白色化

はじめに

白色化とはICA(独立成分分析)の工程の半分を成すとされる[1]．筆者としては, ICAが有効だったという経験がない．理由として，初期値により解が変化する，あるいは解が複数出現する，ということがあった．とはいえ，そもそも ICA が有効となるデータではなかった，あるいは自作プログラム(C++)のバグだったかもしれない．

さておき，一般に ICA の前処理として PCA(主成分分析)，白色化が適用される．本格的な ICA を使わずとも，この前処理にはメリットがある．

＊解の安定性

＊解の一意性

本稿では，この白色化の有効性を述べる．

PCA(主成分分析)

主成分分析について簡単に述べる．主成分分析を一言で述べれば固有値問題である．具体的には，下記の手順である．

① 取得したデータ系列の分散，共分散を算出し，共分散行列Cを構成
共分散の算出式は説明するまでもなく，本稿では割愛する．なお，筆者の場合，PCA では不偏分散(もう一つは母分散)でプログラムの記述を統一している．

② 共分散行列Cの固有値のベクトルLを求める(1)

$$L = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} \quad (1)$$

n: 次数(データ系列の個数)

数値計算ではハウスホルダー変換によりヘッセンベルグ行列を算出し，これを QR 分解する．対称となる共分散では，対称行列版のハウスホルダー変換で十分である(非対称行列版のアルゴリズムもある)．ヤコビ法の実装もあるが，筆者はヤコビ法を用いることはない．アルゴリズムの詳細については割愛する．

③ 固有値から固有ベクトルの行列Eを求める．これは(2)の求解による

$$CV_i = \lambda_i I \quad (2)$$

$$E = [V_1 \quad V_2 \quad \cdots \quad V_n]$$

$$V_i = \begin{bmatrix} v_{i1} \\ v_{i2} \\ \vdots \\ v_{in} \end{bmatrix}$$

数値計算による求解はガウスの消去法でよい．

結局，共分散行列の固有値分解は(3)と表せる．

$$C = EDE^{\top} \quad (3)$$

$$D = \mathrm{diag}(L) = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

ここで，**固有値はデータ系列毎の分散値となっていて，PCA の機能美が現れる**部分である．ここでユニタリ行列Uの性質として(4)がある．

$$(UDU^{\top})^{-1} = UD^{-1}U^{\top} \quad (4)$$

この性質により，固有値分解から逆行列を求めることができる．ちょっとだけ脱線するが，数値計算で逆行列を求めることはあまり歓迎されない．とはいえ，動作確認で使用することもある．しばしば利用するのがガウス・ジョルダンである．これはガウスの消去法の拡張版のような手順である．簡単なアルゴリズムなのだが，ガウスの消去法と同様に枢軸変換は必須である．

白色化

白色化行列Wは(5)で定義される．

$$W = C^{-\frac{1}{2}} = (EDE^{\top})^{-\frac{1}{2}} = ED^{-\frac{1}{2}}E^{\top} \quad (5)$$

$$D^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\sqrt{\lambda_n}} \end{bmatrix}$$

対角成分の平方をとる部分だが，これは分散値 1 とする正規化である．ここで白色化とは観

測データに白色化行列を乗ずることをいう(後述).

カクテルパーティ問題

独立成分分析のお題として有名な，あのカクテルパーティ問題である(**図 1**).

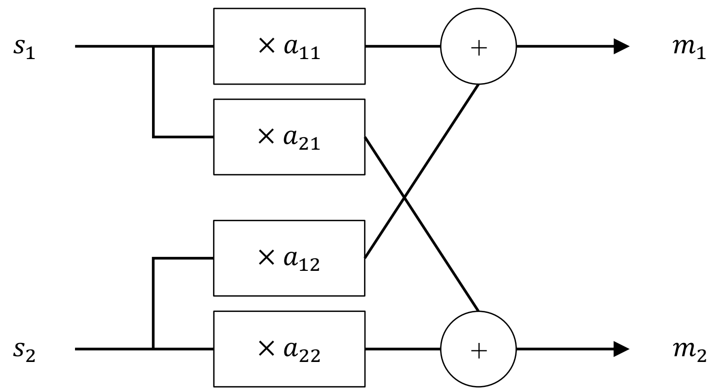

図 1　カクテルパーティ問題

音声の信号源s_1, s_2があり，これを 2 つのマイクm_1, m_2で拾う構成である.これを行列で表せば(6)である.

$$AS = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = M \quad (6)$$

問題としては，MからAとSを推定するということである. (6)では 2 変数であるが任意の多変数としてよい. シミュレーションでは，本題を扱う. 白色化だが，これは観測データを(7)で変換する.

$$WM \quad (7)$$

白色化によりデータは無相関化される. 無相関化とは共分散値を 0(零)とすることである. 一方，対角成分は(5)の正規化により全て 1 となる.

白色化のシミュレーション

筆者には ICA に有効なデータの手持ちがないということがある．適当な周期波形で白色化の効果を確認する(8)～(10).

条件:

$$s_1 = \cos\left(2\pi \frac{i}{125}\right) + 2 \quad (8)$$

$$s_2 = 0.01(i\,\%100) + 2 \quad (9)$$

$$i = 0, 1, 2, \cdots, 499$$

$$A = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.9 \end{bmatrix} \quad (10)$$

各信号を**図 2**～**図 5** で示す.

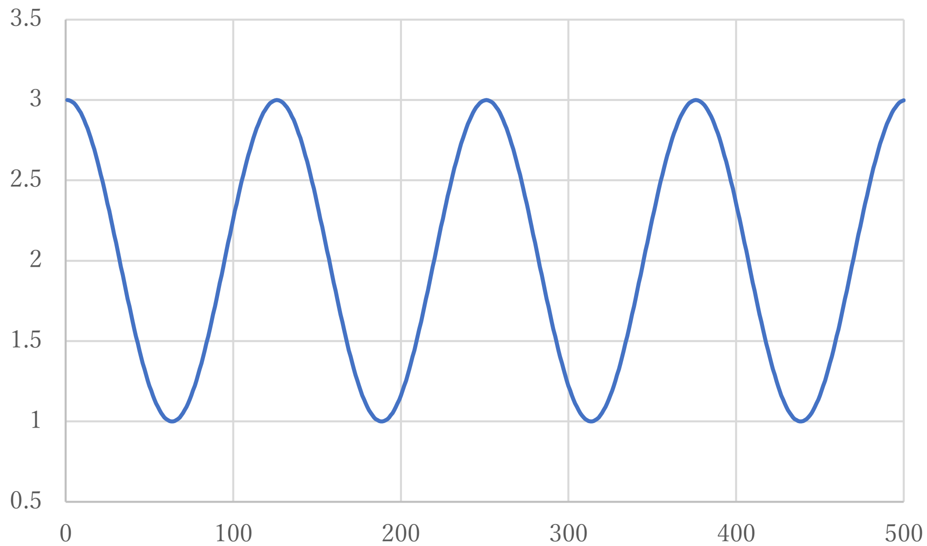

図 2　s_1

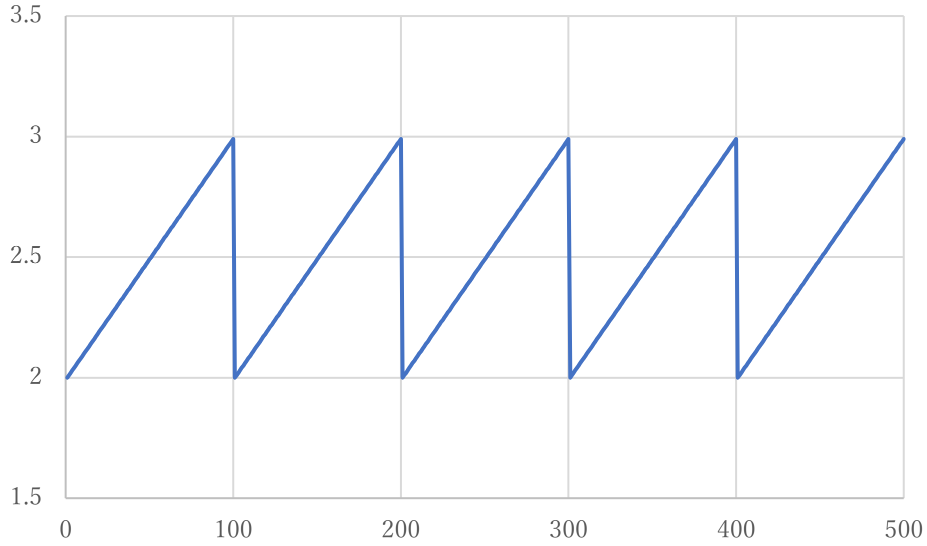

図 3　s_2

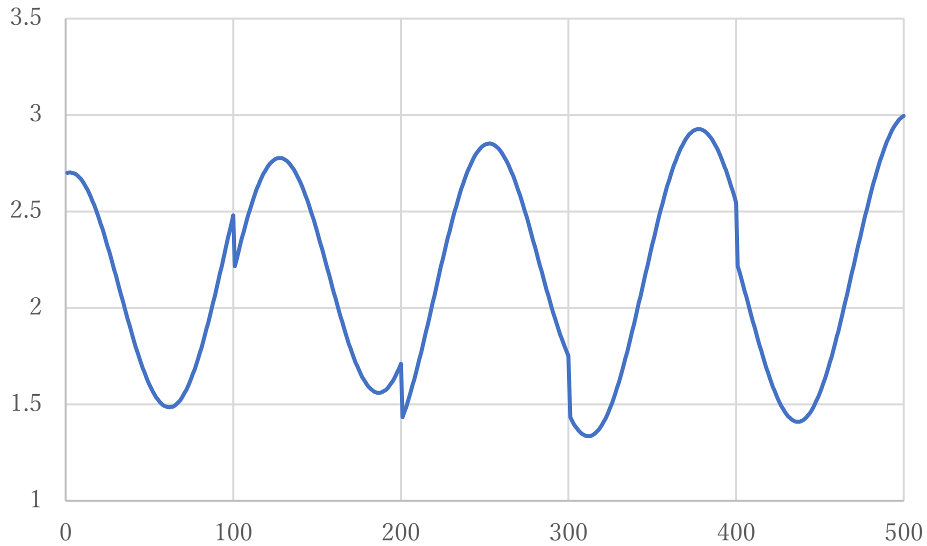

図 4　m_1

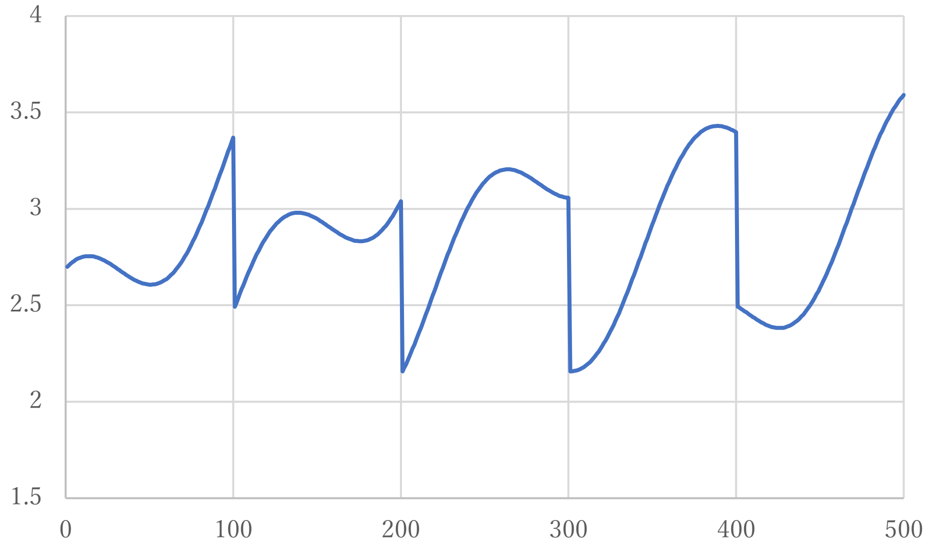

図 5　m_2

シミュレーション結果

白色化の結果を**図 6**，**図 7** で示す．

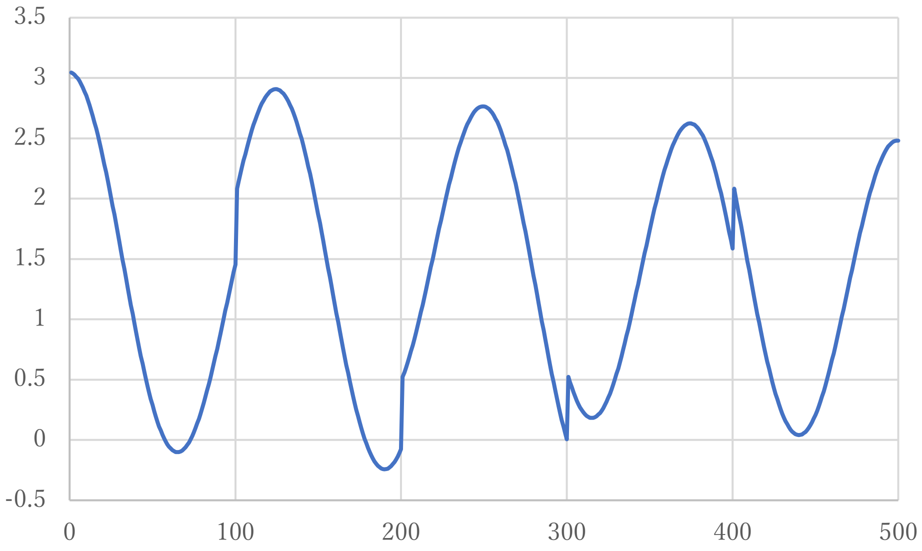

図 6　白色化によるs_1の推定

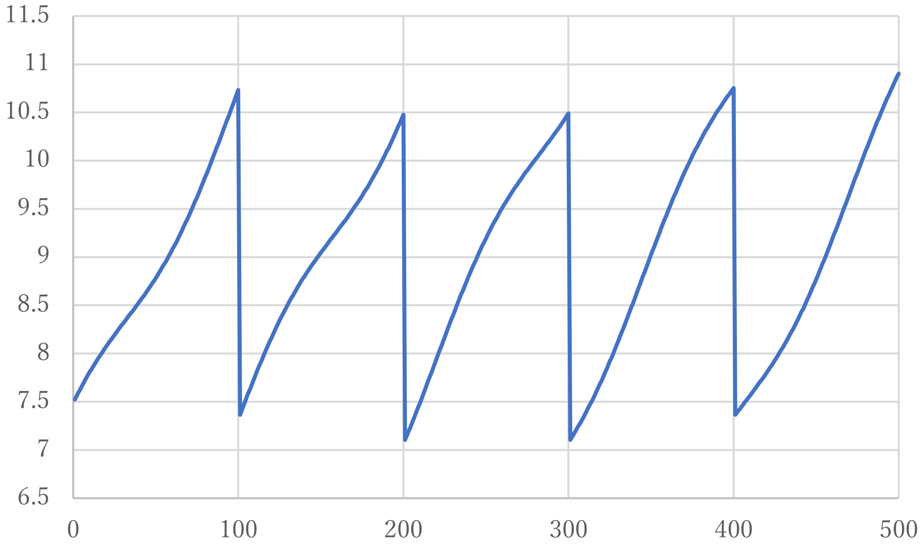

図 7　白色化によるs_2の推定

もともと(6)において，AとSは未知であり，Sの絶対値を知ることはできない．故に信号の大きさではなく，信号の形状に注目いただきたい．白色化により完全ではないまでも，もとの信号に似た形状を推定している．

3 信号のシミュレーション

もう一つ信号を加えて 3 信号とし，シミュレーションを実施する(11), (12)，**図 8〜図 11**.

条件:

$$s_3 = (i\,\%25 > 12)?\,2:1 \quad (11)$$

$$A = \begin{bmatrix} 0.7 & 0.3 & 0.5 \\ 0.2 & 0.5 & 0.8 \\ 0.1 & 0.7 & 0.3 \end{bmatrix} \quad (12)$$

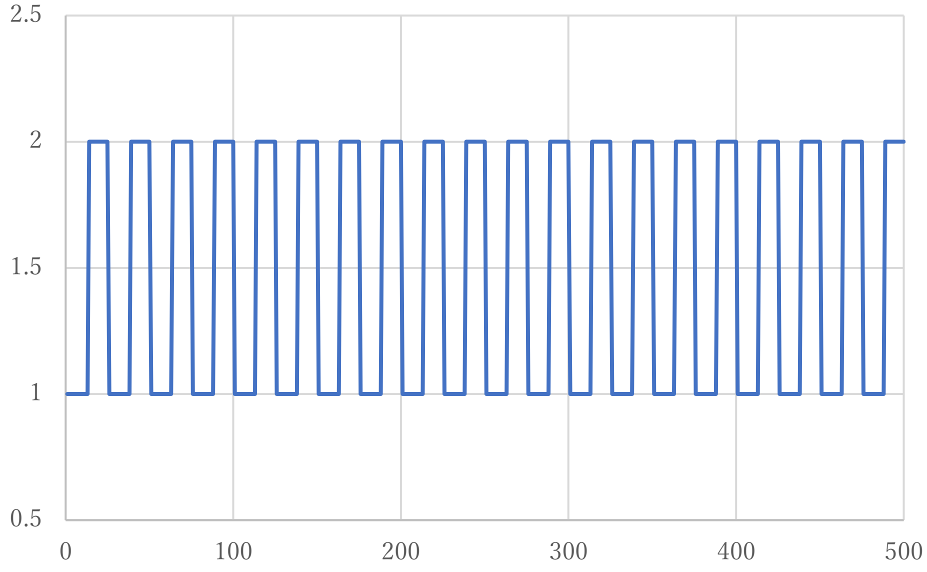
2.5
2
1.5
1
0.5
0
100
200
300
400
500

図 8　s_3

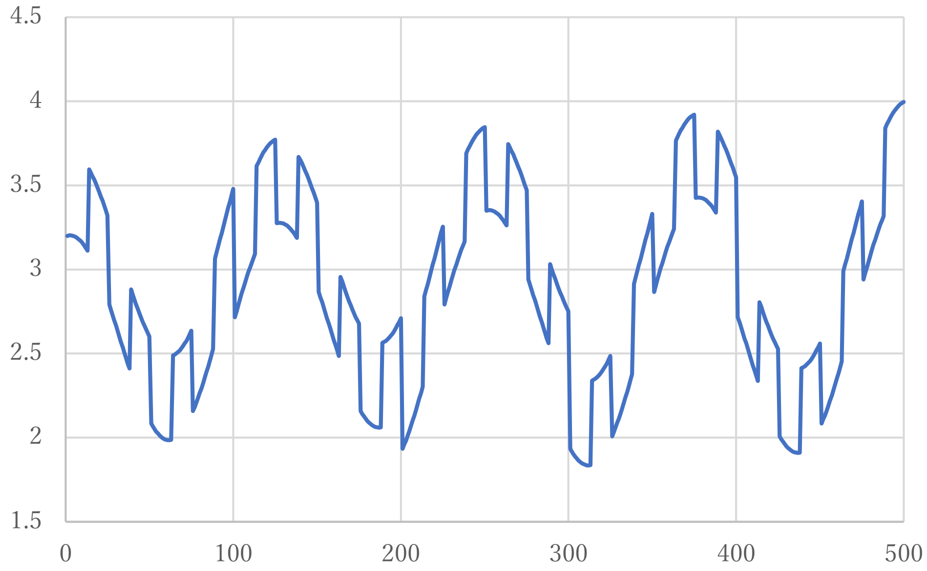
4.5
4
3.5
3
2.5
2
1.5
0
100
200
300
400
500

図 9　m_1

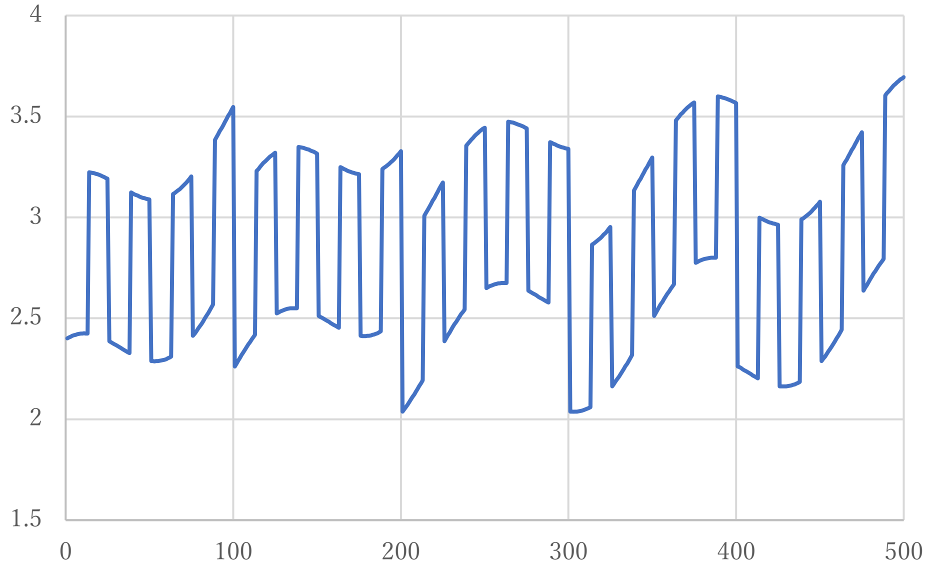

図 10　m_2

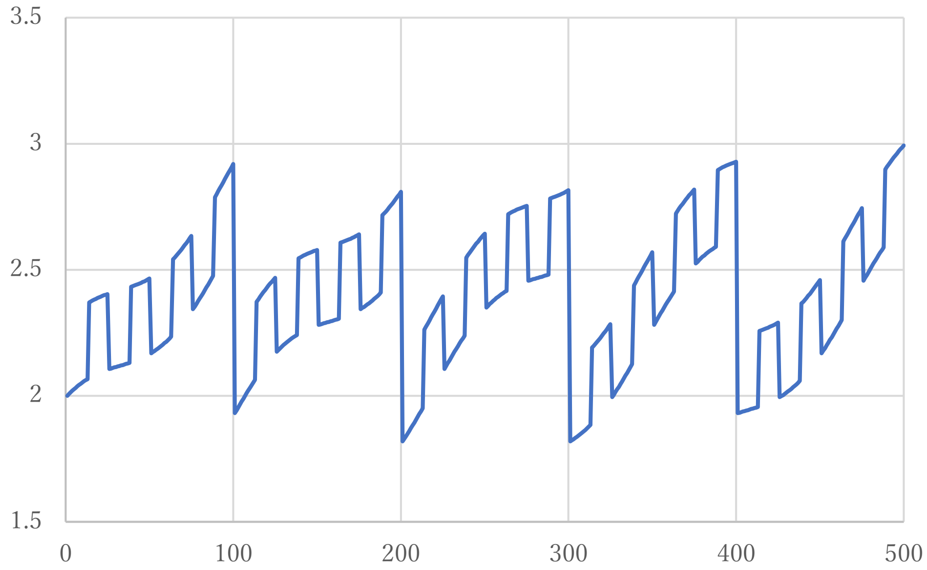

図 11　m_3

3 信号のシミュレーション結果

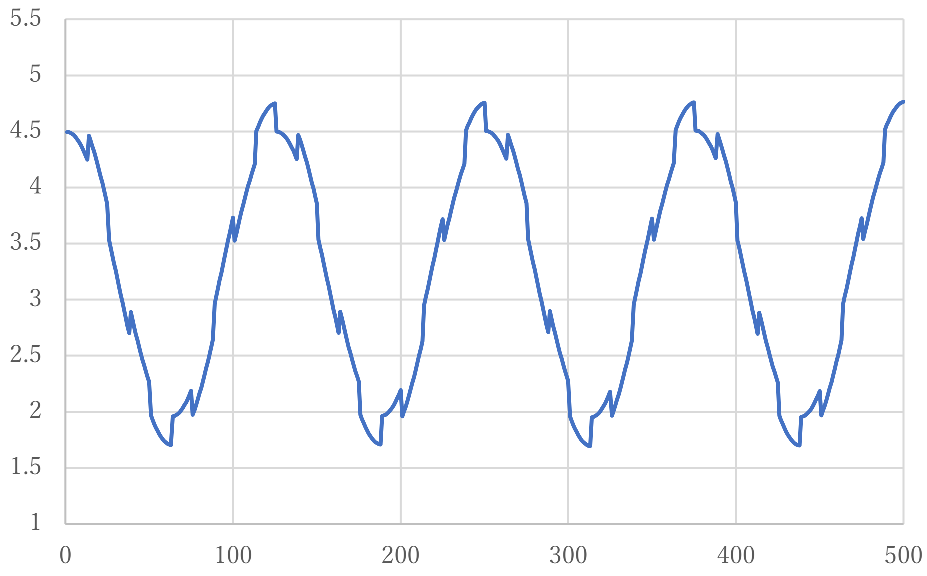

図 12　白色化によるs_1の推定

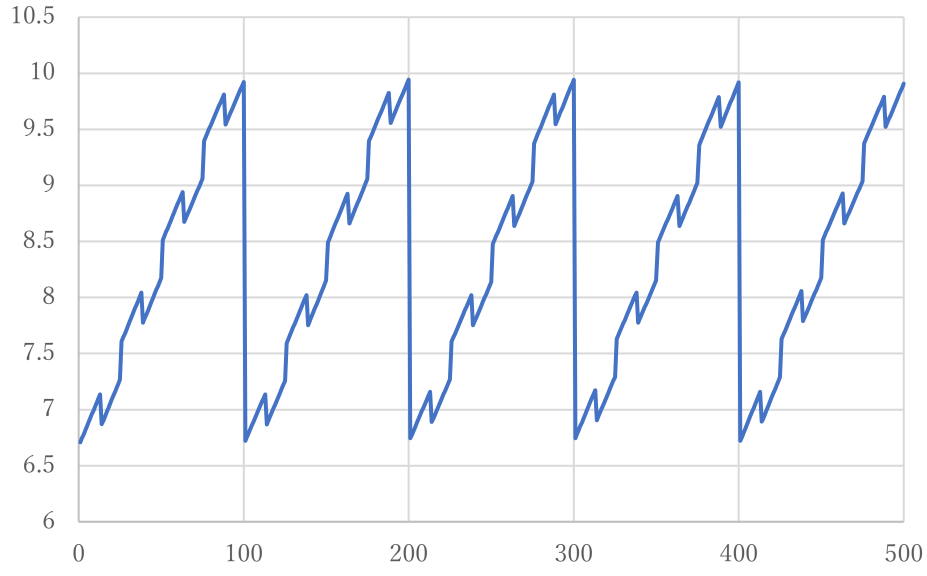

図 13　白色化によるs_2の推定

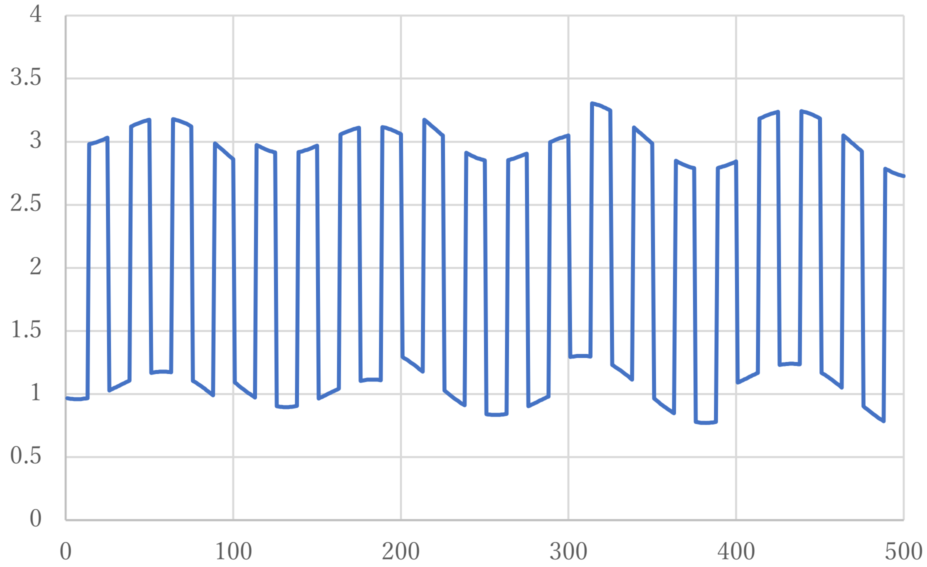

図 14　白色化によるs_3の推定

やはり，完全ではないまでも，もとの信号に似た形状を推定している．

まとめ

完全に分離できてないとの指摘もあろうが，白色化には一定の効果がある．筆者からすると，特徴抽出として十分に有効なことがある．主成分分析と白色化は 信号の相関，無相関に基づく．独立ということではなく，周期波形では周期に依存した相関があり，これが関与するのかと思う．

本件のシミュレーションデータではICA(最尤推定のためのfastICA[1])を適用しても，これ以上の効果が認められなかった．波形データとしては，いずれも劣ガウス分布(ほぼ一様分布)であり，この点に問題はない．自作プログラムなのだが，2つの優ガウス分布(一様乱数から逆関数法により指数分布を作成)からシミュレーションした．初期値にもよるが，独立成分の回転が機能することは確認できており，バグの可能性は低い．現状，筆者としては，これ以上の関心を失っている．

補足として，本稿のシミュレーションデータでは白色化よりも，その周期性からスペクトル(周波数変換)を用いた方がよいかと思う．信号に周期性があり，信号毎に周期が異なるという前提であれば，Aを特定することは難しくない．問題解決にあたり，手法は一つに限定されるものではない．実際，筆者が特許を取得した画像処理では，ほとんどが手法の組み合わせ，応用にすぎない．

その他，主成分分析については[2]を参照いただきたい．筆者が大変お世話になった文献である．和書であるが，これ以上ない，というほどに解りやすい解説となっている．

参考文献

[1] Aapo Hyvärinen, Juha Karuhnen, Erkki Oja　　詳解 独立成分分析
東京電機大学出版局.　2005.

[2] 涌井良幸，涌井貞美　　図解でわかる 多変量解析
日本実業出版社.　2001.

.