

スパース信号表現とその音声・画像処理への応用

中 静 真

大阪大学大学院基礎工学研究科

e-mail: nkszk@sys.es.osaka-u.ac.jp

1 はじめに

近年、信号処理の分野において、信号の次元数を超えた基底系を与えて信号を表現するスパース信号表現が注目を集めている。画像符号化で用いられる画像表現は、情報圧縮を目的として、画像を画像のサンプル数と同数の変換係数で表現する。この変換係数を得るために、離散コサイン変換、ウェーブレット変換を代表として、さまざまな変換法および変換基底が提案されてきた。さらに、信号群に対して、二乗誤差の意味で最良な基底系が得られる主成分分析も、特徴量の次元圧縮を目的としてパターン認識で用いられている。これら信号処理分野で用いられている信号表現は、信号のサンプル数と同数の基底ベクトルの一次結合で信号を表現する。スパース表現では、信号の次元を超える数の基底ベクトルを、信号表現のために用意する。スパース表現は、一つの信号に対してできるだけ少ない数の基底ベクトルを選択し、それらの一次結合で信号表現を行う。基底ベクトルに対する係数を要素とする係数ベクトルの中で、非零要素が疎らに発生することから、スパース性に基づく冗長な基底系による信号分解を、スパース信号分解と呼んでいる。スパースな係数ベクトルを与える方法と共に、信号群に対してスパースな表現を与える基底系の設計方法も研究されている。これら基底系の設計法は、スパースコーディングと呼ばれ、信号の特徴を記述する基底系を得ることができる。本稿では、まずスパース信号分解とその方法を概説し、次に基底系を設計するためのスパースコーディングについて解説する。その後、著者らが実現したシフト不変スパースコーディングと、音声信号の分離、解析のために提案したスパース周期信号分解について解説する。最後に以上をまとめ、今後の展開について述べる。

2 スパース信号分解

N 次元信号 $\mathbf{f} \in \mathbb{R}^N$ の線形生成モデルとして、係数ベクトル $\mathbf{s}$ と行列 $\mathbf{A}$ から

$$\mathbf{f} = \mathbf{A}\mathbf{s} \quad (1)$$

を与える。Aは、信号を表現するために用意された基底ベクトルから構成される行列であり、

$$\mathbf{A} = (\mathbf{a}_1 \cdots \mathbf{a}_M) \quad (2)$$

とする。信号表現や信号分解の分野では、Aを辞書(dictionary)と呼び、 $\mathbf{a}_i$ を原子(atom)と呼ぶことが多い。本稿では、前者を基底行列、後者を基底ベクトルと呼ぶ。

画像処理、音声処理で広く用いられ直交(双直交)変換等では、基底ベクトルの数Mと信号の次元Nは一致し、信号fに対して係数ベクトルsを一意に決定することができる。基底ベクトルの数Mが信号の次元数Nを超え、かつ、N個の互いに一次独立な基底ベクトルが含まれている場合、信号fを表現するための基底ベクトルが過剰に与えられていることになる。この場合、基底行列は過完備(overcomplete)となり、係数ベクトルは一意に定まらない。そのために、係数ベクトルsに対するペナルティを与え、望む性質を持つsを求める。本稿では、ペナルティとして l_p ノルム

$$\|\mathbf{s}\|_p = \left(\sum_n |s_n|^p \right)^{\frac{1}{p}} \quad (3)$$

を与える。ここで $p \geq 0$ である。 s_n は、係数ベクトルsのn番目の要素である。 l_p ノルムの中で、最もよく知られたペナルティは、 l_2 ノルムである。信号の再構成を保障し、かつ、最小の l_2 ノルムを持つ係数ベクトルは、最小ノルム解と呼ばれ、一般化逆行列の解として解析的に計算することができる。一方、多くの画像・信号処理では、情報圧縮を目的として、できるだけ少ない数の基底ベクトルで信号を表現することが望まれる。信号近似に使用する基底ベクトルの個数を最小化する問題は、 $p = 0$ とした l_0 ノルムをペナルティとして与えることで

$$\min \|\mathbf{s}\|_0 \quad \text{subject to } \mathbf{f} = \mathbf{A}\mathbf{s} \quad (4)$$

と表現することができる。 l_0 ノルムの最小化は、係数ベクトルの中の非零の要素数の最小化と一致し、この問題の解で信号を最小の基底数で表現することができる。しかしながら、この問題の解を得るためには、基底ベクトルの1からN個までの組み合わせ順に探索する必要があり、現実的ではない。そこで近似的に l_0 ノルムを最小化する係数ベクトルを求める方法が、いくつか提案されてきた。疎らに非零要素が分布する係数ベクトルによって信号を表現する方法として、これらの方法は、スパース信号分解と呼ばれている。

代表的な方法の一つは、基底ベクトルを繰り返し一つずつ選んで、信号分解を行う方法である。これはMatching pursuits(MP)[1][2]と呼ばれ、さまざまな信号解析、動画像、静止画像の圧縮符号化[3]等へ応用されている。MPでは、 l_2 ノルムが正規化されている基底ベクトルを用いて、反復回数 $i = 0$ から次の反復を繰り返す。

[MPのアルゴリズム]

- 1) 初期残差信号 $\mathbf{R}^{(0)} = \mathbf{f}$ 、初期係数ベクトルsの要素をすべて0に設定する。
- 2) 残差 $\mathbf{R}^{(i)}$ と内積が最大となる基底ベクトル $\mathbf{a}_{j_{\max}}$ を検索し、係数ベクトルの $j_{\max}$ 番目の要素を $s_{j_{\max}} \leftarrow \mathbf{a}_k^T \mathbf{R}^{(i)}$ と更新する。ここで肩符Tは、ベクトルおよび行列の転置を示す。

3) 残差信号を $\mathbf{R}^{(i+1)} \leftarrow \mathbf{R}^{(i)} - s_{j_{\max}} \mathbf{a}_{j_{\max}}$ と更新する.

4) 終了条件を満たさない場合, 2) へ進む.

以上の反復では, N 個の基底ベクトルが選択されても反復が終了するとは限らない. そこで, 反復毎に, 残差信号を既選択された基底ベクトルが張る部分空間に射影し, 選択された基底ベクトルに対応する係数を修正する方法が提案されている. これは直交 Matching Pursuits(OMP)[1][2] と呼ばれ, スパース信号分解法として広く利用されている. 上記の方法は, 逐次的に信号分解に用いる基底ベクトルを探索する方法であるが, 最小化問題 (4) を緩和し, $0 < p < 2$ の範囲で l_p ノルムを定義して信号分解を行う方法が提案されている. $p < 1$ では, l_p ノルムは凹関数となり, その最小化は困難である. そこで $p = 1$ とした l_1 ノルムの最小化が, スパース信号表現では良く用いられる. $p = 1$ とすることで, (4) は

$$\min \|\mathbf{s}\|_1 \quad \text{subject to } \mathbf{f} = \mathbf{A}\mathbf{s} \quad (5)$$

となる. この問題は, Basis Pursuit(BP)[1][4] と呼ばれている. 上記の問題は, 線形計画問題へ変換できる [4]. しかしながら, 画像や音声を基底ベクトルへ分解することを考えた場合, 求めたい係数の数は, 信号のサンプル数を超えた数となる. 画像や音声の処理で, リアルタイムに近い処理を望む場合, 線形計画法の適用は膨大な計算が必要となるために, 実用的ではない. そこで, ラグランジュ乗数を導入し, (5) を無拘束の最小化問題

$$\hat{\mathbf{s}} = \frac{1}{2} \|\mathbf{f} - \mathbf{A}\mathbf{s}\|_2^2 + \lambda \|\mathbf{s}\|_1 \quad (6)$$

へ緩和することで近似解を得る方法 [4][5] が広く用いられている. これは Basis Pursuits Denoising(BPDN)[4] と呼ばれ, 信号に雑音発生を考慮した BP に相当する. ここで BPDN を事後確率最大化の観点から考える [7]. $\mathbf{s}$ を確率ベクトルと考え, それぞれの要素 s_n が, 独立にラプラス分布に従って発生することを仮定する. 仮定により $\mathbf{s}$ の確率密度は

$$p(\mathbf{s}) \propto \prod_n e^{-\frac{\sqrt{2}}{\sigma} |s_n|} \quad (7)$$

と与えられる. 雑音 $\mathbf{e} = \mathbf{f} - \mathbf{A}\mathbf{s}$ が, それぞれの要素で, 独立にガウス分布に従って発生することを仮定すると, $\mathbf{A}\mathbf{s}$ が与えられた上での $\mathbf{f}$ の確率密度は

$$p(\mathbf{f}|\mathbf{A}\mathbf{s}) \propto \prod_n e^{-\frac{1}{2\sigma_f^2} \|\mathbf{f} - \mathbf{A}\mathbf{s}\|_2^2} \quad (8)$$

と与えられる. ベイズの定理から

$$p(\mathbf{s}|\mathbf{f}, \mathbf{A}) \propto \frac{p(\mathbf{f}, \mathbf{A}\mathbf{s})p(\mathbf{s})}{p(\mathbf{f}, \mathbf{A})} \quad (9)$$

が導かれる. ここで $p(\mathbf{f}, \mathbf{A})$ を一定とし, 両辺の対数を求めると

$$\log(p(\mathbf{s}|\mathbf{f}, \mathbf{A})) \propto -\frac{1}{2\sigma_f^2} \|\mathbf{f} - \mathbf{A}\mathbf{s}\|_2^2 - \frac{\sqrt{2}}{\sigma} \|\mathbf{s}\|_1 \quad (10)$$

が導かれる。したがって、BPDNは $\lambda = \sqrt{2}\sigma_r^2/\sigma$ とした場合に、(10)で示した対数尤度を最大化する係数ベクトルを求めることに一致する[7]。ここでは係数分布にラプラス分布を仮定しているが、係数分布に対してもガウス分布を仮定すれば、 l_2 ノルムを最小化する係数ベクトルが事後確率を最大化することになる。この係数ベクトルの導出は、リッジ回帰と呼ばれる正則化項を加えた回帰分析と一致する。

BPDNの解法として、Chenらは内点法を用いている。また、反復再重み付け最小二乗法(IRLS)により、BPDNを最小二乗解の解法を利用して解くことも可能である。本稿では、高速にBPDNの解を得ることができ、画像処理等、大量のサンプル点を持つ信号に対しても適用が容易なBlock Coordinate Relaxation(BCR)法について説明する。実際の信号処理には、ウェーブレット[1]およびウェーブレットパケット基底[1]、離散コサイン基底等を組み合わせて用いることが考えられる。基底系が直交基底系の和で表現されている場合に、Sardyらは、1回の反復毎に一つの直交基底系に対する係数のみを更新するアルゴリズムを提案した[5]。これはBlock Coordinate Relaxation(BCR)法[5]として知られている。もし、基底行列 $\mathbf{A}$ が直交行列である場合には、BPDNの解は解析的に与えられる。基底行列が直交ウェーブレット基底である場合、その解析解の導出は、ウェーブレット雑音除去の分野でソフト縮退(Soft shrinkage)[1][4]と呼ばれる。Sardyらは、一つの直交行列に対してBPDNの解がソフト縮退で得られることを利用することで、BPDNの解を反復的に更新する方法を提案している。ここで基底行列が2つの直交行列 $\mathbf{U}, \mathbf{V}$ から構成されているとする。すなわち、 $\mathbf{A} = (\mathbf{U}, \mathbf{V})$ である。また、 $\mathbf{U}, \mathbf{V}$ に対応して、係数ベクトルも $\mathbf{s}_U, \mathbf{u}_V$ と分割する。BCR法では、以下のアルゴリズムに従い、2つの直交行列に対応する係数ベクトルを交互に更新する。

[BCR アルゴリズム]

- 1) 残差信号 $\mathbf{r}$ を $\mathbf{f}$ に設定する。
- 2) $\mathbf{s}_U$ を $\mathbf{U}^T \mathbf{r}$ のソフト縮退の結果で更新する。更新された係数で、残差信号 $\mathbf{r}$ を $\mathbf{r} - \mathbf{U} \mathbf{s}_U$ で更新する。
- 4) $\mathbf{s}_V$ を $\mathbf{V}^T \mathbf{r}$ のソフト縮退の結果で更新する。更新された係数で、残差信号 $\mathbf{r}$ を $\mathbf{r} - \mathbf{V} \mathbf{s}_V$ で更新する。
- 6) 終了条件を満たさな場合、2)へ進む。

上記の方法では、直交行列として、その行列を使った変換に高速演算法が知られている場合には、基底ベクトルそのものを記録する必要はない。したがって、いくつかの直交変換アルゴリズムを繰り返し実行することで、BPDNの解を得ることができる。オリジナルのBCRアルゴリズムでは、 λ は反復を通じて一定としている。また、基底行列は直交行列の和に限定されている。これを拡張し、参考文献[11][14]では、 λ を反復毎に定数だけ減少させて0に収束させている。さらに、直交行列の他に変換、逆変換を実現する過剰系を許して基底行列を構成している。参考文献[14]では、この方法をモルフォロジカル成分分析と呼び、信号分離法として提案している。モルフォロジ成分分析では、2つの信号、一つは $\mathbf{U}$ でスパースに表現できる信号、他は $\mathbf{V}$ でスパースに表現できる信号の混合信号を、2つ

の係数ベクトルへ分解することができる。画像処理の例では、Curvelet 変換と離散コサイン変換を用いて、画像をテクスチャ成分とエッジから構成されるカートゥーン成分に分離した例がある [14]。また、BCR 法と同じく、ソフト縮退に基づき、さらに収束速度を改善した最適化法が数多く提案されている [19]。

3 スパースコーディング

前章では、基底行列 $\mathbf{A}$ と信号 $\mathbf{f}$ から、非零要素が少ない係数ベクトル $\mathbf{s}$ を求める方法を解説した。本章では、トレーニング用信号群 $\{\mathbf{f}_i\}_{i=1}^I$ を与えて、それぞれをできるだけ少ない基底ベクトルで近似するための基底行列 $\mathbf{A}$ を学習する方法について説明する。この方法はスパースコーディングと呼ばれ、Olshausen らにより、参考文献 [6] で提案された。スパースコーディングでは、信号群が、統計的に独立に疎らに発生する基底ベクトルの線形和で生成されていると仮定し、独立性とスパース性から、個々の基底ベクトルを推定する。基底ベクトルが、信号の独立して発生する特徴を表すことができることから、画像や音声の特徴記述、分離へ応用が期待できる。参考文献 [6] で提案されたスパースコーディングでは、係数ベクトルのスパース性の尺度に l_1 ノルムを用いる場合、コスト関数

$$E = \frac{1}{2} \sum_i \|\mathbf{f}_i - \mathbf{A}\mathbf{s}_i\|_2^2 + \sum_i |\mathbf{s}_i| \quad (11)$$

を最小化する基底行列 $\mathbf{A}$ と、それぞれの信号 $\mathbf{f}_i$ を近似するための係数ベクトル群 $\{\mathbf{s}_i\}_{i=1}^I$ を導出する。このコスト関数の最小化は、以下の2つのステップ

- 1) 基底行列 $\mathbf{A}$ を固定して、コスト関数 E を最小化する係数ベクトル群 $\{\mathbf{s}_i\}_{i=1}^I$ に更新する。
- 2) 係数ベクトル群 $\{\mathbf{s}_i\}_{i=1}^I$ を固定して、コスト関数 E を最小化する基底行列 $\mathbf{A}$ に更新する。

を交互に収束まで繰り返すことで実現されている。1) の更新は、BPDN の問題と同一である。2) の更新は、信号と近似の間での二乗誤差を最小化する問題であり、二次計画問題となる。参考文献 [6] では、自然画像をブロックに分割してトレーニング画像群を生成し、スパースコーディングを適用している。自然画像に対しては、スパースコーディングにより、ガボール関数に類似した基底系が学習されると報告されている。上記の基底行列の更新では、二乗誤差を最小化する意味で基底行列の更新が行われているが、参考文献 [7] では、仮定した係数ベクトルの分布から、基底行列を最尤推定から求める方法が提案されている。また、上記の2つのステップを繰り返すことで、基底行列を学習する方法として、K-SVD 法 [8] が提案されている。K-SVD 法では、トレーニング画像を、同じ基底ベクトルを用いて近似されるグループへ分類し、それらから特異値分解を用いて基底ベクトルを更新する。K-SVD 法は、ベクトル量子化で用いられる k -平均法の拡張として、捉えることができる。

スパースコーディングの応用の一つとして、多重音で構成された音楽信号の解析の例 [9] がある。この例では、時間領域と周波数領域でのスパースコーディングが音楽信号へ適用されている。時間領域のアプローチでは、さまざまな時刻で発生する音に追従することを目的として、基底行列を基本となるいくつかの関数の時間シフトにより生成している。周波

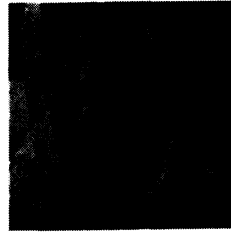


図 1: 原画像 Barbara

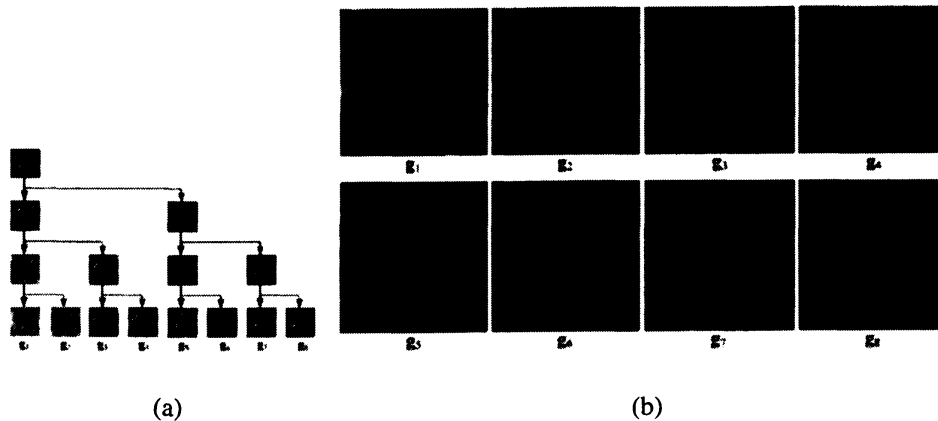


図 2: 原画像から学習された木構造を持つ構成要素とその要素画像

数領域のアプローチでは、音楽信号の振幅スペクトルに対して、基底ベクトルと係数に対して非負拘束を課したスパースコーディングを適用している。いずれの方法でも、スパースコーディングは、複数の音響信号が混合した音楽信号から、個々の音響信号の特徴を捉える基底系を生成することが報告されている。

4 シフト不変スパースコーディング

前節で説明したスパースコーディングの画像への適用では、トレーニング画像を小ブロックへ分割し、ブロックをスパースに表現できる基底行列を導出している。しかしながら、微細な構造が繰り返し現れることで画像が生成されていることを仮定すると、画像の微細構造の発生とブロック分割は独立に起こり、一つの微細構造を一つの基底ベクトルとして表現することはできない。実際に、参考文献 [6][7] で学習された基底系を見ると、ガボール基底 [1] に類似した基底が学習されていることが確認できるが、中心がシフトした複数の類似した基底が発生している。繰り返し現れる微細構造を一つの基本とするベクトルによって表現するために、基底系をいくつかのベクトルの座標シフトによって生成するスパースコーディングが提案されている。これらを、シフト不変性を有するスパースコーディングとして、シフト不変スパースコーディング [10][12][13] と呼ぶ。本節では、著者が参考文献 [12] で提案したシフト不変スパースコーディングの例を説明する。参考文献 [13] では、

画像を構成する微細構造を，マスマティカルモルフォロジによる画像モデルと同じく，構造要素と呼んでいる．シフト不変スパースコーディングでは，基底行列に含まれる基底ベクトルの数が，基本となる構造要素の数と，画像のサイズで定まるシフト座標の総数の積となる．それに対して，通常のスパースコーディングでは，基底行列に含まれる基底ベクトルの数は，近似対象となる信号の次元数の2倍から4倍に設定されている．したがって，シフト不変性を有するスパースコーディングでは，多くの構造要素を仮定した場合，基底ベクトルとそれに対応する係数の数が増大し，演算量，記録容量ともに増加することが問題となる．この問題点を解決するため，参考文献[13]では，画像生成モデルとして，一つの座標で一つの構造要素のみが現れる制限付きのモデルを利用している．このモデルでは，画像 $\mathbf{f}$ が

$$\mathbf{f} = \sum_{(m,n) \in C} \mathbf{T}_{m,n} \mathbf{g}_{\gamma_{m,n}} s_{m,n} + \mathbf{e} \quad (12)$$

と表される．ここで C はシフト座標の総数である． $\{\mathbf{g}_k\}_{k=1}^K$ は， K 個の構造要素を示し，それぞれの構造要素がサポートする領域は画像 $\mathbf{f}$ に比べて十分小さいことを仮定する．具体的には，参考文献[12]においては， 16×16 画素と設定している． $\mathbf{T}_{m,n}$ は，構造要素の座標 (m,n) へのシフト操作を表し， $\gamma_{m,n}$ は座標 (m,n) で発生する構造要素のインデックスを示す． $s_{m,n}$ は，座標 (m,n) で発生する構造要素に対する係数を示す．画像への応用では，画像が非負信号であり，画像を表す構造要素も同じく非負であると仮定し，係数，構造要素はすべて非負としている．また，(14) のモデルでは，係数と構造要素の l_2 ノルムの間で，乗数に関して不定性が生じる．そこで，それぞれの構造要素に対して $\|\mathbf{g}_i\|_2 \leq 1$ の条件を課す．

式(14)で示した画像モデルに対して，スパースに関するペナルティを含むコスト関数

$$Q = \frac{1}{2} \|\mathbf{f} - \hat{\mathbf{f}}\|_2^2 + \lambda \sum_{(m,n) \in C} s_{m,n} \quad (13)$$

を与え，これを減少させながら画像に対して構造要素と係数を導出する．ここで $\hat{\mathbf{f}}$ は，画像 $\mathbf{f}$ の近似を示し

$$\hat{\mathbf{f}} = \sum_{(m,n) \in C} \mathbf{T}_{m,n} \mathbf{g}_{\gamma_{m,n}} s_{m,n} \quad (14)$$

である．式(15)の右辺第1項が近似誤差に関するペナルティ，第2項がスパース性に関するペナルティを示す．

式(15)のコスト関数を減少させる方法として，一般的なスパースコーディングと同じく，係数，構造要素のインデックスの更新と構造要素の更新を繰り返し行う方法が提案されている．係数，構造要素の更新では，BCR法と同じく，繰り返し一つの座標を選択し，その座標における係数と構造要素のインデックスを更新する．この更新では，選択された座標以外の座標における構造要素のインデックスと係数が固定されていると考え，コスト関数を最小化する係数とインデックスに更新を行う．この更新は，選択された座標の構造要素および係数を除いて求められる近似画像と原画像の残差と，内積が最大となる選択された座標へシフトされた構造要素の検索によって，実現することができる．また，構造要素の更新は，構造要素の非負性と l_2 ノルムに関する制約の下で二乗誤差を最小化する構造要素を求める問

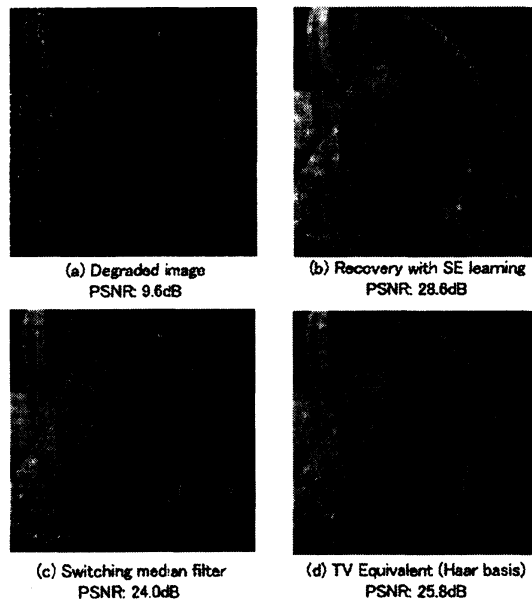


図 3: 画像の欠損補間の例 (a) 劣化画像, (b) 提案法, (c) スイッチングメジアンフィルタ, (d)TV 補間

題であり，凸拘束下での最小二乗法となる．具体的には，勾配の凸空間へ射影と，線形探索を繰り返すことで，コストを最小化する構造要素を求めている．

さて，制限付きの画像生成モデル (14) では，一つの構造要素のみが一つの座標で発生している．そのために，係数の数は構造要素の数に依存せず，画像のサイズのみで決まる．構造要素の更新における勾配の導出，コスト関数の導出に要する演算量は，構造要素の数に依存せず，係数の数で決まる．そのために，提案した制限付きの画像生成モデルでは，構造要素の更新に要する演算量は，画像サイズのみに依存する．また，参考文献 [12] では，構造要素数を一つから順に 2 倍に増加させて学習を繰り返し，構造要素の木構造を作成することを提案している．この木構造を利用し，係数更新における構造要素の探索を，木探索へ変更することで，係数更新における演算量も減少させることができる．

実際に，図 1 に示す画像 *Barbara* から得られる木構造の構造要素群と，それぞれの構造要素に対応する画像成分を図 2 に示す．図 2 では，1 つの構造要素のみから再構成される画像成分を同時に示している．これらの構造要素の中では，コスト関数の減少率により構造要素の増加を打ち切っている．画像は，画像の構造要素の特徴に従い，画像の平坦部，さまざまな周期性，方向性を持つテクスチャ成分へそれぞれ分解されていることがわかる．

次に欠落が発生している画像から，提案法で構造要素と係数を推定し，欠落を補間した例を示す．スパース性に基づく画像補間例として，参考文献 [7][11][8] がある．これらは，スパース信号分解において欠損を受けていない画素をスパース性のペナルティの下で近似することにより，基底ベクトルを用いて欠損を補間する方法である．これらの例では，基底ベクトルとして，欠損の無いトレーニング画像から事前に学習した基底行列，もしくは既

存の変換行列を用いることが前提になっている。提案法では、コスト関数を

$$Q = \frac{1}{2} \|\mathbf{H}\mathbf{f} - \mathbf{H}\hat{\mathbf{f}}\|_2^2 + \lambda \sum_{(m,n) \in C} s_{m,n} \quad (15)$$

と修正し、これを最小化することで欠損のある画像 $\mathbf{H}\mathbf{f}$ から直接、構造要素と補間画像

$$\hat{\mathbf{f}} = \sum_{(m,n) \in C} \mathbf{T}_{m,n} \mathbf{g}_{\gamma_{m,n}} s_{m,n}. \quad (16)$$

を得る。ここで $\mathbf{H}$ は画像に起こった欠損の座標を示す対角行列であり、対角要素が欠損座標で 0、それ以外では 1 となる行列である。補間では、 $\mathbf{H}$ は既知であるとする。劣化画像および補間画像の例を図 3 に示す。この図では、構造要素学習に基づく補間画像と、事前情報を用いない補間法としてスイッチングメジアンフィルタ、TV 補間法 (Haar 変換による近似 [14]) の例を比較対象として示している。提案法は、画像を生成モデル (14) によって近似する過程で、繰り返し現れる画像の構造を学習し、補間を実現する。そのために、他の方法では再現が不可能な細かいテクスチャを良好に復元している。

5 スパース周期信号分解

前章では、信号を基底ベクトルへ分解する観点から、スパース性に関するコストとして l_1 ノルムのペナルティを係数ベクトルに課した。さて、音響信号等では、信号成分の多くが周期性を持ち、短時間の観測区間では、周期信号とみなすことができる。この観点から、著者らは、信号をできるだけ少ない数の周期信号群へ分解することを目的として、周期信号の発生にスパース性に関するペナルティを課したスパース周期信号分解を提案している [15][16]。このスパース周期信号分解では、入力信号 $\mathbf{f}$ を周期信号群 $\{\mathbf{f}_p\}_{p \in \mathbb{P}}$ の和

$$\mathbf{f} = \sum_{p \in \mathbb{P}} \mathbf{f}_p + \mathbf{e} \quad (17)$$

として表現する。ここで $\mathbf{e}$ は、モデル誤差を示す。また、 $\mathbb{P}$ は分解に用いる周期の集合であり、個々の p -周期信号は、

$$f_p(n) = \sum_{k=0}^K a_{p,k} t_p(n - kp). \quad (18)$$

と与えられる。ここで K は、信号の次元を N として $K = \lfloor (N-1)/p \rfloor$ であり、 K は $(N-1)/p$ を超えない最大の整数である。また、信号 t_n は、0 から $p-1$ の範囲で定義され、信号の 1 周期分の波形を示す。また、1 周期の間で t_n の平均は 0 とする。 $a_{p,k}$ は、波形のエンベロープを表す係数であり、非負の値とする。参考文献 [12] では、エンベロープを一定としたモデル、また、参考文献 [15] では、エンベロープ変化を加えた場合のモデルを用いて、信号分解を実現している。

上記の信号モデルを用いて、できるだけ少ない数の周期で信号を分解するために、個々の周期信号の l_2 ノルムの総和をペナルティとしたコスト関数

$$E(\{\mathbf{f}_p\}_{p \in P}) = \frac{1}{2} \|\mathbf{f} - \sum_{p \in P} \mathbf{f}_p\|_2^2 + \lambda \sum_{p \in P} \alpha_p \|\mathbf{f}_p\|_2. \quad (19)$$

の局所最小の一つを与える周期信号群を求める。スパース信号分解における議論に基づけば、上記のペナルティ関数の最小化は、モデル誤差をガウス分布とし、個々の周期信号の l_2 ノルムにラプラス分布を仮定した場合の事後確率最大化に対応する。

各周期信号に対する振幅係数列、および1周期の波形を導出するために、参考文献[15]ではスパースコーディングと類似したアルゴリズムを提案している。この方法では、周期を一つ選択し、その周期の周期信号の振幅係数を固定してコスト関数を最小化する波形を求め、次に波形を固定してコスト関数を最小化する振幅係数を求める。これを順に周期を選択しながら、コスト関数の収束まで繰り返す[15]。

参考文献[15][17]では、音声の混合信号に対してスパース周期信号分解を適用し、音声分離実験を行っている。この実験では、混合信号を観測フレームに分割し、それぞれのフレームに信号分解を適用している。実験結果より、スパース周期信号分解は、2つの音声の混合信号を、1フレームあたり平均20個程度の周期信号に分解し、それを理想的に振り分けることで音声分離が実現できることがわかった。スパース周期信号分解による音声分離の精度は、離散フーリエスペクトルを理想的に振り分けた場合に比べて、3dBほど低い。しかしながら、分解で生じる信号が少なく、分離後の信号を話者へ、容易に振り分けることができる。分解後に得られる周期信号のクラスタリングと、話者に関する先見情報に基づき、参考文献[17]では、単一マイクロホンからのセミブラインド音声分離を実現している。

6 まとめ

本稿では、スパース性のペナルティの下で信号分解を行う方法、および、スパースに信号を表現するための基底ベクトルの学習法としてスパースコーディングを解説した。事前に対象となる信号をスパースに表現できる基底行列を知ることができれば、スパース性のペナルティを用いて、混合信号からの信号分離、雑音除去、補間・復元等を実現できる。さらに信号のスパース性に注目することで、少ない数の観測値を用いて信号を復元する Compressive Sampling[18]も現在、注目を集めている。これは、乱数で生成された基底ベクトルと信号の内積を観測値として、少ない数の観測値から、信号の持つスパース性により信号を復元する考え方である。

また、スパースコーディングは、信号をスパースに表現できる基底ベクトルを学習することから、信号の特徴を捉えた基底ベクトルを導出することができる。スパースコーディングにより得られた基底ベクトルは、画像の特徴を記述しているため、これら为先見情報として画像処理へ応用することが期待できる。

今後の発展として、応用に応じた信号モデルの拡張、スパース性のペナルティ関数の拡張等が挙げられる。特にスパース性のペナルティ関数については、係数に対して仮定する

事前確率に応じて、いろいろな関数を考えることができるが、それらに応じた最適化法を検討する必要がある。また、学習サンプルからの基底ベクトルの識別可能性、分解の一意性についても、今後、議論が続けられていくことが予想される。

参考文献

- [1] S. Mallat, A WAVELET TOUR OF SIGNAL PROCESSING, Academic Press, London, 1998.
- [2] S. Mallat and Z. Zhang, "Matching pursuit in a time-frequency dictionary," IEEE Trsns. on Signal Processing, vol. 41, no. 12, pp. 3397-3415, Dec. 1993.
- [3] R. Neff, A. Zakhor, "Very low bit-rate video coding based on matching pursuits," IEEE Trans. on Circuits and Syst. Video Tech., vo. 7, no. 1, pp. 158-171, July 1997.
- [4] S. S. Chen, D. L. Donoho and M. A. Saunders, "Atomic decomposition by basis pursuit," in SIAM Journal on Scientific Computing, vol. 20, no. 1, pp. 33-61, 1998.
- [5] S. Sardy, A. G. Bruce and P. Tseng, "Block coordinate relaxation methods for nonparametric wavelet denoising," Journal of Computational and Graphical Statistics, vol. 9, no. 2, pp. 361-379, 2000.
- [6] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive-field properties by learning a sparse code for natural images," Nature, vol. 381, no. 13, pp. 607-609, July 1996.
- [7] M. S. Lewicki and B. A. Olshausen, "A probabilistic framework for the adaptation and comparison of image codes," J. Opt. Soc. Amer. A, Opt. Image Sci., vol. 16, no. 7, pp. 1587-1601, 1999.
- [8] M. Aharon, M. Elad and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," IEEE Trans. on Signal Processing, vol. 54, no. 11, pp. 4311-4322, Nov. 2006.
- [9] M. D. Plumbley, S. A. Abdallah, T. Blumensath and M. E. Davies, "Sparse representation of polyphonic music," Signal Processing, vol. 86, pp. 417-431, 2006.
- [10] T. Blumensath and M. Davies, "Sparse and shift-invariant representations of music," IEEE Trans, on Audio, Speech and Language Processing, pp. 50-57, vol. 14, no.1 Jan. 2006.
- [11] M. Elad, J. -L. Stark, P. Querre and D. L. Donoho, "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)," Appl. Comput. Harmon. Anal., vol. 19, pp. 340-358, 2005.

- [12] M. Nakashizuka, H. Nishiura and Y. Iiguni, "Sparse image representations with shift-invariant tree-structured dictionary, " in Proc. IEEE Int'l Conf. on Image Processing, Cairo, Nov, 2009.
- [13] M. Nakashizuka, H. Nishiura, Y. Iiguni, "Shift-invariant sparse image representations with tree-structured dictionary, " IEICE Trans. on Fundamentals, vol. E92-A, no. 11, pp. 2809-2818, Nov. 2009.
- [14] J. -L. Stark, M. Elad and D. L. Donoho, "Image decomposition via the combination of sparse representation and a variational approach, " IEEE Trans. on Image processing, vol. 14, no. 10, pp. 1570-1582, Oct. 2005.
- [15] M. Nakashizuka, "A sparse decomposition method for periodic signal mixtures, " IEICE Trans. on Fundamentals, vol. E91-A, no. 3, pp.791-800, March 2008.
- [16] M. Nakashizuka, H. Okumura and Y. Iiguni, "A sparse periodic decomposition and its application to speech representation, " in Proc. on EUSIPCO 2008, Lausanne, Aug. 2008.
- [17] M. Nakashizuka, H. Okumura and Y. Iiguni, "Single-channel speech separation by using a sparse periodic decomposition, " in Proc. EUSIPCO 2009, Glasgow, Aug. 2009.
- [18] J. Romberg, "Imaging via compressive sampling, " IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 12-20, March 2008.
- [19] M. Zibulevsky and M. Elad, "L1-L2 Optimization in signal and image processing, " IEEE Signal Processing Magazine, pp. 76-88, May 2010.