Vol.2012-CVIM-183 No.20
2012/9/3

スパース表現の数理とその応用

日野 英逸[1,a)]　村田 昇[1,b)]

概要：スパースコーディングは生物の一次視覚野の情報処理を数学的にモデル化したものであり，与えられた画像を少数の基底の線型結合で表現する手法である．観測信号のスパース表現は，工学的にも効率的な情報の保持・伝達，あるいはノイズに対して頑健な情報表現を実現する手法として注目を集めている．本稿では，スパースコーディングを始めとする種々の行列分解手法の数理的側面を，その確率モデルを介して統一的に論じる．また，スパースコーディングの代表的なアルゴリズムと幾つかの応用を紹介する．

1. はじめに

生物の一次視覚野の細胞は，方向と空間スケールに関して選択的な空間フィルタを有していることが古くから知られていた．一次視覚野における受容細胞は，ある方向を向いたある空間周波数成分が網膜上の特定の領域に出現すると，選択的に反応する性質を持っている．細胞が反応する方向と周波数は受容細胞毎に異なるため，視覚野全体として空間フィルタバンクを形成していると考えることが出来る．一次視覚野の細胞がこうした特徴を持つに至った理由を説明するために様々な試みがなされてきたが，ここ十数年の間に，自然画像の統計的構造を積極的に利用することによって自然画像を効率的に符号化 (コーディング) するための仕組みとして獲得されたとする考え方が提案され脚光を浴びている [1, 2]．Olshausen らによる一連の研究では，「自然画像を基底画像の線型結合で表した時，その結合係数が疎である」という制約の下で基底を学習することで，生物の一次視覚野における受容細胞と同様の特徴を有する基底が得られることが実験的に示されている．本稿では画像の小領域の画素値を観測信号として扱い，信号のスパース表現の問題を解説する．例えば一辺が $\sqrt{d}$ ピクセルの画像の小領域を扱う場合には，$\sqrt{d} \times \sqrt{d}$ の画素値をベクトル表現した $\boldsymbol{x} \in \mathbb{R}^d$ が一つの観測信号ということになる．Olshausen & Field [1] は，画像 $\boldsymbol{x}$ が基底ベクトル[*1] $\boldsymbol{d}_k,\ k = 1, \ldots, m$ の線型結合

$$\boldsymbol{x} = \sum_{k=1}^{m} c_k \boldsymbol{d}_k,\ c_k \in \mathbb{R} \qquad (1)$$

で表されるという仮定を置いた．この基底系 $\{\boldsymbol{d}_k\}$ は，画像を「効率的に」表現することが出来るように構成する．ここでいう効率的とは，ある画像を表現する際に係数ベクトル $\boldsymbol{c} = (c_1, \ldots, c_m) \in \mathbb{R}^m$ のうちの少数の c_k のみが非ゼロの値を取り，残りの大部分はゼロとなることを意味する．もし非ゼロの c_k の数が d より少なく，かつ基底系 $\{\boldsymbol{d}_k\}$ を予めうまく選ぶことができているとすれば，d 個の値を持つ画像 $\boldsymbol{x}$ をより少ない個数の非ゼロの c_k の集合で効率良く表現することができるからである．このように，意味のある非ゼロ要素が全体に対して少数である状態を，スパース (sparse; 疎) であると呼ぶ．適切な基底系 $\{\boldsymbol{d}_k\}$ とその結合係数 $\{c_k\}$ を得るには，観測信号の情報を保存する項 (文献 [1] では l_2 ノルムを考えている) と，結合係数がスパースになることを促進する正則化項 Ψ の和

$$\min_{\{\boldsymbol{d}_k\},\{c_k\}} \quad \|\boldsymbol{x} - \sum_{k=1}^{m} c_k \boldsymbol{d}_k\|_2^2 + \Psi(\boldsymbol{c}) \qquad (2)$$

の最小化を行えば良い．例えば [1] では Ψ として $\Psi(\boldsymbol{c}) = \lambda \sum_{k=1}^{d} e^{-c_k^2},\ \lambda > 0$ を用いることを提案している．こうして得られた基底は，生物の一次視覚野細胞とよく似た局所的な空間周波数特性を持つ Gabor wavelet 基底と類似したものになる [3]．

なお，Olshausen & Field [1] 以前から，ニューラルネットワークや連想記憶の研究において，スパースな符号化によって情報の表現及び保持を効率的に行うことが出来ることが知られていた．例えば [4–6] では，工学的観点からスパース表現の利点の定量的議論が行われている．

以下，本稿の構成を記す．第 2 節では画像の生成モデルを簡単に説明する．第 3 節ではスパースコーディングの定式化を行ない，スパース表現において中心的役割を果たすノルムを導入する．第 4 節ではスパース表現を観測信号行列の分解手法として捉え，関連する種々の手法を概説し，

[1] 早稲田大学理工学術院　〒 169–8555 東京都新宿区大久保 3–4–1
[a)] hideitsu.hino@toki.waseda.jp
[b)] noboru.murata@eb.waseda.ac.jp

[*1] 必ずしも一次独立なベクトルとは限らず，基底という表現は正確ではないが，慣習上「基底」と呼ぶことにする．

第 5 節でこれらの行列分解手法の背後にある確率モデルを説明する．第 6 節及び第 7 節ではそれぞれスパース表現のための係数，基底の最適化手法を紹介し，第 8 節ではスパース表現の画像処理における応用例を挙げる．

2. 画像の生成モデル

本稿ではスパース表現の具体的な例として画像を取り上げる．その準備のために，画像の生成モデルとして重要な２つの考え方を簡単に纏める．

2.1 ピクセルベースモデル

多数の変数とその変数間の相互作用からなる系のうち，近傍のみで条件付けられる単純なモデルの一つとしてはマルコフ確率場 (Markov Random Field; MRF) があり，画像の確率モデルとしても広く利用されている [7]．このモデルは，あるピクセルのとる値 v の確率がその周辺のピクセル値で決まる条件付き確率

$$p(v|\text{ 周辺のピクセル値})$$

によって表されると考える，ピクセルベースのモデルである．マルコフ確率場は事前分布として画像の滑らかさなどに対応する制約を容易に取り込むことができる．例えば劣化画像を観測した場合に，観測した画像で条件付けた原画像の生成確率 (事後確率) を利用して画像復元を行う方法や，超解像への応用が知られている [8,9]．また，局所的な相互作用を記述した確率モデルを導入することで，統計力学の分野で発展した平均場近似を始めとする各種の近似計算手法が利用できる [8]．しかし，実用に供するレベルでの画像の生成機構をモデル化するためには，事前分布の適切な設定と，本質的に非凸な問題の最適化技術が必要となり，綿密な設計と高度なノウハウが要求される．

2.2 基底ベースモデル

画像を基底画像の線型結合により表現するアプローチを，以下では基底ベースモデルと呼ぶ．画像全域を表現する基底を利用することもあるが，一般には画像の小領域を対象としてモデル化を行い，小領域をずらしつつ画像全域を表現することが多い．特に自然画像はある程度小さな領域に限ると類似したパターンを示すことが多く，従って

$$\boldsymbol{x} = D\boldsymbol{c}, \quad \#\{i \big| c_i \neq 0\} \ll d(= \text{領域の画素数}),$$

のように少数の基本的な基底の組合せで表現できる．ここで $\#S$ は集合 S の要素数を表す．また，多くの画像処理手法の計算量は観測信号の次元に対して指数関数的に増加するため，小領域毎に処理を行う基底ベースモデルには計算量の低減という利点もある．本稿ではスパース表現の数理的側面の説明を，基底ベースモデルによる画像の表現を前提として行う．

3. スパースコーディングの定式化

本節ではスパース表現の代表例としてスパースコーディング問題を定式化する．また，スパース性と関連の深いベクトル及び行列のノルムについて説明する．

3.1 定式化

d 次元観測信号ベクトルを $\boldsymbol{x} \in \mathbb{R}^d$ とし，m 個の基底を並べた辞書 $D = (\boldsymbol{d}_1, \cdots, \boldsymbol{d}_m)$ の線型結合により $\boldsymbol{x} = D\boldsymbol{c}$ のように表現することを考える．この時，スパースコーディングの枠組みでは $m \gg d$ という過剰決定系[*2]による近似を考える．信号の次元より多い基底による表現 $\boldsymbol{x} = D\boldsymbol{c}$ では $\boldsymbol{c}$ の一意性を保証することが出来ないので，通常は観測信号 $\boldsymbol{x}$ の表現に利用される基底を D のうちの一部に制限する．つまり，$\|\boldsymbol{c}\|_0$ で $\boldsymbol{c}$ の l_0 ノルム，すなわちベクトル $\boldsymbol{c}$ の非ゼロ成分の数を表すとして，スパースコーディングは典型的には最適化問題

$$\min_{\boldsymbol{c}} \quad \|\boldsymbol{x} - D\boldsymbol{c}\|_2^2 + \lambda\|\boldsymbol{c}\|_0, \quad \lambda > 0 \qquad (3)$$

として定式化される．しかしながら，この問題は全ての基底の組み合わせを試さないと最適解が得られない組合せ最適化問題であり，NP 困難であることが知られている [10]．そこで，l_1 ノルムへの緩和問題

$$\min_{\boldsymbol{c}} \quad \|\boldsymbol{x} - D\boldsymbol{c}\|_2^2 + \lambda\|\boldsymbol{c}\|_1, \quad \lambda > 0 \qquad (4)$$

を考えることが多い．この l_1 ノルム正則化問題は線型計画問題として表現することが可能である．

より一般に，辞書と係数を求める問題は統計的推測の枠組みで次のように表される．n 個の観測信号を並べた行列を $X = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$，辞書行列を $D = (\boldsymbol{d}_1, \ldots, \boldsymbol{d}_m) = (\boldsymbol{d}^1, \ldots, \boldsymbol{d}^d)^\top$，各信号を表現するための辞書の係数をならべた係数行列を $C = (\boldsymbol{c}_1, \cdots, \boldsymbol{c}_n) = (\boldsymbol{c}^1, \cdots, \boldsymbol{c}^m)^\top$ とする．ここで，$\boldsymbol{d}_i \in \mathbb{R}^d$ は行列 D の第 i 列，$\boldsymbol{d}^j \in \mathbb{R}^m$ は行列 D の第 j 行を表し，同様に $\boldsymbol{c}_i \in \mathbb{R}^m$ は行列 C の第 i 列，$\boldsymbol{c}^j \in \mathbb{R}^n$ は第 j 行を表すものとする．$L(D, C; X)$ で近似の損失関数を表し，信号の表現がスパースならば小さい値を取る正則化項 $\Psi(D, C)$ を用いて，n 個の観測信号に対するスパースコーディングは最適化問題

$$\min_{D,C} \quad L(D, C; X) + \Psi(D, C) \qquad (5)$$

として定式化される．後述するように，損失関数は負の対数尤度，Ψ は基底あるいは係数に対する事前分布の負の対数尤度として確率モデルでの解釈が可能であることが多い．本稿で主に扱う式 (5) の定式化は，損失関数 $L(D, C; X)$ を可能な限り小さくすることで信号の再構成を精度よく行いつつ，正則化項 $\Psi(D, C)$ の最小化により表現のスパース性

[*2] 機械学習の文脈では過完備基底 (overcomplete basis) とも呼ぶ．

を確保しようというものである．問題を実際に解く際や特定の理論解析においては，例えば

$$\min_{D,C} \quad \Psi(D,C) \quad \text{subject to} \quad L(D,C;X) < \epsilon,$$
$$\min_{D,C} \quad L(D,C;X) \quad \text{subject to} \quad \Psi(D,C) < \epsilon$$

のような定式化が便利なこともある．種々の表現の同等性，及び用途に応じた適切な表現に関しては [11] に詳しくまとめられている．

3.2 ベクトルのノルム

基底による信号の表現でのスパース性は，基本的には結合係数に対するノルム制約から導かれる．本稿では一般にベクトル $\boldsymbol{x} \in \mathbb{R}^d$ に対して，

$$\|\boldsymbol{x}\|_p = \left(\sum_{i=1}^{d} |x_i|^p \right)^{1/p} = \sqrt[p]{|x_1|^p + \cdots + |x_d|^p} \quad (6)$$

で l_p ノルム (l_p norm) を表現する．特に $p = 0$ の時は

$$\|\boldsymbol{x}\|_0 = \lim_{p \to 0} \|x\|_p^p = \lim_{p \to 0} \sum_{i=1}^{d} |x_i|^p = \#\{i \big| x_i \neq 0\} \quad (7)$$

とする．定義 (6) において $0 \leq p < 1$ の範囲では厳密には $\|\boldsymbol{x}\|_p$ はノルムの公理を満たさない．すなわち，$p = 0$ の時は斉次性

$$\|a\boldsymbol{x}\|_p = |a| \|\boldsymbol{x}\|_p, \quad a \in \mathbb{R}$$

が満たされず，$0 < p < 1$ の時は三角不等式 (劣加法性)

$$\|\boldsymbol{x}_1 + \boldsymbol{x}_2\|_p \leq \|\boldsymbol{x}_1\|_p + \|\boldsymbol{x}_2\|_p, \quad \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^m$$

が満たされないので，$\|\boldsymbol{x}\|_p$, $0 \leq p < 1$ は正確にはノルムとはならないが，慣習上 $0 \leq p < \infty$ に対して $\|\boldsymbol{x}\|_p$ を l_p ノルムと呼ぶ．

3.3 行列のノルム

行列 $M \in \mathbb{R}^{n \times m}$ に対するノルムにも種々の定義がある．行列特有のノルムとして特に重要なものは核ノルム (nuclear norm) あるいはトレースノルム (trace norm) と呼ばれるもので，

$$\|M\|_* = \mathrm{Tr}(\sqrt{M^\top M}) = \sum_{i=1}^{m} \sigma_i \quad (8)$$

で定義される．ここで，σ_i は行列 M の特異値である．ベクトルのノルムにおいて $p = 0$ の場合に対応するのが，行列のランク (rank)

$$r = \#\{i | \sigma_i \neq 0\} \quad (9)$$

である．すなわち，行列 M の非ゼロ特異値の数が小さいという制約は，行列にスパースな構造を与える．しかし，ベクトルに対する l_0 ノルム制約が組み合せ最適化問題を伴い計算が困難であるように，行列に対するランク制約は非凸な制約となり取り扱いが難しい．そこで，ベクトルの l_0 ノルムを l_1 ノルムで緩和するように，ランク制約を核ノルム制約で緩和することが多い．

また，行列の成分毎に定義した l_p ノルム

$$\|M\|_p = \left(\sum_{i=1}^{n} \sum_{j=1}^{m} |M_{ij}|^p \right)^{1/p} \quad (10)$$

もよく利用される．特に $p = 2$ の場合をフロベニウスノルム (Frobenius norm) と呼ぶ．ベクトルあるいは行列のノルムに関する議論は，例えば [12] に詳しい．本稿では行列のノルムとしては核ノルムあるいは成分毎の l_p ノルムを主に利用する．

4. 行列分解としての理解

本節では，代表的なスパース表現を観測信号行列の分解手法という観点から整理する．一般に，スパース表現は観測信号行列 X を

$$X = DC + E, \quad (11)$$

のように辞書行列 D とその係数行列 C で表現するものである．ここで，$E = (\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n), \boldsymbol{\epsilon}_i \in \mathbb{R}^d$ であり誤差を表す行列であるが，一般に各 $\boldsymbol{\epsilon}_i$ は適当な平均ゼロの多変量正規分布 $\mathcal{N}(\boldsymbol{0}, \sigma^2 I_d)$ に従うとすることが多い．ここで I_d は d 次元単位行列を表す．

通常は係数行列 C の各列がスパース，すなわち利用される基底が少数であるという状況を考えるが，辞書行列 D の各列，すなわち各基底がスパースである，あるいは D, C の両方がスパースであるという問題設定もある．

4.1 スパースコーディング

スパースコーディング (Sparse Coding; SC) では多くの場合損失関数としてフロベニウスノルム $L(D,C;X) = \|X - DC\|_2^2$ を考え，正則化項 $\Psi(D,C)$ としては係数ベクトル $\boldsymbol{c}$ の l_p ノルムで $0 \leq p \leq 1$ の場合を考える:

$$\min_{D,C} \quad \sum_{i=1}^{n} \|X - DC\|_2^2 + \lambda \sum_{i=1}^{m} \|\boldsymbol{c}_i\|_p^p, \quad \lambda \geq 0. \quad (12)$$

狭義のスパースコーディングは，与えられた辞書 D を用いて信号 $\boldsymbol{x}$ を近似するためのスパースな係数 $\boldsymbol{c}$ を求める問題であるが，辞書の最適化も含めて考えることもある．係数 C，辞書 D の具体的な最適化手法についてはそれぞれ第 5 節，第 6 節で簡単に説明する．

4.2 主成分分析

ここでは，観測信号の各次元は平均ゼロに正規化されているとする．すなわち，観測行列 $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ の各行の和は 0 であると仮定する．X のランクを r として，X

の特異値分解を

$$X = U\Sigma V^\top$$

とする．ここで，

$$\begin{aligned} U &= (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_r),\ \boldsymbol{u}_i \in \mathbb{R}^d, U^\top U = I_r \\ V^\top &= (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n),\ \boldsymbol{v}_i \in \mathbb{R}^r, V^\top V = I_r \\ \Sigma &= \mathrm{diag}(\sigma_1, \cdots, \sigma_r), \sigma_i \in \mathbb{R} \end{aligned}$$

として，特異値は降順 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$ に並んでいるものとする．正規直交基底系をなす辞書行列 $D = U$ を固定した時，係数行列は $C = \Sigma V^\top$ であり，信号 $\boldsymbol{x}_i$ を $\boldsymbol{x}_i = D\boldsymbol{c}_i$ で表すとき，その係数ベクトルは

$$\boldsymbol{c}_i = \Sigma \boldsymbol{v}_i = (\sigma_1 v_{1i}, \ldots, \sigma_r v_{ri})^\top$$

で与えられる．主成分分析 (Principal Component Analysis; PCA [13]) では $L(D, C; X) = \|X - DC\|_2^2$ を想定し，辞書 D は正規直交基底系をなすという制約がある．この制約を $\Psi(D, C) = \|D^\top D - I_r\|_2^2$ で表すと，主成分分析は

$$\min_{D,C} \quad \|X - DC\|_2^2 + \lambda \|D^\top D - I_r\|_2^2 \qquad (13)$$

という行列分解問題として表現できる．辞書を $D = U$ とした時にこの目的関数を最小にするような係数 C は $C = \Sigma V^\top$ で与えられる．また，X をランク $k < r$ の行列の積で近似するような場合には，小さな特異値に対応する σ_i の $k+1$ から r までを 0 にした行列 $\tilde{\Sigma} = \mathrm{diag}(\sigma_1, \cdots, \sigma_k, 0, \cdots, 0)$ を用いてスパースな係数が

$$\boldsymbol{c}_i = \tilde{\Sigma} \boldsymbol{v}_i = (\sigma_1 v_{1i}, \sigma_2 v_{2i}, \ldots, \sigma_k v_{ki}, 0, \ldots, 0)^\top$$

で与えられる．主成分分析そのものは，観測信号行列の低ランク行列による近似という意味でのスパース表現であるが，C の各列に対する l_p ノルム制約を考えることで積極的にスパースな係数を得る主成分分析の拡張法も提案されている [14].

4.3 独立成分分析

独立成分分析 (Independent Component Analysis; ICA [15]) においては，D は辞書というよりは混合行列であり，n 個の m 次元原信号 $C = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n) = (\boldsymbol{c}^1, \ldots, \boldsymbol{c}^m)^\top$，$\boldsymbol{c}_i \in \mathbb{R}^m, \boldsymbol{c}^k \in \mathbb{R}^n$ が，混合行列 $D = (\boldsymbol{d}^1, \ldots, \boldsymbol{d}^d)^\top$ によって混合した結果 X が観測されると考える．すなわち，$\boldsymbol{x}_k$ の第 i 成分は，m 次元の原信号 $\boldsymbol{c}_k$ が混合比 $\boldsymbol{d}^i = (d_1^i, \ldots, d_m^i)^\top$ で混合された信号 $(\boldsymbol{d}^i)^\top \boldsymbol{c}_k$ となる．このとき，$d \geq m$，すなわち観測信号の次元が原信号の次元以上で，原信号 $\boldsymbol{c}^k, k = 1, \ldots, m$ が非正規で互いに独立であるならば，観測信号 X から原信号 C を復元することが出来る．簡単のために，$D \in \mathbb{R}^{m \times m}$ かつ正則であるとする，すなわち観測信号の次元と原信号の次元は一致しており，混合行列 D は逆行列を有すると仮定する．このとき，復元行列を $B = D^{-1}$ と置くと，$C = BX = (\boldsymbol{c}^1, \ldots, \boldsymbol{c}^m)^\top$ であり，この C の各行 $\boldsymbol{c}^k$ が独立になるように B を推定する．スパースな行列分解とは異なるように見えるが，次節で述べるように係数 C のエントロピー最小化の観点からスパース表現としての理解が可能である．例えば $L(D, C; X) = \|X - DC\|_2^2$ とおき，確率密度関数 p を持つ確率変数のエントロピーを

$$H(p) = -\int p(\boldsymbol{x}) \log p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$$

と記すことにする．混合行列 D の関数としての原信号 $C = D^{-1}X$ の第 j 行が従う分布を p_{c^j} とし，$\Psi(D, C) = \sum_{j=1}^m H(p_{c^j})$ を用いると，独立成分分析の一つの定式化として次の最適化問題

$$\min_D \quad \|X - DC\|_2^2 + \sum_{j=1}^m H(p_{c^j}) \qquad (14)$$

が得られる．

4.4 非負行列因子分解

データ解析において，データ，基底，結合係数ともに非負に制限することが望ましいことがある．例えば，画像のピクセル値，エネルギー，イベントの生起回数などは非負値のみを取るため，これらのデータの基本要素となる非負の基底の加法のみで観測信号が表現される分解が自然である．非負行列因子分解 (Non-negative Matrix Factorization; NMF [16, 17]) は，非負データを，非負成分のみからなる辞書と係数に分解する手法である．通常，非負行列因子分解は正則化項 $\Psi(D, C)$ に直接非負性は含めず，最適化問題 (5) の定義域を制限して

$$\min_{[D]_{ij} \geq 0, [C]_{ij} \geq 0} \quad \|X - DC\|_2^2 + \Psi(D, C) \qquad (15)$$

という問題を考える．ただし $[M]_{ij}$ は行列 M の (i, j) 成分である．損失関数 $L(D, C; X)$ として例えば [17] では，

$$L(D, C; X) = \|X - DC\|_2^2 \qquad (16)$$

あるいは

$$L(D, C; X) = \sum_{i=1}^d \sum_{j=1}^n \left([X]_{ij} \log \frac{[X]_{ij}}{[DC]_{ij}} - [X]_{ij} + [DC]_{ij} \right) \qquad (17)$$

を用いて，定義域に関する制約のみで正則化項を含まない単純な最適化アルゴリズムを導出している．これらの損失関数は，それぞれ，近似誤差の分布に正規分布を仮定した場合の負の対数尤度と，平均 $[DC]_{ij}$ のポアソン分布を仮定した場合の一般化 Kullback-Leibler (KL) divergence に対応する．

非負行列因子分解では非負成分のみを有する基底の加

法のみで信号を表現するため，局所性を有する辞書と，スパースな係数が得られやすい．ただし，問題によっては解釈が容易なスパース表現が得られにくいため，D と C に同時に l_1 ノルム正則化を施し，例えば

$$\Psi(D,C) = \lambda_1 \sum_{j=1}^{m} \|\boldsymbol{d}_j\|_1 + \lambda_2 \sum_{i=1}^{n} \|\boldsymbol{c}_i\|_1,\ \lambda_1, \lambda_2 > 0 \quad (18)$$

とすることで，積極的にスパースな表現を得る手法も提案されている [18]．また，$L(D,C;X)$ として通常の KL divergence に替えて α-divergence や Bregman divergence を用いた一般化も行われている [19, 20]．

5. 確率モデルとしての理解

本節では，前節で説明した各種のスパース行列分解の確率モデルとしての解釈を与える．なお，各手法は異なる複数の確率モデルとしての解釈が可能であり，本節で示す解釈はその一つであることに注意されたい．

5.1 スパースコーディング

簡単のため，まず辞書 D は与えられている状況を考える．損失関数を $L(D,C;X) = \|X - DC\|_2^2$ としたとき，これは各信号 $\boldsymbol{x}_i, i = 1, \ldots, n$ の近似誤差 $\boldsymbol{\epsilon}_i$ に正規分布を仮定していることになる．すなわち $p(\boldsymbol{x}_i - D\boldsymbol{c}_i) = p(\boldsymbol{x}_i|\boldsymbol{c}_i) \propto \exp(-\frac{1}{\sigma^2}\|\boldsymbol{x}_i - D\boldsymbol{c}_i\|_2^2)$ である．係数の分布としては，$0 < p \leq 2$ の時は一般化ガウス分布

$$p(\boldsymbol{c}_i) \propto \exp\left(-\frac{\lambda}{\sigma^2}\|\boldsymbol{c}_i\|_p^p\right) \quad (19)$$

を仮定していると考えられる．$0 < p < 2$ のときに，この分布は正規分布と比較して裾の重く尖度が高い優ガウス的分布となり，スパースな解を与える[*3]．これにより，観測信号と基底の結合係数の負の対数尤度は，定数項を除いて

$$-\sum_{i=1}^{n} \log p(\boldsymbol{x}_i|\boldsymbol{c}_i)p(\boldsymbol{c}_i) = \|X - DC\|_2^2 + \lambda \sum_{i=1}^{n} \|\boldsymbol{c}_i\|_p^p \quad (20)$$

となり，これを C に関して最小化することは結合係数の MAP 推定と等価である．$p = 0$ の時に確率的解釈を行うためには非ゼロの値を取る係数の数，係数の添字，係数の値に対する分布を定める必要があり，特に係数の添字に関する分布の評価が困難となる．Bayes 統計の文脈で変数選択に利用されてきた Spike & Slab モデル [21] が，係数に l_0 正則化を課したスパースコーディングの確率モデルとして注目を集めている (図 1)．これは，係数ベクトル $\boldsymbol{c}$ の分布として適当な連続分布 (正規分布を用いることが多い) と，中心をゼロとするディラックのデルタ関数 $\delta_0(\cdot)$ の混合分布

[*3] 典型的には，$p = 1$ の時にラプラス分布になる．

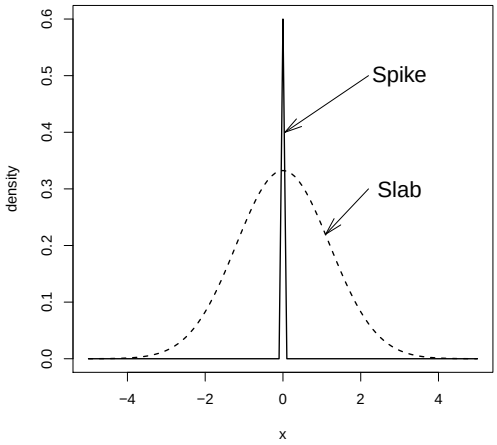

図 1 l_0 正則化に対応する Spike & Slab 事前分布．
Fig. 1 Spike & Slab prior distribution of coefficients for l_0 regularization.

$$p(c_i|\sigma^2, q) = q\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}c_i^2\right) + (1-q)\delta_0(c_i) \quad (21)$$

を用いるものである．このモデルでは非ゼロのパラメタの数を混合比 q でコントロールすることが可能であり，l_0 ノルムを模していると解釈できる．基底 D の事前分布にこのモデルを仮定するもの [22] も，係数 C の事前分布に仮定するもの [23] もある．例えば係数 C の事前分布に Spike & Slab モデルを仮定する場合は，

$$p([C]_{ij}|\sigma_c^2, q) = q\frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left(-\frac{1}{2\sigma_c^2}[C]_{ij}^2\right) + (1-q)\delta_0([C]_{ij}) \quad (22)$$

であり，式 (5) における正則化項 $\Psi(D,C)$ としては

$$\Psi(D,C) = -\log\left\{q\frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left(-\frac{[C]_{ij}^2}{2\sigma_c^2}\right) + (1-q)\delta([C]_{ij})\right\} \quad (23)$$

が考えられる．

5.2 主成分分析

主成分分析は近似誤差として正規分布を仮定しているが，辞書 D 及び係数行列 C としては直交性の制約のみを考えており，陽に D, C の分布を考えたモデルではない．そのため式 (5) に則って主成分分析の確率モデルを表現することは困難であるが，以下のようにして確率モデルとして解釈することが可能である [24][*4]．観測信号は平均がゼロに

[*4] この表現は因子分析 (Factor Analysis; FA [25]) と関連が深い．

正規化されていると仮定して，信号の表現

$$\boldsymbol{x} = D\boldsymbol{c} + \boldsymbol{\epsilon} \tag{24}$$

における誤差 $\boldsymbol{\epsilon}$ は正規分布 $\mathcal{N}(\mathbf{0}, \sigma^2 I_d)$ に従い，さらに $\boldsymbol{c}$ が正規分布 $\mathcal{N}(\mathbf{0}, I_m)$ に従うと仮定する．このとき，$p(\boldsymbol{x}|\boldsymbol{c})$ は $\mathcal{N}(D\boldsymbol{c}, \sigma^2 I_d)$ であり，$p(\boldsymbol{x}) = \int p(\boldsymbol{x}|\boldsymbol{c})p(\boldsymbol{c})\mathrm{d}\boldsymbol{c}$ は共分散行列 $\Sigma = DD^\top + \sigma^2 I_d$ の正規分布 $\mathcal{N}(\mathbf{0}, \Sigma)$ に従うことが分かる．このモデルを，確率的主成分分析 (Probabilistic PCA; PPCA) と呼ぶ．確率的主成分分析では，観測信号 $\boldsymbol{x}_i, i = 1, \ldots, n$ の尤度を最大化するように D 及び σ^2 を定めることが出来る．すなわち，辞書 D 及び残差の分散 σ^2 をパラメタとするモデル $p(\boldsymbol{x})$ に観測信号行列 X を代入して得られる負の対数尤度

$$-\log p(X|D, \sigma^2) = 2n \log |\Sigma| + \mathrm{Tr}(\Sigma^{-1} X X^\top), \tag{25}$$

ただし

$$\Sigma = DD^\top + \sigma^2 I_d$$

を D, σ^2 に関して最小化すればよい．ただし，対数尤度において最適化に寄与しない定数項は省略した．

確率的主成分分析には，欠損データの取り扱い，モデル選択，混合モデルへの拡張などが自然に行えるという利点がある．なお，[14] に対応して，特に係数にスパース性の制約を加えた確率的主成分分析も提案されている [26].

5.3 独立成分分析

中心極限定理によれば，有界な分散を持つ独立な確率変数の和の分布は，足し合わせる変数の数の増加とともに正規分布に近づいていく．一方，同じ分散を持つ連続確率変数の中で最大のエントロピーを持つ分布は正規分布である [27]．これらの事実から，原信号が重なりあうとその分布は正規分布に近づき，エントロピーは増大することが分かる．従って，直観的には観測信号 $\boldsymbol{x}$ から推定される復元信号 $B\boldsymbol{x}$ の各成分のエントロピーを減少させるように復元行列 $B = D^{-1}$ を選べば良いことが分かる．また，ラプラス分布を代表とする一般化ガウス分布 (19) など，平均ゼロの尖度の高い優ガウス的分布はゼロ周辺に高い確率密度を有することから，スパースな信号となる．

今，観測信号 $\boldsymbol{x} \in \mathbb{R}^d$ の従う確率分布の密度関数 p は推定出来ているとする．このとき，観測信号 $\boldsymbol{x}$ と原信号 $\boldsymbol{c} \in \mathbb{R}^m$ の関係から，$\boldsymbol{c}$ の確率分布は復元行列 B と確率密度関数 p で表される．これを $p_{\boldsymbol{c}}$ と書くことにする．また，復元信号 $\boldsymbol{c}$ の各成分 c^j の周辺分布を p_{c^j} と書くことにすると，原信号の各成分が独立であるとは，同時分布の密度関数と周辺分布の密度関数の積が等しいこと，すなわち

$$p_{\boldsymbol{c}} = p_{c^1} \cdots p_{c^m}$$

が成り立つことである．ここで，密度関数 p, q で表される２つの分布の「離れ具合」の尺度として KL divergence [27]

$$KL(p|q) = \int p(\boldsymbol{x}) \log \frac{p(\boldsymbol{x})}{q(\boldsymbol{x})} \mathrm{d}\boldsymbol{x}$$

を採用すると，独立性の必要十分条件は同時分布と周辺分布のエントロピーを用いて

$$KL(p_{\boldsymbol{c}} || \prod_{j=1}^m p_{c^j}) = \sum_{j=1}^m H(p_{c^j}) - H(p_{\boldsymbol{c}}) = 0$$

と表すことが出来る．同時分布とその周辺分布のエントロピーの間には不等式

$$H(p_{\boldsymbol{c}}) \leq \sum_{j=1}^m H(p_{c^j})$$

が成立するため，周辺分布のエントロピーの和が出来るだけ小さくなり，$H(p_{\boldsymbol{c}})$ と一致すれば $\boldsymbol{c}$ の各成分は独立であるといえる．以上より，独立成分分析の定式化の一つとして，

$$\min_D \quad \|\boldsymbol{x} - D\boldsymbol{c}\|_2^2 + \sum_{j=1}^m H(p_{c^j}), \tag{26}$$

が得られる．ここで，$H(p_{c^j})$ は $c^j, j = 1, \ldots, m$ の密度関数で値が定まる汎関数であり，$\boldsymbol{c} = D^{-1}\boldsymbol{x}$ なので結局は D^{-1} の関数であることに注意する．原信号が優ガウス的な分布に従うときに，エントロピーは低い値を取る．その意味で，エントロピー最小化による独立成分分析は原信号にスパース性の仮定をおいて信号分離を行なっていると解釈できる．例えば観測信号が n 個ある時，係数ベクトルの第 j 成分がラプラス分布 $p(c^j) \propto \exp(-\lambda|c^j|)$ に従うとして，各成分の観測値を $c_i^j, i = 1, \ldots, n$ とすると，$p(c^j)$ に関する期待値を経験平均で置き換えることでエントロピー $H(p_{c^j})$ は

$$H(p_{c^j}) = -\frac{1}{n} \sum_{i=1}^n \log p_j(c_i^j) \propto \lambda \sum_{i=1}^n |c_i^j| = \lambda \|\boldsymbol{c}^j\|_1 \tag{27}$$

のように推定できる．これは，$\Psi(D, C)$ として行列の成分毎の l_1 ノルムを用いることに相当する．以上より，問題 (26) は正則な行列 D に関する最適化問題

$$\min_D \quad \|X - DC\|_2^2 + \lambda \|D^{-1}X\|_1 \tag{28}$$

になる．ただし，原信号と観測信号の間には $C = D^{-1}X$ なる関係があることに注意する．

5.4 非負行列因子分解

非負行列因子分解の場合，ポアソン分布に基づく式 (17) のような損失関数も用いられるが，ここでは簡単のため近似誤差として正規分布を仮定した時の負の対数尤度で損失関数 $L(D, C; X) = \|X - DC\|_2^2$ を考える．D 及び C の各成分には非負という制約があるため，これらの事前分布と

しては非負の値を取る分布を設定する必要がある．ここでは一つの確率的解釈として，D, C の各成分がそれぞれ独立にパラメタ λ_1 と λ_2 の指数分布に従うとする．このとき，

$$p(D) = \prod_{i=1}^{d}\prod_{j=1}^{m} \lambda_1 \exp(-\lambda_1 [D]_{ij}), \quad [D]_{ij} \geq 0,$$
$$p(C) = \prod_{j=1}^{m}\prod_{k=1}^{n} \lambda_2 \exp(-\lambda_2 [C]_{jk}), \quad [C]_{jk} \geq 0$$

であり，D, C の MAP 推定は

$$\begin{aligned} &\min_{[D]_{ij}\geq 0, [C]_{jk}\geq 0} \|X - DC\|_2^2 \\ &\qquad + \lambda_1 \sum_{i=1}^{d}\sum_{j=1}^{m} [D]_{ij} + \lambda_2 \sum_{j=1}^{m}\sum_{k=1}^{n} [C]_{jk} \\ &= \min_{[D]_{ij}\geq 0, [C]_{jk}\geq 0} \|X - DC\|_2^2 + \lambda_1 \|D\|_1 + \lambda_2 \|C\|_1 \end{aligned} \tag{29}$$

と等価である．さらに残差の分布及び $[D]_{ij}, [C]_{jk}$ の従う指数分布のパラメタにも事前分布を設定し，Gibbs サンプリングにより Bayes 的に非負行列因子分解を行う手法が提案されている [28].

6. 係数選択の手法

観測信号 $\boldsymbol{x}$ と辞書 D が与えられた時，$\boldsymbol{x}$ を $D\boldsymbol{c}$ で近似するような係数 $\boldsymbol{c}$ を求める問題を，(狭義の) スパースコーディング問題と呼ぶ．ここでは最適化問題 (3) を，再構成誤差を一定の閾値以下に抑えた上で出来るだけ少ない数の基底の線型結合で信号を近似する問題

$$\min_{\boldsymbol{c}} \quad \|\boldsymbol{c}\|_0 \quad \text{subject to} \quad \|\boldsymbol{x} - D\boldsymbol{c}\|_2 < \epsilon \tag{30}$$

の形に書き直して考える．先に述べたように，スパース性の制約が l_0 ノルム制約の場合には (30) は組合せ最適化問題であり，NP 困難な問題である [10]．この問題に対する解法として，貪欲法に基づく方法や，l_0 制約を l_1 制約で緩和した上で解く方法など，数多くのアルゴリズムが提案されている．本節では，信号処理及び画像処理の分野で広く用いられているスパースコーディングアルゴリズムとして，
(1) Orthogonal Matching Pursuit [29]
(2) Iterative Reweighted Least Squares [30]
を紹介する．

Orthogonal Matching Pursuit (OMP)
OMP は，観測信号の近似に利用する係数の添字集合の中から「サポート」，すなわち非ゼロ係数の添字集合 $\mathcal{S}$ を見つけ出すアルゴリズムである．初めはサポートは空集合であるとして，観測信号 $\boldsymbol{x}$ を基底の線型結合で近似した時の残差を最小にするように新たな基底をサポート集合に一つ一つ追加していき，サポートに含まれる基底のみで信号を近似した時の残差が ϵ 以下になったら停止する．残差の低減に寄与する基底を順次選択していく貪欲法であり，解の最適性は保証されないが，多くの場合優れた近似を与えることが知られている．OMP の手続きを Algorithm1 にまとめる．

Algorithm 1 OMP: (Orthogonal Matching Pursuit)

input:
辞書行列 $D \in \mathbb{R}^{d\times m}$
観測信号 $\boldsymbol{x} \in \mathbb{R}^d$
近似閾値 $\epsilon > 0$
initialize:
繰り返しのカウンタ $t = 0$
初期係数ベクトル $\boldsymbol{c}^0 = \boldsymbol{0}$
初期残差 $\boldsymbol{r}^0 = \boldsymbol{x} - D\boldsymbol{c}^0 = \boldsymbol{x}$
初期サポート集合 $\mathcal{S} = \mathrm{supp}\{\boldsymbol{c}^0\} = \emptyset$
while $\|\boldsymbol{r}^t\|_2 \geq \epsilon$ **do**
残差計算: $j \notin \mathcal{S}$ に対して，基底 $\boldsymbol{d}_j$ を追加した時の残差を計算:
$$\epsilon(j) = \min_{z_j} \|\boldsymbol{d}_j z_j - \boldsymbol{r}^t\|_2^2$$
サポート更新: $j_0 = \arg\min_{j\notin\mathcal{S}} \epsilon(j)$ をサポートに追加:
$$\mathcal{S} = \mathcal{S} \cup \{j_0\}$$
暫定解の計算:
$$\boldsymbol{c}^{t+1} = \arg\min_{\boldsymbol{c}} \|D\boldsymbol{c} - \boldsymbol{x}\|_2^2 \quad \text{subject to} \quad \mathrm{supp}\{\boldsymbol{c}\} \subset \mathcal{S}$$
残差更新: $\boldsymbol{r}^{t+1} = \boldsymbol{x} - D\boldsymbol{c}^{t+1}$.
end while

Iterative Reweighted Least Square (IRLS)
IRLS は，$p \in [0, 1]$ に対して l_p ノルムを重み付き l_2 ノルムで繰り返し近似することで，l_p ノルム正則化問題を近似的に解く手法である．t 回目のくり返しで得られている係数ベクトルを $\boldsymbol{c}^t \in \mathbb{R}^m$ とする．これを用いて，重み行列を $W_t = \mathrm{diag}(|c_1^t|^{1-p/2}, \cdots, |c_m^t|^{1-p/2})$ で定義する．対角成分が 0 の時は逆行列の対応する成分もゼロにすることにして W_t^{-1} を定義すると，$\|W_t^{-1}\boldsymbol{c}^t\|_2^2 = \|\boldsymbol{c}^t\|_p^p$ となり，$\|W_t^{-1}\boldsymbol{c}\|_2^2$ は $\boldsymbol{c}$ の l_p ノルムを模していることが分かる．そこで，問題

$$\min_{\boldsymbol{c}} \quad \|W_t^{-1}\boldsymbol{c}\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{x} - D\boldsymbol{c}\|_2 = \boldsymbol{0} \tag{31}$$

を考えて，ラグランジュの未定乗数法により最適解

$$\boldsymbol{c}^{t+1} = W_t^2 D^\top (D W_t^2 D^\top)^\dagger \boldsymbol{x} \tag{32}$$

を得る．ただし，$M^\dagger$ は行列 M の一般化逆行列を表す．これを，残差のノルム $\|\boldsymbol{r}^t\|_2 = \|\boldsymbol{x} - D\boldsymbol{c}^t\|$ が ϵ 以下になるまで繰り返す．IRLS による係数最適化の手続きを Algorithm2 にまとめる．

OMP を始めとするスパース表現のための係数選択手法は広く用いられており，高速・高精度な計算方法や，理論的に興味深い手法が現在も開発されている．例えば [31] では，OMP を初期解として，指定した計算時間内で OMP による解の改善を木探索アルゴリズムに基づいて行う方法

Algorithm 2 IRLS: Iterative Reweighted Least Square

input:
辞書 $D \in \mathbb{R}^{d\times m}$
観測信号 $\boldsymbol{x} \in \mathbb{R}^d$
近似閾値 $\epsilon > 0$
ノルムの指数 $p \geq 0$
initialize:
繰り返しのカウンタ $t = 0$
初期係数ベクトル $\boldsymbol{c}^0 = \mathbf{1}$
初期重み行列 $W^0 = I_m$
while $\|\boldsymbol{r}^t\|_2 \geq \epsilon$ **do**
暫定解の計算: 次式で係数ベクトルを更新:

$$\boldsymbol{c}^{t+1} = W_t^2 D^\top (D W_t^2 D^\top)^\dagger \boldsymbol{x}$$

重み更新: $W_{jj}^{t+1} = (c_j^{t+1})^{1-p/2}, \quad j = 1, \ldots, m$
残差更新:

$$\boldsymbol{r}^{t+1} = \boldsymbol{x} - D\boldsymbol{c}^{t+1}$$

end while

が提案されており，計算時間と近似精度のトレードオフを利用者が調整できる様になっている．また，観測信号を基底表現における係数の情報が欠損した不完全データをみなすことで，EM アルゴリズムの枠組みで係数推定を捉える手法 [32] や，Bayes 推定の枠組みとして OMP を捉えた研究 [33] も行われている．

7. 基底学習の手法

基底の線型結合による信号の表現において，辞書のデザインは非常に重要である．離散コサイン変換や Fourier 変換，wavelet [3]，あるいは curvelet [34] のように予め決められた手段に従って基底を用意しておく方法と，信号から基底を学習する方法がある．一般に基底の学習には大きな計算コストがかかるが，近似対象の信号と類似した特徴を有する信号から学習した辞書を利用することで，スパース性が高く誤差の少ない近似が得られることが期待できる．

初期の研究においては，基底の学習は勾配法によって行われた [1]．計算効率や収束性の観点からは改善の余地の大きい方法であるが，Gabor wavelet 基底に似た局所的な基底が得られている．その後，信号処理の観点から，Engan らにより k-means 法に基づく基底学習手法である Method of Optimal Directions [35] が提案された．さらに k-means 法の一般化として Aharon らにより K-SVD [36] と呼ばれる基底学習アルゴリズムが提案された．K-SVD を含め多くの基底学習アルゴリズムは，固定された基底のもとで係数の最適化を行い，その係数を固定して基底を最適化することを繰り返す．このように基底と係数の交互最適化手法を用いる K-SVD の基本原理を以下に示す．係数行列 C は例えば OMP などを利用して求められているとする．基本的な考え方は，l 番目の基底 $\boldsymbol{d}_l$ を更新する際，$\boldsymbol{d}_l$ を利用しないで信号を近似した時の誤差を考え，この誤差を表現する基底を新たな $\boldsymbol{d}_l$ とするというものである． 具体的な手段としては，まず観測信号 $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ の中で，現在の基底 $\boldsymbol{d}_l$ が表現に用いられているような信号の添字集合

$$\Omega_l = \{i \in \{1, \ldots, n\} | [C]_{li} \neq 0\}$$

を求め，Ω_l に含まれる観測信号のみからなる観測信号行列，すなわち Ω_l に含まれる添字の列ベクトルからなる X の部分行列 X^{Ω_l} を構成する．更新対象の基底 $\boldsymbol{d}_l$ を利用しないで観測信号を近似した時の残差を

$$R_l = X^{\Omega_l} - \sum_{j\neq l} \boldsymbol{d}_j \boldsymbol{c}^j \tag{33}$$

とすると，$\|X^{\Omega_l} - DC\|_2^2 = \|R_l - \boldsymbol{d}_l \boldsymbol{c}^l\|_2^2$ であり，最適な基底 $\boldsymbol{d}_l$ は残差をランク 1 で l_2 ノルムの意味で最良近似すれば得られることが分かる．この最良近似は，残差行列 R_l に特異値分解を施し，その第一左特異ベクトル $\boldsymbol{u}_1$ を $\boldsymbol{d}_l$ とすれば良い．K-SVD は基底の学習アルゴリズムであるが，基底学習の際に合わせて係数の修正も $\boldsymbol{c}^j = \sigma_1 \boldsymbol{v}_1$ で行う．これを全ての基底ベクトル $\boldsymbol{d}_l, l = 1, \ldots, m$ について順次行うことで辞書 D を更新する．この基底の最適化と係数の最適化を，予め定めた基準を満たすまで繰り返す．この手続きを Algorithm3 にまとめる．この基底学習アルゴリズムと，OMP などの係数選択アルゴリズムとを繰り返し行うことで，観測信号に適した辞書を学習することが出来る．

Algorithm 3 K-SVD Dictionary Learning Algorithm

input:
辞書 $D \in \mathbb{R}^{d\times m}$
観測信号 $X = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$
係数行列 C
for $l = 1, \ldots, m$ **do**
添字集合 Ω_l を計算する
残差行列 $R_l = X^{\Omega_l} - \sum_{j\neq l} \boldsymbol{d}_j \boldsymbol{c}^j$ を計算する
特異値分解 $R_l = U\Sigma V^\top$ を行い，

$$\boldsymbol{d}_l \leftarrow \boldsymbol{u}_1, \quad \boldsymbol{c}_l \leftarrow \sigma_1 \boldsymbol{v}_1$$

で基底，係数を更新する
end for

K-SVD アルゴリズムは実問題に対して優れた性能を示しており，その数学的なシンプルさ，計算効率の高さも相まって，画像処理，音声信号処理におけるスパース表現のための基底学習手法として広く用いられている．

8. 画像処理への応用

本稿の最後に，画像のスパース表現の応用例を幾つか挙げる．基底展開による画像の表現の代表的な応用は画像圧縮及びノイズ除去であるが，他にも以下のような応用が可能である．

画像分離，画像修復 [37, 38]
２つの異なる画像が重なりあった画像 $\boldsymbol{x}$ が観測されるとする．この時，２つの異なる画像がそれぞれ異なる辞書

D_a, D_b と係数 $\boldsymbol{c}_a, \boldsymbol{c}_b$ で $D_a\boldsymbol{c}_a, D_b\boldsymbol{c}_b$ のように表現されているとすると，次の問題

$$\min_{\boldsymbol{c}_a,\boldsymbol{c}_b} \|\boldsymbol{c}_a\|_0 + \|\boldsymbol{c}_b\|_0 \text{ subject to } \|\boldsymbol{x} - D_a\boldsymbol{c}_a - D_b\boldsymbol{c}_b\|_2^2 \leq \epsilon$$

を解くことで重なりあった画像の分離を行うことが出来る．この問題は Morphological Component Analysis (MCA) と呼ばれ，特にテクスチャーと線素の分離，あるいは単純な独立ノイズではなく，構造を持ったノイズにより劣化した画像の修復などに応用されている．

超解像手法 [39, 40]

まず，高解像度の学習用画像から辞書 D_H を学習しておき，その辞書に対してボケ，ダウンサンプリングによる擬似的な劣化を施すことで低解像度画像を表現するための辞書 D_L を作成する．観測した複数枚の低解像度画像を辞書 D_L を用いて表現し，非ゼロの係数を記録しておく．この非ゼロの係数に対応する高解像度画像用の基底に係数をかけて足し合わせることで，高解像度の画像を得る．この超解像手法の概念図と，この手法により得られた高解像度画像の例を図 2 に示す．

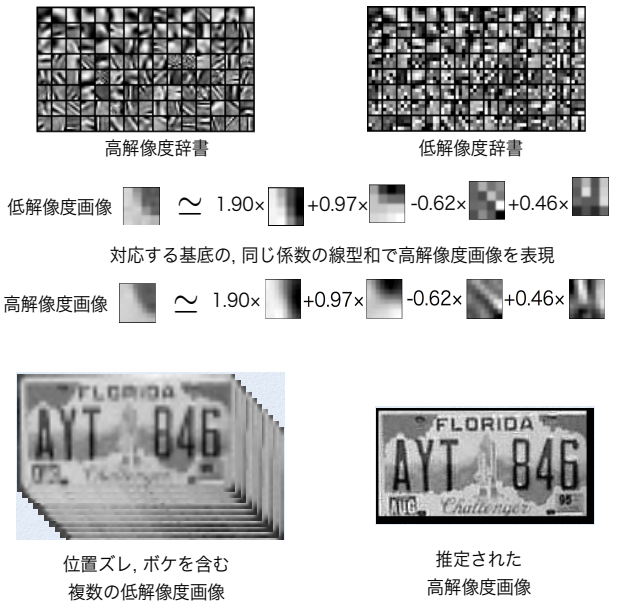

図 2 スパースコーディングによる超解像の概念図.

Fig. 2 A conceptual diagram of super-resolution based on sparse coding.

スパース表現を利用したロバストな顔認識 [41, 42]

顔画像による認証は古くからコンピュータビジョン及びセキュリティの重要な研究課題であり，スパース表現を利用した顔画像認証手法も多数提案されている．特に [41] では，スパース表現を直接的に利用した顔認識手法を提案している．認証システムに登録した人物が H 人いるとして，各人物は複数の顔画像 $\boldsymbol{f}_1^h, \ldots, \boldsymbol{f}_{n_h}^h,\ h = 1, \ldots, H$ を登録しているとする．これらの画像を基底として扱い，辞書

$$D = (\boldsymbol{f}_1^1, \ldots, \boldsymbol{f}_{n_1}^1; \cdots; \boldsymbol{f}_1^H, \ldots, \boldsymbol{f}_{n_H}^H)$$

を構成し，認証対象の画像をこの基底のスパースな線型結合で表現する．このときの結合係数は認証対象の人物の画像が辞書に含まれていれば，その人物の登録画像に対応する少数の係数のみが非ゼロの値を取ることが期待されるため，基底の結合係数から認証スコアを算出して認証を行うことが出来る．さらに [42] ではロバスト統計で用いられるような外れ値に対して頑健な損失関数を採用することで，より安定した認識を実現している．

9. 終わりに

本稿では，信号のスパース表現を行列の分解の形式で整理し，その確率モデルとしての解釈を試みた．信号をスパース表現する際に重要な係数及び基底の学習アルゴリズムを紹介し，画像処理における幾つかの応用例を挙げた．本稿では説明出来なかったが，スパースにサンプリングした信号からの原信号の復元可能性 [43] や，スパース正則化回帰問題におけるオラクル性 [44] など，圧縮センシングや統計学の分野ではスパース表現からの情報復元の可能性の研究が盛んに行われている．機械学習においてここ 10 年で大きく発展した Multiple Kernel Learning (MKL [45]) も，多数のカーネル関数によって表現される特徴量から有用なものをスパースな凸結合により選択する手法と捉えることが可能であり [46]，画像処理にも応用されている．スパース表現，あるいは行列の低ランク・スパース分解は現在でも盛んに研究されている分野である．スパース表現の範疇として取り扱える問題は非常に広く，画像処理の諸問題への有効なアプローチとしてもさらに発展が期待される．

参考文献

[1] B. A. Olshausen and D. J. Field: “Emergence of simple-cell receptive field properties by learning a sparse code for natural images”, Nature, **381**, pp. 607–609 (1996).
[2] B. A. Olshausen and D. J. Field: “How Close Are We to Understanding V1?”, Neural Computation, **17**, pp. 1665–1699 (2005).
[3] T. S. Lee: “Image representation using 2D Gabor wavelets”, IEEE Trans. Pattern Anal. Mach. Intell., **18**, 10, pp. 959–971 (1996).
[4] G. Palm: “On associative memory”, Biological Cybernetics, **36**, pp. 19–31 (1980).
[5] S. Amari: “Characteristics of sparsely encoded associative memory”, Neural Networks, **2**, 6, pp. 451–457 (1989).
[6] S. Akaho: “Capacity and error correction ability of sparsely encoded associative memory”, ICANN’93, pp. 707–710 (1993).
[7] S. Z. Li: “Markov random field modeling in image analysis”, Springer-Verlag (2001).
[8] K. Tanaka: “Statistical-mechanical approach to image processing”, Journal of Physics A: Mathematical and General, **35**, 37, pp. R81–R150 (2002).
[9] M. E. Tipping and C. M. Bishop: “Bayesian image super-resolution”, NIPS15, pp. 1303–1310 (2003).
[10] B. K. Natarajan: “Sparse approximate solutions to linear

systems", SIAM J. Comput., **24**, 2, pp. 227–234 (1995).
[11] M. Elad: "Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing", Springer-Verlag (2010).
[12] D. Harville: "Matrix Algebra from a Statistician's Perspective", Springer-Verlag (2008).
[13] I. Jolliffe: "Principal Component Analysis", Springer-Verlag (1986).
[14] H. Zou, T. Hastie and R. Tibshirani: "Sparse principal component analysis", Journal of Computational and Graphical Statistics, **15**, pp. 265–286 (2004).
[15] A. Hyvärinen, J. Karhunen and E. Oja: "Independent Component Analysis", John Wiley Sons, Inc. (2001).
[16] D. D. Lee and H. S. Seung: "Learning the parts of objects by non-negative matrix factorization", Nature, **401**, 6755, pp. 788–791 (1999).
[17] D. D. Lee and H. S. Seung: "Algorithms for non-negative matrix factorization", NIPS13, pp. 556–562 (2000).
[18] P. O. Hoyer: "Non-negative matrix factorization with sparseness constraints", J. Mach. Learn. Res., **5**, pp. 1457–1469 (2004).
[19] A. Cichocki, H. Lee, Y.-D. Kim and S. Choi: "Non-negative matrix factorization with alpha-divergence", Pattern Recognition Letters, **29**, 9, pp. 1433–1440 (2008).
[20] Y. Fujimoto and N. Murata: "Nonnegative matrix factorization via generalized product rule and its application for classification", LVA/ICA2012, pp. 263–271 (2012).
[21] T. J. Mitchell and J. J. Beauchamp: "Bayesian Variable Selection in Linear Regression", Journal of the American Statistical Association, **83**, 404, pp. 1023–1032 (1988).
[22] S. Mohamed, K. Heller and Z. Ghahramani: "Bayesian and l1 approaches to sparse unsupervised learning", ICML2012, pp. 721–728 (2012).
[23] M. K. Titsias and M. Lázaro-Gredilla: "Spike and slab variational inference for multi-task and multiple kernel learning", NIPS24, pp. 2339–2347 (2011).
[24] M. E. Tipping and C. M. Bishop: "Probabilistic principal component analysis", Journal of the Royal Statistical Society, Series B, **61**, pp. 611–622 (1999).
[25] B. Everitt: "An Introduction to Latent Variable Models", Monographs on Statistics and Applied Probability, Chapman and Hall (1984).
[26] Y. Guan and J. G. Dy: "Sparse probabilistic principal component analysis", J. Mach. Learn. Res. - Proceedings Track, **5**, pp. 185–192 (2009).
[27] T. M. Cover and J. A. Thomas: "Elements of information theory", John Wiley and Sons, Inc. (1991).
[28] M. N. Schmidt, O. Winther and L. K. Hansen: "Bayesian non-negative matrix factorization", ICA2009, pp. 540–547 (2009).
[29] Y. C. Pati, R. Rezaiifar, Y. C. P. R. Rezaiifar and P. S. Krishnaprasad: "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition", Asilomar1993, pp. 40–44 (1993).
[30] R. Chartrand and W. Yin: "Iteratively reweighted algorithms for compressive sensing", ICASSP2008, pp. 3869–3872 (2008).
[31] R. Rei, J. P. Pedroso, H. Hino and N. Murata: "A tree search approach to sparse coding", LION6 (2012, to appear).
[32] A. C. Gurbuz, M. Pilanci and O. Arikan: "Expectation maximization based matching pursuit", ICASSP2012, pp. 3313–3316 (2012).
[33] A. Dremeau, C. Herzet and L. Daudet: "Structured Bayesian orthogonal matching pursuit", ICASSP2012, pp. 3625–3628 (2012).
[34] E. Candès and D. Donoho: "Curvelets: A surprisingly effective nonadaptive representation for objects with edges", Curves and Surfaces (Ed. by L. L. Schumaker et al.), Vanderbilt University Press (1999).
[35] K. Engan, S. O. Aase and J. Hakon Husoy: "Method of optimal directions for frame design", ICASSP1999, pp. 2443–2446 (1999).
[36] M. Aharon, M. Elad and A. Bruckstein: "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation", IEEE Trans. Sig. Proc., **54**, 11, pp. 4311–4322 (2006).
[37] J. Bobin, J. L. Starck, J. M. Fadili, Y. Moudden and D. L. Donoho: "Morphological component analysis: An adaptive thresholding strategy", IEEE Trans. Img. Proc., **16**, 11, pp. 2675–2681 (2007).
[38] M. Elad, J. Starck, P. Querre and D. Donoho: "Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA)", Applied and Computational Harmonic Analysis, **19**, 3, pp. 340–358 (2005).
[39] Y. Ueda, T. Ohta, A. Iwase, M. Seki and N. Murata: "Super-resolution based on sparse coding", Proceedings of Computational Algebraic Statistics, Theories and Applications (2008).
[40] J. Yang, J. Wright, T. S. Huang and Y. Ma: "Image super-resolution via sparse representation", IEEE Trans. Image Proc., **19**, 11, pp. 2861–2873 (2010).
[41] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry and Y. Ma: "Robust face recognition via sparse representation", IEEE Trans. Pattern Anal. Mach. Intell., **31**, 2, pp. 210–227 (2009).
[42] M. Yang, L. Zhang, J. Yang and D. Zhang: "Robust sparse coding for face recognition", CVPR2011, pp. 625–632 (2011).
[43] E. J. Candès and T. Tao: "Decoding by linear programming", IEEE Trans. Inf. Theo., **51**, 12, pp. 4203–4215 (2005).
[44] J. Fan and R. Li: "Variable selection via nonconcave penalized likelihood and its oracle properties", Journal of the American Statistical Association, **96**, 456, pp. 1348–1360 (2001).
[45] A. Rakotomamonjy, F. R. Bach, S. Canu and Y. Grandvalet: "SimpleMKL", J. Mach. Learn. Res., **9**, pp. 2491–2521 (2008).
[46] F. R. Bach: "Consistency of the group lasso and multiple kernel learning", J. Mach. Learn. Res., **9**, pp. 1179–1225 (2008).