$.

The operator-based approach can be used to decompose the residual signal $\hat{U}$ repeatedly. Hence, S can be represented as the summation of subcomponents in the null spaces of a sequence of operators derived from the corresponding sequence of residual signals. There are similarities between the operator-based approach, the MP approach and the EMD approach, which we elaborate on in Section II. Here we remark that, similar to the approach in [13], we define $\|f\|_{\mathcal{O}}$ as the norm of signal f to the null space of operator $\mathcal{O}$. As a result, (3) can be written as $\|S-U\|_{\mathcal{T}_s} + \lambda \|U\|_D$, which is a form of (2).

When an operator-based method is applied to a real-life signal, it is usually difficult to determine the Lagrangian multiplier λ in (3). In addition, during the optimization of the equation, most of the signal in the null space of $\mathcal{T}_s$ is removed from S. A small portion of S is required to regulate $\hat{U}$, so it is retained in the residual signal $\hat{U}$. However, we found that, for many signals, better solutions can usually be obtained if we remove less information from the null space of $\mathcal{T}_s$ than that required by (3). This suggests that we should use a less greedy approach so as to preserve more information in the null space of $\mathcal{T}_s$ in the residual signal. Thus, we propose the following solution to the optimization problem:

$$\hat{U} = \arg\min_U \Big\{ \|\mathcal{T}_{\mathcal{S}}(S-U)\|^2 + \lambda(\|D(U)\|^2 + \gamma \|S-U\|^2) + F(\mathcal{T}_{\mathcal{S}}) \Big\}. \tag{6}$$

The first and the second terms of (6) correspond to the terms in (3). The parameter γ in the third term of (6) determines the amount of $S-U$ to be retained in the null space of $\mathcal{T}_s$; and the last term is the Lagrange term for the parameters of the operator $\mathcal{T}_s$.

In this paper, we show that the optimal solution of (6) can be derived analytically, and the parameters of the operator $\mathcal{T}_s$ as well as the parameters λ and γ in the equation can be estimated adaptively. The proposed algorithm is called the Null Space Pursuit algorithm. We demonstrate the algorithm's robustness and accuracy in decomposing noisy signals as well as its efficacy when applied to some real-life signals.

The remainder of this paper is organized as follows. In Section II, we review the operator-based approach and compare it with the MP and the EMD approaches. In Section III, we present the proposed Null Space Pursuit algorithm. Experiments on simulated and real-world signals are reported in Section IV. We then summarize our conclusions in Section V.

II. Comparison With the MP and EMD Approaches

In this section, we review the operator-based approach proposed in [14] and compare it with the MP approach and the EMD approach. We show that the operator-based approach can be regarded as a generalization of the MP approach, and that a particular case of EMD decomposition is in the null space of an adaptive operator.

A. Operator-Based Approach

The operator-based approach uses adaptive operators to decompose a signal into additive subcomponents. The basic operation involves estimating an operator from a signal S and decomposing the signal into two components, U and V, such that V is in the null space of the operator and U is the residual signal. The operation can be formulated as an optimization problem by (3).

In [14], two types of local operators are used: an integral operator and a differential operator. The differential operator is

$$\mathcal{T}_s = \left(\frac{1}{\varpi(t)^2} \frac{d^2}{dt^2} + 1 \right)^2 \tag{7}$$

where $\varpi(t)$ is the instantaneous frequency of the component V [8]. It is estimated from the positions of the local extrema of the signal S. The null space of the above differential operator contains the narrow band signal $P_2(t)\cos(wt)$, where $P_2(t)$ is a polynomial of order 1.

The steps of the signal decomposition algorithm proposed in [14] are as follows.

Step 1) Input the signal S and the stopping threshold ϵ. Set $V_0 = S$ and $i = 0$.

Step 2) Estimate the instantaneous frequency from V_i and derive the operator $\mathcal{T}_i$ by (7).

Step 3) Choose a λ value and use (4) to obtain U_{i+1}:

$$U_{i+1} = \left(\mathcal{T}_s^* \mathcal{T}_s + \lambda D^* D\right)^{-1} \mathcal{T}_s^* \mathcal{T}_s V_i. \tag{8}$$

Step 4) If $\|U_{i+1}\| > \epsilon$, then let $V_{i+1} = V_i - U_{i+1}$ and $i = i+1$. Go to Step 2.

In Step 2, the instantaneous frequency of a signal is estimated from the locations of its local extrema. However, in some signals comprised of two additive subcomponents, the correct instantaneous frequency of each subcomponent cannot be estimated

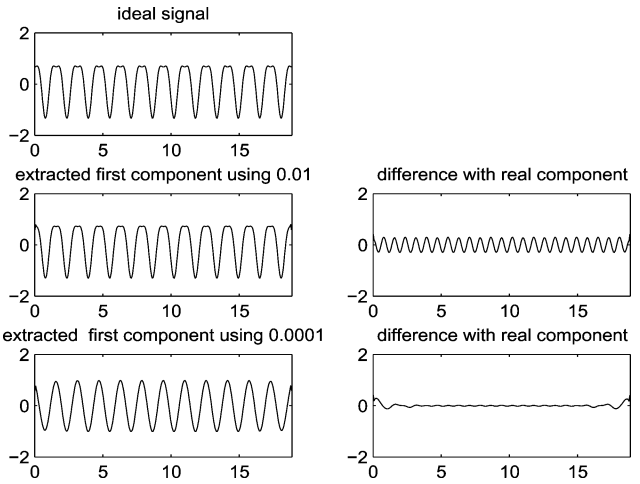

Fig. 1. Top row: Signal $S(t) = \cos(4t) - (1/3)\cos(8t)$ with $t \in [0, 6\pi]$. Middle row: the left-hand and right-hand frames show, respectively, the extracted signal and the error signal derived by setting $\lambda = 0.01$. Bottom row: the left-hand and right-hand frames show, respectively, the extracted signal and the error signal derived by setting $\lambda = 0.0001$. The extracted components vary because we use different λ values.

accurately from the extrema of the signals [14]. Moreover, selection of the value of λ in Step 3 is difficult. In the following, we describe a simple signal that is formed by summing two single-tone signals, as shown in the top subfigure of Fig. 1. The two subfigures in the middle row of the figure are, respectively, the extracted component and the error of the component if we set $\lambda = 0.01$; while the bottom subfigures show, respectively, the extracted component and the error of the component if we set $\lambda = 0.0001$ instead. An automatic and robust estimation of λ is extremely important because if an inappropriate λ value is chosen in one iteration, it will affect the residual signal and be propagated to subsequent iterations. The above algorithm decomposes S into a sum of subcomponents, which are in the null spaces of a sequence of operators. Then, we have

$$S = \sum_{i=1}^{k} V_i + U_k \tag{9}$$

where V_i is in the null space of the operator $\mathcal{T}_{i-1}$, which is derived from $S - \sum_{k=1}^{i-1} V_k$. The procedure is similar to that of the MP algorithm, where the basis that best matches the signal is selected, and the component that the signal projects onto the basis is extracted from the signal in each iteration of the algorithm. The procedure is also similar to that of the EMD algorithm, where an IMF derived from the signal is extracted from the signal in each iteration. Next, we compare the operator-based approach with the MP and EMD approaches.

B. Comparison With the MP Approach

The MP algorithm decomposes a signal into a linear expansion of the bases, g_j, in an overcomplete dictionary D by a succession of greedy steps [3]. The signal S is first decomposed into

$$S = \langle S, g_{j_0} \rangle g_{j_0} + R^1 S$$

where $g_{j_0} = \arg_{g_j \in D} \max\{|\langle f, g_j \rangle|\}$, $R^1 S$ is the residual signal after approximating S in the direction of g_{j_0}, and $\langle R^1 S, g_{j_0} \rangle = 0$. The dictionary element, g_{j_0}, combined with the inner product value $\langle S, g_{j_0} \rangle$ is called an atom. The matching pursuit algorithm then decomposes the residual $R^1 S$ by projecting it on to the basis functions of D, as was done for S. After M iterations, we have

$$S = \sum_{k=0}^{M-1} \langle R^k S, g_{j_k} \rangle g_{j_k} + R^M S \tag{10}$$

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

Next, we show that the operator-based approach can be regarded as a generalization of the MP approach. We associate each $g_j \in D$ with an operator T_{g_j} such that

$$T_{g_j} : \cdot \to \cdot - \langle \cdot, g_j \rangle g_j. \tag{11}$$

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

$$T_{g_j}(S - \langle S, g_j \rangle g_j) = S - \langle S, g_j \rangle g_j \text{ and } T_{g_j}(g_j) = 0. \tag{12}$$

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

$$\min_{T \in \{T_{g_j} | g_j \in D\}} \|TS\|^2. \tag{13}$$

Let T_g be the above solution. Then, based on the definition of T_g in (11), we have

$$T_g S = S - \langle S, g \rangle g. \tag{14}$$

Because $T_g S$ has the minimum norm, $\langle S, g \rangle g$ 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.

C. Comparison With the EMD Approach

The EMD algorithm decomposes a signal into a sum of intrinsic mode functions (IMFs). Each IMF must satisfy two conditions: 1) the number of extrema and the number of zero-crossings can differ by one, at most; and 2) the mean value of the envelopes defined by the local maxima and the local minima must be zero. An IMF can be obtained by the following sifting procedure. Given a signal S, the procedure first finds the extremal points, and then computes the mean value M of the envelopes of the extrema. If the mean value is not zero, the procedure is applied to the new signal $S - M$. The procedure is repeated several times until the mean value of a signal is zero and the signal is an IMF. The sifting operation of the EMD is defined and derived from the local extrema of a signal, while the operator of the operator-based approach can be derived from a regularization approach by solving an optimization problem. Usually, the latter approach obtains a better result when the signal is contaminated with noise. Examples of the signal separation results

derived by the operator-based and the EMD algorithm are compared in [4].

The approach described in [14] considers a special case of IMF computation in which 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. Note that the derivation can still be applied when the first few iterations of the extrema positions vary, but the rest of the iterations are invariant during the sifting process.

III. Null Space Pursuit

In the previous sections, we reviewed the operator-based approach proposed in [14] and identified some difficulties that have yet to be resolved. To address those difficulties, we propose a modification of the approach and show that the new parameters can be estimated adaptively at the same time.

If we set the regulation operator D in (3) as the identity matrix, we obtain

$$\min_U \left\{ ||\mathcal{T}_s(S-U)||^2 + \lambda ||U||^2 \right\}. \tag{15}$$

This is essentially a greedy approach since (15) extracts the most suitable component in the null space of $\mathcal{T}_s$ from S by minimizing $||U||^2$. It can remove more than enough information from the null space of $\mathcal{T}_s$ in the residual signal. Let us compare the following two decompositions of the signal S: $S = V + U$, and $S = (1-\beta)V + \beta V + U$, where V is in the null space of $\mathcal{T}_s$, and $\beta \neq 0$. In the first case, V is extracted and, at the next iteration, the operator $\mathcal{T}_1$ is estimated from the residual U; and in the second case, $(1-\beta)\,V$ is extracted and, at the next iteration, the operator $\mathcal{T}_2$ is estimated from the residual signal $\beta V + U$. The operators $\mathcal{T}_1$ and $\mathcal{T}_2$ are not necessarily the same. Thus, we propose using a controlled greedy approach based on the following formula:

$$\min_U \left\{ ||\mathcal{T}_s(S-U)||^2 + \lambda_1(||U||^2 + \gamma ||S-U||^2) + F(\mathcal{T}_s) \right\} \tag{16}$$

where γ denotes the leakage parameter and the last term is the Lagrange term of the parameters of the operator $\mathcal{T}_s$. We call γ the leakage parameter because its value determines the amount of information about $S - U$ that is retained in the null space of $\mathcal{T}_s$.

To demonstrate the proposed approach, we choose the following differential operator:

$$\mathcal{T}_s = \frac{d^2}{dt^2} + \alpha(t). \tag{17}$$

If the input is a single-tone signal $e^{i\phi(t)}$, then $\mathcal{T}_s e^{i\phi(t)} = 0$ implies that the parameters $\alpha(t) = \phi'(t)^2$ are the square of the instantaneous frequency of the signal. To ensure that $\alpha(t)$ is a smooth function, we choose the second differential operator D in $\alpha(t)$ and the parameter λ_2 for the Lagrangian term $F(\mathcal{T}_s)$. The optimization problem in (16) then becomes

$$\min_{\alpha(t),U,\lambda_1,\gamma,\lambda_2} \left\{ \left\| \left(\frac{d^2}{dt^2} + \alpha(t) \right)(S-U) \right\|^2 + \lambda_1(||U||^2 + \gamma ||S-U||^2) + \lambda_2 ||D\alpha(t)||^2 \right\}. \tag{18}$$

To separate a multi-tone signal S, we must apply (18) several times, in the same way as the algorithm described in Section II-A. In the following, we show that the parameters in (18), except for λ_2, can be estimated simultaneously.

A. Discrete Representation

In a discrete case, S, U and α are column vectors of length L, and D is the matrix of the second difference. The optimization problem is rewritten as

$$\min_{\alpha,U,\lambda_1,\gamma,\lambda_2} \left\{ ||(D + P_\alpha)(S-U)||^2 + \lambda_1(||U||^2 + \gamma ||S-U||^2) + \lambda_2 ||D\alpha||^2 \right\} \tag{19}$$

where P_α is a diagonal matrix whose diagonal elements are equal to α. Let $\hat{\alpha}$, $\hat{U}$, $\hat{\lambda}_1$, $\hat{\gamma}$ and $\hat{\lambda}_2$ be the solution of the above equation. Given $\hat{\lambda}_1$, $\hat{\gamma}$, and $\hat{\lambda}_2$. Equation (19) becomes

$$F(\alpha, U, \hat{\lambda}_1, \hat{\gamma}, \hat{\lambda}_2) = ||(D + P_\alpha)(S-U)||^2 + \hat{\lambda}_1(||U||^2 + \hat{\gamma} ||S-U||^2) + \hat{\lambda}_2 ||D\alpha||^2. \tag{20}$$

Then, we have

$$\frac{\partial F}{\partial \alpha} = 2A^T(A\alpha + D(S-U)) + 2\lambda_2 D^T D\alpha \tag{21}$$

where A is a diagonal matrix whose diagonal elements are equal to $(S-U)$. Let $\hat{\lambda}_2 = \lambda_2$. To estimate $\hat{\alpha}$, we use the equation $(\partial F/\partial \alpha) \mid_{U=\hat{U}} = 0$ and obtain

$$\hat{\alpha} = -\left(A^T A + \hat{\lambda}_2 D^T D \right)^{-1} A^T D(S - \hat{U}). \tag{22}$$

Similarly, to estimate $\hat{U}$, we use the equation $(\partial F/\partial U) \mid_{\alpha=\hat{\alpha}} = 0$ and obtain

$$\begin{aligned} \hat{U} &= \left(Q^T Q + (1+\hat{\gamma})\hat{\lambda}_1 I \right)^{-1} \left(Q^T Q S + \hat{\lambda}_1 \hat{\gamma} S \right) \\ &= M(\hat{\lambda}_1, \hat{\gamma}) \left(Q^T Q S + \hat{\lambda}_1 \hat{\gamma} S \right) \end{aligned} \tag{23}$$

where $Q = D + P_{\hat{\alpha}}$ and $M(\hat{\lambda}_1, \hat{\gamma}) = \left(Q^T Q + (1+\hat{\gamma})\hat{\lambda}_1 I \right)^{-1}$.

The optimal value of $\hat{\alpha}$ depends on the parameter $\hat{\lambda}_2$. Our numerical experiments show that $\hat{\lambda}_2$ is insensitive to the optimal solution; thus, we assign it a fixed value. The optimal residual $\hat{U}$ is dependent on the parameters $\hat{\lambda}_1$ and $\hat{\gamma}$, and the value of $\hat{\lambda}_1$ is sensitive to the solution. Next, we propose a signal model and use it to estimate the parameters $\hat{\lambda}_1$ and $\hat{\gamma}$.

B. Signal Model

We assume that a signal S can be decomposed into two components $\tilde{U}$ and $\tilde{V}$ that are orthogonal to each other, and let $\tilde{V}$

be in the null space of the operator $D + P_{\hat{\alpha}}$. The signal model assumes that

$$S = \tilde{V} + \tilde{U} \tag{24}$$

where $\tilde{V}^T\tilde{U} = 0$ and $(D + P_{\hat{\alpha}})\,\tilde{V} = 0$. It also assumes that the optimal residual signal $\hat{U}$ is a linear mixture of $\tilde{V}$ and $\tilde{U}$:

$$\hat{U} = \beta_1\tilde{V} + \beta_2\tilde{U} \tag{25}$$

where β_1 and β_2 are coefficients. Substituting (24) and (25) into (20) and using $\tilde{V}^T\tilde{U} = 0$ and $(D + P_{\hat{\alpha}})\tilde{V} = 0$, we have (26)

$$\begin{aligned} F(\hat{\alpha}, \hat{U}, \lambda_1, \gamma, \hat{\lambda}_2) &= \|(D + P_{\hat{\alpha}})(S - \hat{U})\|^2 \\ &\quad + \lambda_1(\|\hat{U}\|^2 + \gamma\|S - \hat{U}\|^2) + \hat{\lambda}_2\|D\hat{\alpha}\|^2 \\ &= (1 - \beta_2)^2\|(D + P_{\hat{\alpha}})\tilde{U}\|^2 \\ &\quad + \lambda_1\left(\beta_1^2 + \gamma(1 - \beta_1)^2\right)\|\tilde{V}\|^2 \\ &\quad + \lambda_1\left(\beta_2^2 + \gamma(1 - \beta_2)^2\right)\|\tilde{U}\|^2 + \hat{\lambda}_2\|D\hat{\alpha}\|^2. \end{aligned} \tag{26}$$

The above equation shows that the terms related to β_1 and β_2 can be separated, so the optimization of $F(\hat{\alpha}, \hat{U}, \lambda_1, \gamma, \hat{\lambda}_2)$ can be divided into two components as follows:

$$F_1 = \lambda_1\left(\beta_1^2 + \gamma(1 - \beta_1)^2\right)\|\tilde{V}\|^2 \tag{27}$$

and $F - F_1$. By taking the partial derivative of F_1 with respect to β_1 and setting the result to zero, we obtain

$$\beta_1 = \frac{\gamma}{1 + \gamma}. \tag{28}$$

Similarly, by taking the partial derivative of $F - F_1$ with respect to β_2 and setting the result to zero, we obtain

$$\beta_2 = \frac{\|(D + P_{\hat{\alpha}})\tilde{U}\|^2 + \lambda_1\gamma\|\tilde{U}\|^2}{\|(D + P_{\hat{\alpha}})\tilde{U}\|^2 + \lambda_1(1 + \gamma)\|\tilde{U}\|^2}. \tag{29}$$

If we assume that λ_1 has a very small value, then we have $\beta_2 \simeq 1$ because $\|(D + P_{\hat{\alpha}})\tilde{U}\| > 0$. Substituting the above results for β_1 and β_2 into (25), we obtain

$$\hat{U} = \frac{\gamma}{1 + \gamma}\tilde{V} + \tilde{U}. \tag{30}$$

By (24) and (30), we have

$$S - \hat{U} = \frac{1}{1 + \gamma}\tilde{V}. \tag{31}$$

The component removed from S by solving (19) with the operator $D + P_{\hat{\alpha}}$ is $S - \hat{U}$, which is a $1/(1 + \gamma)$ fraction of $\tilde{V}$.

Based on the assumptions that $S = \tilde{V} + \tilde{U}$ and $\tilde{V}^T\tilde{U} = 0$, as well as the result of (31), we can derive the optimal value of the parameter γ by

$$\frac{(S - \hat{U})^T S}{\|S - \hat{U}\|^2} = 1 + \gamma = 1 + \hat{\gamma}. \tag{32}$$

On the other hand, from (23), we have

$$\begin{aligned} S - \hat{U} &= S - \left(Q^TQ + (1 + \hat{\gamma})\hat{\lambda}_1 I\right)^{-1}\left(Q^TQS + \hat{\lambda}_1\hat{\gamma}S\right) \\ &= \hat{\lambda}_1(Q^TQ + (1 + \hat{\gamma})\hat{\lambda}_1 I)^{-1}S \\ &= \hat{\lambda}_1 M(\hat{\lambda}_1, \hat{\gamma})S \end{aligned} \tag{33}$$

where $M\left(\hat{\lambda}_1, \hat{\gamma}\right) = \left(Q^TQ + (1 + \hat{\gamma})\hat{\lambda}_1 I\right)^{-1}$. By (33), we have

$$\frac{(S - \hat{U})^T S}{\|S - \hat{U}\|^2} = \frac{1}{\hat{\lambda}_1}\frac{S^T M(\hat{\lambda}_1, \hat{\gamma})^T S}{S^T M(\hat{\lambda}_1, \hat{\gamma})^T M(\hat{\lambda}_1, \hat{\gamma})S}. \tag{34}$$

From (32) and (34), the optimal parameters $\hat{\lambda}_1$ and $\hat{\gamma}$ are related by

$$\frac{1}{\hat{\lambda}_1}\frac{S^T M(\hat{\lambda}_1, \hat{\gamma})^T S}{S^T M(\hat{\lambda}_1, \hat{\gamma})^T M(\hat{\lambda}_1, \hat{\gamma})S} = 1 + \hat{\gamma}. \tag{35}$$

The optimal value of $\hat{\lambda}_1$ can be estimated by solving the fixed point of the equation

$$\frac{1}{1 + \hat{\gamma}}\frac{S^T M(\lambda_1, \hat{\gamma})^T S}{S^T M(\lambda_1, \hat{\gamma})^T M(\lambda_1, \hat{\gamma})S} = \lambda_1. \tag{36}$$

From above derivation, we note that, given the value of λ_2, the optimal values of the parameters $\hat{\alpha}(t)$, $\hat{U}$, $\hat{\lambda}_1$, and $\hat{\gamma}$ in (19) can be estimated. However, the optimal value of λ_2, which regularizes the parameters of the operator [see (17)], cannot be estimated by the proposed procedure. Instead, one can try several values for λ_2 and choose the one that gives the best estimation of the other parameters. However, in practice, we have found that the solution of (19) is insensitive to the value of λ_2 (see Section IV). Thus, its value can be fixed.

C. Null Space Pursuit Algorithm

We summarize the derivations in Sections III-A and B to solve the optimization of (19) in the following algorithm.

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 $j = 0$, $\hat{U}_j = 0$, $\lambda_1^j = \lambda_1^0$, and $\gamma^j = \gamma^0$.

Step 3) Compute $\hat{\alpha}_j$ according to (22) as follows:

$$\hat{\alpha}_j = -\left(\left(A_j^TA_j + \hat{\lambda}_2 D^TD\right)^{-1}A_j^TD(S - \hat{U}_j)\right) \tag{37}$$

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

Step 4) Compute λ_1^{j+1} according to (36) as follows:

$$\lambda_1^{j+1} = \frac{S^T M\left(Q_j, \lambda_1^j, \gamma^j\right)^T S}{(1 + \gamma^0)S^T M\left(Q_j, \lambda_1^j, \gamma^j\right)^T M\left(Q_j, \lambda_1^j, \gamma^j\right)S} \tag{38}$$

where $M\left(Q_j, \lambda_1^j, \gamma^j\right) = \left(Q_j^TQ_j + (1 + \gamma^j)\lambda_1^j I\right)^{-1}$ and $Q_j = D + P_{\hat{\alpha}_j}$, $P_{\hat{\alpha}_j}$ is a diagonal matrix whose diagonal elements are equal to $\hat{\alpha}_j$.

Step 5) Compute $\hat{U}_{j+1}$ according to (23) as follows:

$$\hat{U}_{j+1} = \left(Q_j^T Q_j + (1+\gamma^j)\lambda_1^{j+1} I\right)^{-1} \times \left(Q_j^T Q_j S + \gamma^j \lambda_1^{j+1} S\right). \quad (39)$$

Step 6) Compute γ^{j+1} according to (32) as follows:

$$\gamma^{j+1} = \frac{(S-\hat{U}_{j+1})^T S}{\|S-\hat{U}_{j+1}\|^2} - 1. \quad (40)$$

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

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

When the algorithm terminates, the component that is removed by the process is $S - \hat{U} = (1/1+\hat{\gamma})\tilde{V}$ (see (31)). According to our signal model (see (24)), the component $\tilde{V}$ of S is in the null space of the optimal operator $D + P_{\hat{\alpha}}$. Hence, our algorithm keeps a

$$\tilde{V} - \frac{1}{1+\hat{\gamma}}\tilde{V} = \frac{\hat{\gamma}}{1+\hat{\gamma}}\tilde{V} \quad (41)$$

portion of $\tilde{V}$ in the residual signal $\hat{U}$. In fact, the leakage parameter γ controls the greediness when a component is removed from the null space of the operator in a signal. In [14], the value $\hat{\gamma} = 0$ corresponds to an extremely greedy case because the operator $D + P_{\hat{\alpha}}$ removes the signal $\tilde{V}$ completely. On the other hand, if $\hat{\gamma} \approx \infty$, then all of $\tilde{V}$ is retained in the residual signal. This corresponds to an extremely lazy case because the operator $D + P_{\hat{\alpha}}$ does not remove any information. Thus, we enforce the $\hat{\gamma}$ value in the segment between 0 and 1, which corresponds to the operator removing at least 1/2 of $\tilde{V}$. This constraint can be implemented in Step 6 of the algorithm by setting any γ^{j+1} value outside the segment to be the same as the value at the closest end point of the segment.

IV. Implementation and Experiment Results

In this section, we consider some implementation issues of the proposed algorithm and present the results of experiments on various signals.

Every point in a signal has an α value. The optimal value of parameter $\hat{\alpha}$ is estimated in Step 3 of the proposed algorithm. The step uses (37) to estimate all the α values of the signal simultaneously in each iteration. However, we found that the estimation is unstable at points where $(S-\hat{U}_j)$ is very small [15]. Thus, in our implementation, we use the following method to estimate the α values. At each point t, we select a neighborhood B_t of the point and calculate

$$\hat{\alpha}_j(t) = -\frac{c_t^T d_t}{c_t^T c_t + \lambda_2} \quad (42)$$

where c_t is the restriction of $(S-\hat{U}_j)$ on the interval B_t; d_t is the second difference of $(S-\hat{U}_j)$ on B_t; and the parameter λ_2 ensures that the denominator is not zero. In all the experiments, the neighborhood B_t is 31 points for any t. We also impose a constraint to resolve the difficulty that arises due to the unstable estimation of α in some signals. Specifically, we constrain the support of the spectrum of the extracted component $\hat{V}$ to be in a subset of the spectrum of the signal S. This constraint corresponds to the projection of the spectrum of $\hat{V}$ into the support of the spectrum of S. It is straightforward to incorporate the constraint into our algorithm. After Step 5, we compute the spectrum of $S-\hat{U}$, and remove the part that does not appear in the support of the spectrum of S to obtain the modified spectrum of $S-\hat{U}$. The residual signal is obtained by applying the inverse FFT on the modified spectrum of $S-\hat{U}$. Our experiments show that incorporating the constraint in the implementation makes very little difference to the signal separation results. However, incorporating it can reduce the variation of the estimated instantaneous frequency when a signal's SNR is low.

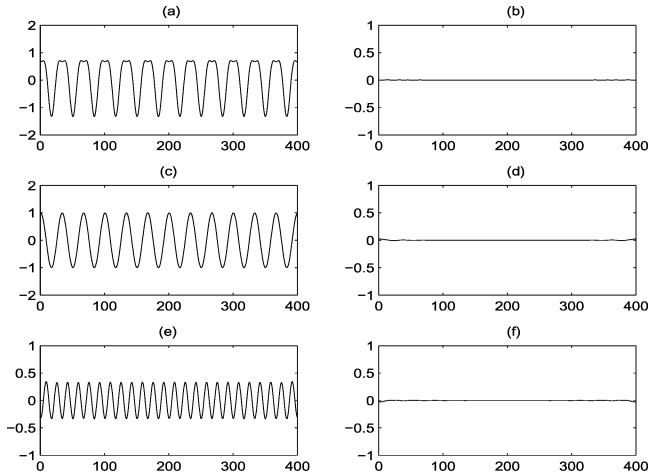

Fig. 2. Using the NSP algorithm to decompose $\cos(4t) - (1/3)\cos(8t)$. Top left: The input signal. Top right: the residual signal after the first and second subcomponents are extracted from the input signal. Middle left: the first extracted component. Middle right: the error signal obtained by subtracting 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 $-(1/3)\cos(8t)$. We use $\lambda_1^0 = 1e-6$ as the initial value to extract both subcomponents. (a) Original signal. (b) The residual. (c) Extracted first component. (d) Difference with real component. (e) Extracted second component. (f) Difference with real component.

The threshold ϵ can be set as low as $1e{-}7$ for most signals, but the algorithm cannot be stopped if the noise is high. Thus, for a noisy signal, we set the threshold value at $1e-3$. Although the Lagrangian multiplier λ_1 can be estimated adaptively, the initial value of λ_1^0 may affect the results. Our experiment results show that, for a broad range of initial λ_1^0 values (from 0.01 down to 0.0000001), the algorithm decomposes a signal well. Different initial values in the range affect the convergence rate of the decomposition. The parameter λ_2 is insensitive to various signals, so its value is fixed. In all our experiments, $\hat{\lambda}_2$ is set at 0.0001 and the initial value of γ^0 is set at 1.

In the following, we provide some examples to demonstrate the results achieved by our algorithm when decomposing various signals. The first example shows that the algorithm can separate harmonic signals. We separate the signal $S(t) = \cos(4t) - (1/3)\cos(8t)$ into two subcomponents, $\cos(4t)$ and $-(1/3)\cos(8t)$. The extracted subcomponents and the residual signals are shown in Fig. 2. It is interesting to note

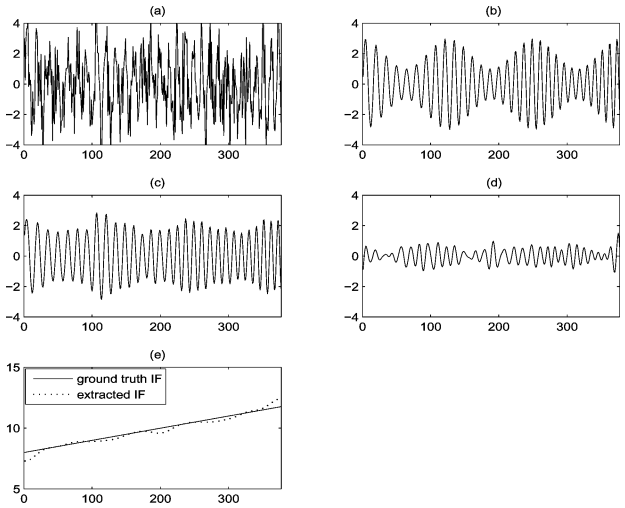

Fig. 3. Removing the noise from a noisy chirp signal. Top left: the noisy chirp signal, which has an SNR of -0.12 dB. Top right: the clean chirp signal. Middle left: the extracted subcomponent, which has an SNR of 9 dB. Middle right: the error signal obtained by subtracting the extracted component from 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. (a) Noisy signal. (b) Clean signal. (c) Extracted first component. (d) Difference with real component. (e) Extracted instantaneous frequency.

that the first extracted component is the low frequency subcomponent of $S(t)$. By contrast, in the approach in [14], which uses the extrema to estimate the instantaneous frequency, the first extracted component is the high frequency subcomponent of $S(t)$.

In the second example, we show that the proposed Null Space Pursuit (NSP) algorithm can remove the noise from a noisy chirp signal. We experimented by embedding the chirp signal $s(t) = (2+\cos(t))\sin(8t+0.1*t^2)$ in additive Gaussian random noise so that the noisy chirp signal's SNR was -0.12 dB, as shown in Fig. 3. The decomposition results derived by the NSP algorithm are also shown in the figure. The SNR of the extracted subcomponent is 9 dB, which means there is a 9 dB gain in the noise reduction of this signal. The estimated instantaneous frequency and the correct instantaneous frequency are superimposed in the bottom left subfigure of Fig. 3. The error is relatively large at the beginning and end of the signal because of the boundary effect in our process.

The method in [14] can only be used to estimate the instantaneous frequency of an oscillatory signal. For a non-oscillatory signal, the parameters of the proposed operator cannot be derived from the extrema of the signal. In contrast, our algorithm can estimate the parameters of the proposed operators, even when the signal is not oscillatory. This is because we use a variational approach to estimate the parameters [see (22)]. Fig. 4 shows that our algorithm can remove the noise from a piecewise smooth signal. Because of the regularization of our approach, the extracted component is a smooth function. When compared to the original signal, the maximal error appears as a singularity, as shown in the bottom-right subfigure of Fig. 4.

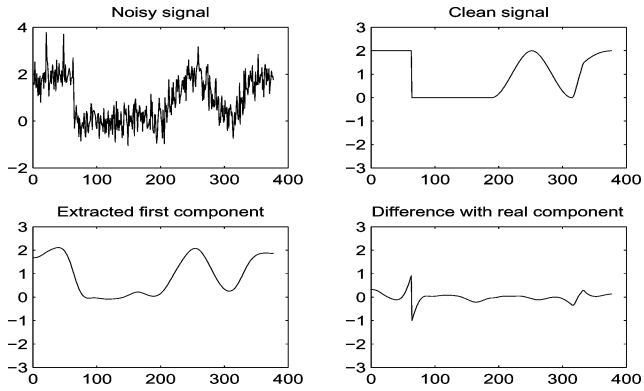

Fig. 4. Extracting the piecewise smooth component from a noisy signal. Top left: the noisy signal. Top right: The clean piecewise smooth signal. Bottom left: The extracted component. The singularities are oversmoothed by our approach. Bottom right: the error signal obtained by subtracting the extracted component from the clean signal. The maximal errors occur at the singularities. The initial value of λ_1^0 is 0.00001.

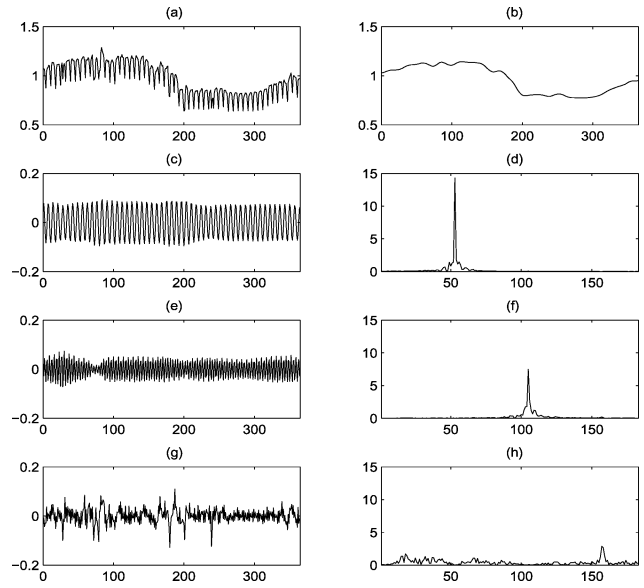

Fig. 5. The decomposition of Poland's electricity consumption signal. Top left: The signal. Top right: The trend of the signal (the first extracted subcomponent). Left-hand rows 2, 3, and 4: the second, third, and residual subcomponents respectively. The spectrum of each extracted component is shown in the right-hand subfigure of the corresponding row. The unit of the horizontal axes in the right-hand subfigures is $2\pi/365$ and the initial values of λ_1^0 for the extraction of the first, second, and third subcomponents are set at 0.001, 0.01, and 0.1 respectively. The second and third extracted subcomponets relate, respectively, to a one-week cycle and a half-week cycle, which might correlate with people's working patterns over a week. (a) Original signal. (b) First component. (c) Second component. (d) Spectrum of second component. (e) Third component. (f) Spectrum of third component. (g) Residual signal. (h) Spectrum of residual.

We also study two real-life signals. The first is Poland's daily electricity consumption in 1990 [10]. Fig. 5 shows the decomposition results derived by our algorithm. A similar decomposition of oscillatory components is reported in [14], and the trend component is estimated by calculating the local mean of the signal [16]. The optimal value of the Lagrangian parameter in [14] is selected manually by trial and error. In contrast, all the parameters of our decomposition (except $\hat{\lambda}_2$) are estimated

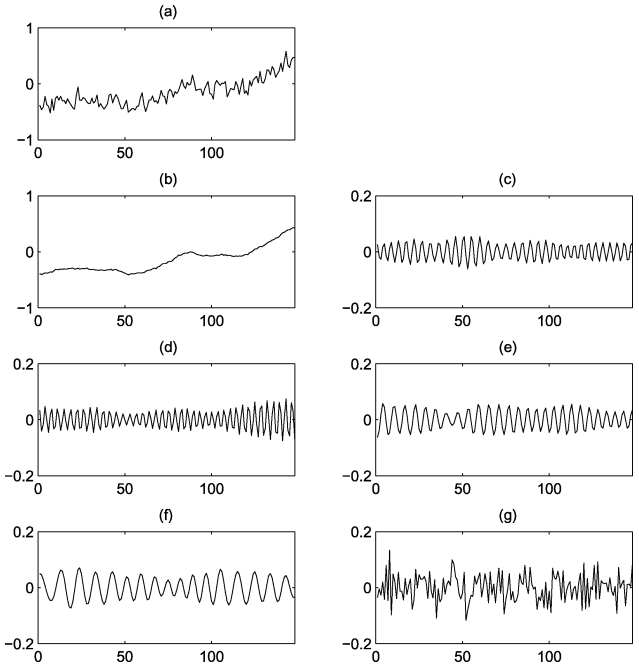

Fig. 6. Top left: The signal of the anomaly in the annual mean global surface temperature. The first five components and the residual signal are shown in subfigures (b) to (g), respectively. The third component, plotted in subfigure (d), is a chirp signal with a higher variation than the others. Note that the chirp signal decreases and then increases as the year progresses. The initial values of λ_1^0 for the extraction of the first, second, third, fourth, and fifth subcomponents are set at 0.001, 0.001, 0.01, 0.01, and 0.1, respectively. When the NSP algorithm stops, these values are 0.0018, 0.0023, 0.0081, 0.0030, and 0.0059, respectively. Note that the first extracted component of the NSP algorithm is the local mean of the input signal. (a) The input signal. (b) Extracted first component. (c) Extracted second component. (d) Extracted third component. (e) Extracted fourth component. (f) Extracted fifth component. (g) The residual signal.

adaptively from the signal and the residual signals of the decomposition.

The last example is the anomaly in the annual mean global surface temperature reported in [21]. The purpose of processing the signal is to decompose it into additive subcomponents and then observe the anomaly of variations in the global temperature from the subcomponents. Fig. 6 shows the decomposition results derived by the NSP algorithm. The first subcomponent extracted by the algorithm is the trend. The energy of the subcomponents extracted after the fifth subcomponent is extremely small; therefore, we do not show them. The third subcomponent is of the most interest because the variation in the temperature is relatively high compared to that in the other subcomponents. In addition, the frequency of the third subcomponent decreases and then increases as the year progresses. The fifth subcomponent is almost a periodic signal over a period of ten years. It might be correlated with the periodic fluctuations in the length of a day (LOD) over a decade period in [23] and [24]. The results obtained by the EMD decomposition approach are reported in [21] and [22]. To compare them with the results of the NSP algorithm, we show the first five subcomponents of the Ensemble Empirical Mode Decomposition (EEMD) algorithm in Fig. 7. Clearly, decomposition results derived by the NSP and EMD algorithms are different. It would therefore be very interesting to investigate the correlation between the decomposed data derived by the two methods.

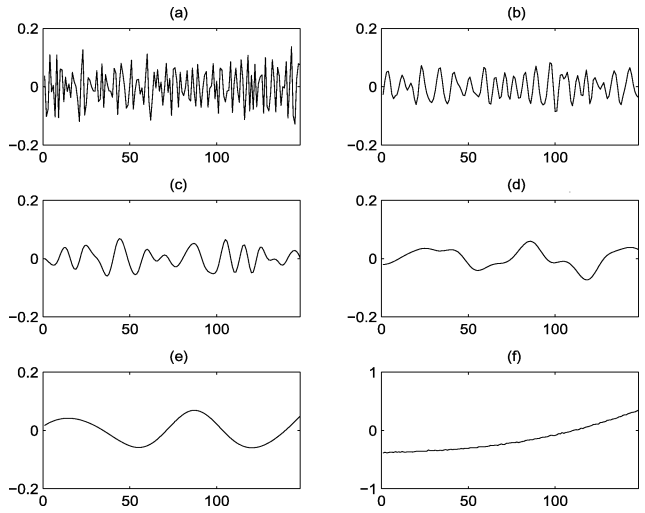

Fig. 7. Separation results of the global surface temperature data [see subfigure (a) of Fig. 6] derived by the EEMD algorithm. The codes of the algorithm can be found in [21]. Subfigures (a), (b), (c), (d), (e), and (f) are the extracted first, second, third, fourth and fifth components and the residual signal respectively. (a) Extracted first component. (b) Extracted second component. (c) Extracted third component. (d) Extracted fourth component. (e) Extracted fifth component. (f) The residual signal.

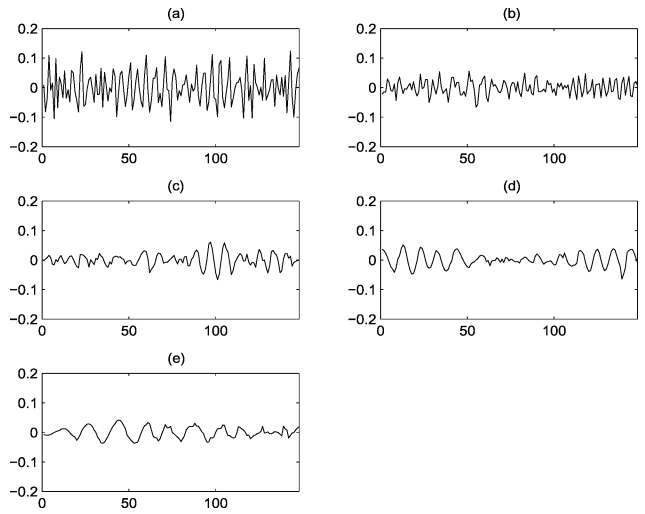

Fig. 8. Separation results of the global surface temperature data (see subfigure (a) of Fig. 6) derived by using the separation algorithm in [14]. Subfigures (a), (b), (c), (d), and (e) are the extracted first, second, third, fourth, and fifth components respectively. The λ values used by Algorithm08 to extract the first, second, third, fourth, and fifth components are taken from the final λ_1 values of the second, third, fourth, fifth, and sixth components derived by the NSP algorithm. The λ values of Algorithm08 are 0.0023, 0.0081, 0.0030, 0.0059 and 0.0093 for the extraction of the first, second, third, fourth, and fifth components, respectively (see the caption of Fig. 6). (a) Extracted first component. (b) Extracted second component. (c) Extracted third component. (d) Extracted fourth component. (e) Extracted fifth component.

Finally, we compare the results of the NSP algorithm with those of the algorithm in [14]. Since the parameters of the latter algorithm (called Algorithm08 for convenience) must be determined manually, we use the optimal values of λ_1 obtained by the NSP algorithm as the corresponding λ values of Algorithm08. First, to obtain the detrend signal, we remove the trend of the global surface temperature signal by subtracting the local mean subcomponent obtained by the NSP algorithm from the signal. Fig. 8 shows the results of applying Algorithm08 to extract the

first five subcomponents of the detrend signal. Since the corresponding λ and λ_1 values of Algorithm08 and the NSP algorithm are the same, the differences in the separation results are mainly due to the parameter γ in the third term of (6), which determines the amount of $S-U$ to be retained in the null space of the operator $\mathcal{T}_s$, and the estimated parameters of $\mathcal{T}_s$. Algorithm08 solves the optimization problem in (3), which does not have the γ parameter, by estimating the operator from the extrema of a signal. The separation results of the NSP algorithm and Algorithm08 are also different.

V. Conclusion

We have proposed an approach that uses an adaptive operator to separate a signal into additive subcomponents. Basically, we generalize the original operator-based approach by achieving better control of the amount of information to be removed from the null space of the optimal operator in the signal. We show that, under our approach, the operators' parameters as well as the Lagrangian multipliers can be estimated adaptively. This overcomes the difficulties encountered when implementing the original approach on real-life signals. We compare the proposed NSP approach to the MP approach and show that it is a generalization of the MP approach. In addition, we provide several examples, including real-life signals, to demonstrate the separation results derived by our algorithm. In our future work, we will investigate various issues, such as developing operators to preserve singularities and extending the method to images.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable comments. They would also like to thank X. Hu, Institute of Automation, Chinese Academy of Sciences, who prepared Figs. 7 and 8.

References

[1] E. Candes and M. B. Wakin, "An introduction to compressive sensing," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.
[2] D. L. Donoho and M. Elad, "Optimally sparse represenation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, pp. 2197–2202, 2003.
[3] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, 1993.
[4] X. Hu, S. L. Peng, and W. L. Hwang, "Estimation of instantaneous frequency parameters of the operator-based signal separation method," *Adv. Adapt. Data Analy.*, vol. 1, no. 13, pp. 1–14, 2009.
[5] J. Bobin, J.-L. Starck, J. Fadili, Y. Moudden, and D. Donoho, "Morphological component analysis: An adaptive thresholding strategy," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2675–2681, Nov. 2007.
[6] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.
[7] J. Garnett, T. Le, Y. Meyer, and L. Vese, "Image decompositions using bounded variation and generalized homogeneous Bosov spaces," *Appl. Comput. Harmon. Anal.*, vol. 23, pp. 25–56, 2007.
[8] R. Carmona, W. L. Hwang, and B. Torresani, "Multiridge detection and time-frequency reconstruction," *IEEE Trans. Signal Process.*, vol. 47, no. 2, pp. 480–492, 1999.
[9] L. Rudin, S. Osher, and E. Fatemi, "Nonliner total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.
[10] A. C. Harvey and T. Trimbur, "General model-based filters for extracting cycles and trends in economic time series," *Rev. Econom. Statist.*, vol. 85, no. 2, pp. 244–255.
[11] N. E. Huang, Z. Shen, S. R. Long, M. L. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis," *Proc. R. Soc. Lond. A*, vol. 454, pp. 903–995, 1998.
[12] Y. Meyer, *Oscillating Patterns in Image Processing and Non-Linear Evolution Equations*, ser. University Lecture Series. Providence, RI: AMS, 2002, vol. 22.
[13] E. Tadmor, S. Nezzar, and L. Vese, "Multiscale hierarchical decomposition of images with applications to deblurring, denoising and segmentation," *Commun. Math. Sci.*, vol. 6, no. 2, pp. 281–307, 2008.
[14] 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.
[15] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*. Cambridge, U.K.: Cambridge Univ. Press, 1992.
[16] D. S. G. Pollock, "Methodology for trend estimation," *Econom. Model.*, vol. 18, no. 1, pp. 75–96, 2001.
[17] G. Rilling and P. Flandrin, "One or two frequencies? the empirical mode decomposition answers," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 85–95, 2008.
[18] J. Starck, M. Elad, and D. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Trans. Image Process.*, vol. 14, no. 10, pp. 1570–1582, Oct. 2005.
[19] L. Vese and S. Osher, "Modeling textures with total variation minimization and oscillating patterns in image processing," *J. Sci. Comput.*, vol. 19, pp. 553–577, 2003.
[20] Z. Wu and N. E. Huang, "A study of the characteristics of white noise using the empirical mode decomposition method," *Proc. R. Soc. London*, vol. 460, pt. Ser. A, pp. 1597–1611, 2004.
[21] Research Center and Adaptive Data Analysis, National Central Univ. [Online]. Available: http://rcada.ncu.edu.tw/research1_clip_ex.htm
[22] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: A noise-assisted data analysis method," *Adv. Adapt. Data Anal.*, vol. 1, no. 1, pp. 1–41, 2009.
[23] Q. L. Wang, Y. T. Chen, D. X. Cui, W. P. Wang, and W. F. Liang, "Decadal correlation between curstal deformation and variation in length of day of the earth," *Earth Planets Space Lett.*, vol. 52, pp. 989–992, 2000.
[24] A. Paris and G. Hulot, "Length of day decade variations, torsional oscillations and inner core superrotation: Evidence from recovered core surface zonal flows," *Phys. Earth Planet. Int.*, vol. 118, pp. 291–316, 2000.

Silong Peng was born in Sixian, Anhui, in 1971. He received the B.S. degree in mathematics from Anhui University, Hefei, China, in 1993 and the M.S. and Ph.D. degrees in mathematics from the Institute of Mathematics, Chinese Academy of Sciences (CAS), Beijing, in 1995 and 1998 respectively.

From 1998 to 2000, he worked as a Postdoctoral in the Institute of Automation, CAS. In 2000, he became a Full Professor of signal processing and pattern recognition in the Institute of Automation, CAS. His research interests include wavelets, multirate signal processing, and digital image processing.

Wen-Liang Hwang received the B.S. degree in nuclear engineering from the National Tsing Hua University, Hsinchu, Taiwan, R.O.C., the M.S. degree in electrical engineering from the Polytechnic Institute of New York, and the Ph.D. degree in computer science from New York University in 1993.

He was a Postdoctoral Researcher with the Department of Mathematics, University of California, Irvine, in 1994. In January 1995, he became a member of the Institute of Information Science, Academia Sinica, Taipei, Taiwan, where he is currently a Research Fellow. He is a coauthor of the book *Practical Time-Frequency Analysis* (Academic, 1998). His research interests include wavelet analysis, signal and image processing, and multimedia compression and transmission.

Dr. Hwang currently is an Associate Editor of the *Journal of Wavelet Theory and Applications* and the *International Journal of Wavelets, Multiresolutions and Information Processing*. He was awarded the Academia Sinica Research Award for Junior Researchers in 2001.