

残差ベクトルの ℓ_1 ノルムを最小化する線形予測分析の解法*

亀岡弘和, 鎌本優, 原田登, 守谷健弘

(NTT コミュニケーション科学基礎研究所)

1 序論

線形予測分析 (LPC) に基づく時系列信号の可逆圧縮符号化方式では、LPC により生成される残差信号の符号化にエントロピー符号化 (Golomb-Rice 符号など) がしばしば用いられ、出現頻度の高い振幅値ほど短い符号が割り当てられる仕組みとなっている。Golomb-Rice 符号 (以後、Rice 符号) の場合、残差振幅に割り当てられる符号量は、振幅の絶対値とほぼ比例関係にあるため、残差符号量は残差振幅の絶対値和である程度良く近似される。しかしながら、従来の LPC による可逆圧縮符号化方式では、残差振幅の二乗和を最小化するように予測係数を求めるため、残差符号量を直接的に最小化する規準での最適化となっていない。従って、予測係数の算出方法には符号量削減の余地が少なからず残されている。

LPC を含む線形回帰分析において、誤差ベクトルの要素の絶対値和を最小化する回帰係数の解法は、シンプレックス法に基づくもの [1] が古典的で有名である。しかし、この方法には、説明変数行列や回帰係数ベクトルのサイズが大きいうちに計算量が膨大となるという特徴があり、即時性が求められる圧縮符号化への応用には適していないという問題があった。

本稿では、誤差ベクトルの要素の絶対値和を最小化する回帰係数を効率的に求める手法の導出を出発点とし、Rice 符号量関数を絶対値和に比べてより良く表現するコスト関数をいくつか例示したのちに、同じ原理に基づいてこれらを最小化する回帰係数も同様に求められることを示す。

2 絶対誤差を最小化する回帰係数の解法

2.1 線形回帰分析

\mathbf{x}, \mathbf{s} をそれぞれ I 次元および J 次元ベクトル、 \mathbf{A} を $I \times J$ 行列とし、次のような線形回帰モデル

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{z} \quad (1)$$

を仮定する。ただし、 \mathbf{z} は雑音を表す確率ベクトルである。 \mathbf{x}, \mathbf{A} が与えられたもとで、 \mathbf{s} の最尤推定値を求める線形回帰分析の問題を考える。 \mathbf{z} にもし Gauss 性白色雑音を仮定すると、これは ℓ_2 ノルム $\|\mathbf{x} - \mathbf{A}\mathbf{s}\|_2^2$ を \mathbf{s} に関して最小化する問題であり、 $I \geq J$ ならば

$$\hat{\mathbf{s}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{x} \quad (2)$$

が最適解となる。これに対し、我々がここで考えるのは、 \mathbf{z} の各要素が互いに独立に Laplace 分布に従う場

合を仮定した \mathbf{s} の最尤解を求める問題である。すなわち、 ℓ_1 ノルム (絶対誤差)

$$L(\mathbf{s}) \triangleq \|\mathbf{z}\|_1 = \|\mathbf{x} - \mathbf{A}\mathbf{s}\|_1 = \sum_{i=1}^I |x_i - (\mathbf{A}\mathbf{s})_i| \quad (3)$$

を最小化する \mathbf{s} がこの場合の最尤解に対応する。この推定量のことは ℓ_1 推定量と呼び、 \mathbf{x} の外れ値に頑健なロバスト推定量の 1 つとして知られる。

2.2 線形予測分析 (LPC)

線形予測モデルは式 (1) の特殊例である。線形予測モデルは、あるデジタル信号の時刻 i におけるサンプル y_i を、時刻 $i-1$ から過去 J 個のサンプル $y_{i-1}, y_{i-2}, \dots, y_{i-J}$ の線形結合で表現するモデル

$$y_i \approx \sum_{j=1}^J s_j y_{i-j} \quad (4)$$

のことであり、自己回帰モデルとも呼ばれる。上記のモデルのもとで、 I 個の連続するサンプル系列 y_1, \dots, y_I を最も良く表す線形予測モデルの J 個の係数を推定する問題を線形予測分析と呼ぶ。式 (1) に対応させると、線形予測モデルは、

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_I \end{pmatrix} = \begin{pmatrix} y_0 & y_{-1} & \cdots & y_{1-J} \\ y_1 & y_0 & \cdots & y_{2-J} \\ \vdots & \vdots & & \vdots \\ y_{I-1} & y_{I-2} & \cdots & y_{I-J} \end{pmatrix} \begin{pmatrix} s_1 \\ \vdots \\ s_J \end{pmatrix} + \mathbf{z} \quad (5)$$

と置いたことに相当し、 \mathbf{z} が予測残差である。

2.3 補助関数法の原理

提案法の特徴は、式 (3) の最小化に、補助関数法と呼ぶ原理を用いる点にある。補助関数法とは、適切に設計した補助関数を反復的に減少させていくことで目的関数を間接的に最小化する手法であり、EM アルゴリズムを一般化した原理である。EM アルゴリズム同様に補助関数法はいかなる最適化問題に適用可能というわけではないが、ある要件を満たす補助関数が設計できれば大変効率の良い最適化アルゴリズムが導ける場合がある。以下に、補助関数法の原理を示し、補助関数が満たすべき要件を明らかにする。

目的関数 $L(\mathbf{s})$ に対し、

$$L(\mathbf{s}) = \min_{\mathbf{B}} L^+(\mathbf{s}, \mathbf{B}) \quad (6)$$

* A Solution to the Linear Predictive Coding Minimizing ℓ_1 Norm of the Residual Vector. by KAMEOKA Hirokazu, KAMAMOTO Yutaka, HARADA Noboru and MORIYA Takehiro (NTT Communication Science Laboratories)

が成り立つとき、 $L^+(s, \mathbf{B})$ を $L(s)$ の補助関数、 \mathbf{B} を補助変数と定義すると、次の定理が成り立つ。

定理 1 (補助関数法). 補助関数 $L^+(s, \mathbf{B})$ を、 \mathbf{B} に関して最小化するステップと、 s に関して最小化するステップを繰り返すことで、目的関数 $L(s)$ を単調減少させていくことができる。

証明: $\mathbf{B}_{t+1} = \operatorname{argmin}_{\mathbf{B}} L^+(s_t, \mathbf{B})$ とし、 $s_{t+1} = \operatorname{argmin}_s L^+(s, \mathbf{B}_{t+1})$ とする。ただし、 t は反復計算のステップ数を表す。 $s = s_t$, $\mathbf{B} = \mathbf{B}_t$ から $s = s_{t+1}$, $\mathbf{B} = \mathbf{B}_{t+1}$ に更新されたときに、 $L(s)$ が非増加することを示す。明らかに $L(s_t) = L^+(s_t, \mathbf{B}_{t+1})$ であり、 $s_{t+1} = \operatorname{argmin}_s L^+(s, \mathbf{B}_{t+1})$ より、 $L^+(s_t, \mathbf{B}_{t+1}) \geq L^+(s_{t+1}, \mathbf{B}_{t+1})$ である。さらに補助関数の定義より、 $L^+(s_{t+1}, \mathbf{B}_{t+1}) \geq L(s_{t+1})$ なのだから、結局、 $L(s_t) \geq L(s_{t+1})$ である。□

以上より、補助関数法を効果的に適用するには、次の 2 点を満たす補助関数を設計する必要がある。

1. $\operatorname{argmin}_{\mathbf{B}} L^+(s, \mathbf{B})$ が解析的に求められる。
2. $\operatorname{argmin}_s L^+(s, \mathbf{B})$ が解析的に求められる。

これを踏まえ、式 (3) の最小化問題を解くための補助関数を以下で設計する。

2.4 2 次形式の補助関数の導出

式 (3) 最小化のための補助関数を設計するにあたり、絶対値関数より小さくなることなく (補助関数の定義より)、かつ、絶対値関数に接する (要件 1 より) 2 次関数 (要件 2 より) を見つけたい。この中で 2 次の係数が最小となるのは、原点に関して対称なものである。絶対値関数 $f_0: \mathbb{R} \rightarrow \mathbb{R}_+$

$$f_0(z) = |z| \quad (7)$$

に $b \neq 0$ で接する原点に関して対称な 2 次関数は

$$h_0(z) = \frac{z^2}{2|b|} + \frac{|b|}{2} \quad (8)$$

で与えられる。 $\forall z, f_0(z) \leq h_0(z)$ であるので、

$$|z| \leq \frac{z^2}{2|b|} + \frac{|b|}{2} \quad (9)$$

が成り立つ。なお、言うまでもなく上不等式の等号成立は接点 $z = b$ においてである。

この不等式を式 (3) に当てはめると、

$$\begin{aligned} L(s) &\leq \sum_{i=1}^I \left(\frac{(x_i - (\mathbf{A}\mathbf{s})_i)^2}{2|b_i|} + \frac{|b_i|}{2} \right) \quad (10) \\ &= (\mathbf{x} - \mathbf{A}\mathbf{s})^T \mathbf{B}(\mathbf{x} - \mathbf{A}\mathbf{s}) + d \equiv L^+(s, \mathbf{B}) \end{aligned}$$

のような不等式が立てられる。ただし、

$$\mathbf{B} = \begin{pmatrix} \frac{1}{2|b_1|} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{2|b_I|} \end{pmatrix}, \quad d = \frac{1}{4} \operatorname{Tr}(\mathbf{B}^{-1}) \quad (11)$$

である。以上より、 $L^+(s, \mathbf{B})$ は補助関数の定義を満たす。次に、 $L^+(s, \mathbf{B})$ が補助関数の要件を満たすかどうかを確認する。

2.5 補助関数法の更新則

まず、 $L(s) = L^+(s, \mathbf{B})$ となる \mathbf{B} は明らかに、

$$\mathbf{B} = \begin{pmatrix} \frac{1}{2|x_1 - (\mathbf{A}\mathbf{s})_1|} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{2|x_I - (\mathbf{A}\mathbf{s})_I|} \end{pmatrix} \quad (12)$$

である。一方、 \mathbf{B} 固定のときに $L^+(s, \mathbf{B})$ を最小化する s は、式 (12) より、大きい残差には小さい重みを、小さい残差には大きい重みを課す重みつき最小二乗解に相当し、正規方程式の解

$$\mathbf{s} = (\mathbf{A}^T \mathbf{B} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{B} \mathbf{x} \quad (13)$$

で与えられる。 $\mathbf{A}^T \mathbf{B} \mathbf{A}$ は正定値対称行列であるので、 $(\mathbf{A}^T \mathbf{B} \mathbf{A})^{-1}$ は Cholesky 分解で求めると良い。以上より、 $L^+(s, \mathbf{B})$ は補助関数の要件を満たし、次の手順

1. s の初期値設定
2. s と \mathbf{B} を収束するまで以下のように交互に更新

$$\begin{aligned} \mathbf{B} &\leftarrow \begin{pmatrix} \frac{1}{2|x_1 - (\mathbf{A}\mathbf{s})_1|} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{2|x_I - (\mathbf{A}\mathbf{s})_I|} \end{pmatrix} \\ \mathbf{s} &\leftarrow (\mathbf{A}^T \mathbf{B} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{B} \mathbf{x} \end{aligned}$$

により、回帰係数の絶対誤差最小解を得ることができる。式 (3) は s に関して凸なので、収束解は大域最適解に一致する。ただし、大域最適解は一意とは限らず (大域最適解の集合は解空間上で凸集合をなし)、提案法はその中の 1 点を求めるものである。以上のアルゴリズムは、ロバスト推定の 1 つである Huber の M 推定における反復重みつき最小二乗誤差推定 (IRLS) 法 [2] と同じ手続きをとる。 ℓ_1 推定量を得るための上記手順においては Huber の IRLS 法との間に違いはないが、次章で両者の違いを明らかにする。

3 Rice 符号量モデルとその最小化

残差振幅 $z \in \mathbb{Z}$ に対する Rice 符号量 $R(z)$ は、

$$R(z) = \begin{cases} \lfloor \frac{z}{2^{r-1}} \rfloor + r + 1 & (z \geq 0) \\ \lfloor \frac{-z-1}{2^{r-1}} \rfloor + r + 1 & (z < 0) \end{cases} \quad (14)$$

と表せる。ただし、 $\lfloor \cdot \rfloor$ は切捨て整数化演算子を表し、 $r \in \mathbb{N}$ は Rice パラメータである。2 章で最小化しようとした残差振幅の絶対誤差は、 $r = 1$ と置いた場合の Rice 符号量のモデルに相当する。しかし Rice 符号化では実際のところ $r > 1$ と設定することが多く、式 (14) は図 1 のように階段形の関数となる。ここで特

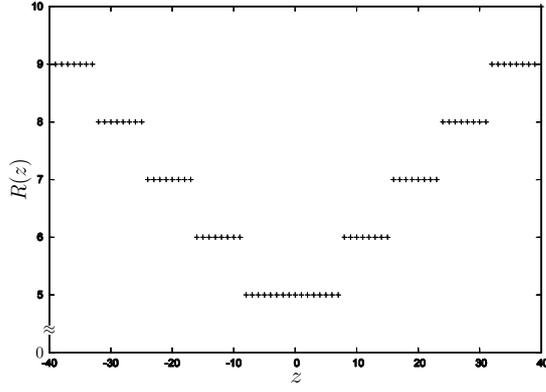


Fig. 1 Rice 符号量 $R(z)$ ($r = 4$)

に注目すべきは、 $z = 0$ 周辺の一定範囲内の値には等しい符号量が割り当てられる点であり、このことは、残差振幅の絶対値をできるだけ小さくすることだけに専念するよりも、残差振幅に適度な「遊び」を持たせ、その自由度の分だけ極端な外れ値を少なくする方が全体の符号量をより小さくできることを示している。残差振幅が通常 0 付近に集中する点を考慮すると、コスト関数をモデル化する際、 $z = 0$ 周辺の $R(z)$ を良く近似したものであった方が良い。

以上の問題意識に基づき、絶対値関数に代わる次のような 3 種類の Rice 符号量モデル $f_1(z)$, $f_2(z)$, $f_3(z)$

$$f_1(z) \triangleq \begin{cases} \frac{z^2}{4\beta\gamma} + \frac{\beta}{\gamma} & (|z| \leq 2\beta) \\ |z|/\gamma & (|z| > 2\beta) \end{cases}$$

$$f_2(z) \triangleq \frac{\sqrt{z^2 + \beta^2}}{\gamma}$$

$$f_3(z) \triangleq \begin{cases} \beta/\gamma & (|z| \leq \beta - c) \\ \frac{(z - \text{sign}(b)(\beta - c))^2}{4c\gamma} + \frac{\beta}{\gamma} & (\beta - c < |z| \leq \beta + c) \\ |z|/\gamma & (|z| > \beta + c) \end{cases}$$

を導入する。 $\beta/\gamma = r + 1$ 、 $\gamma = 2^{r-1}$ であり、 β, γ は Rice パラメータに依存して一意に決まる定数である。 c は $0 < c \leq \beta$ の範囲で適当に定める定数であり、 $c \rightarrow 0$ で $f_3(z)$ は Vapnik's ε -insensitive [3] に漸近する。また、 $c = \beta$ で $f_1(z) = f_3(z)$ となる。各々の概形を図 2 に示す。以上の符号量モデルを用い、残差符号の全符号量を近似した 3 種類の目的関数

$$L_n(\mathbf{s}) = \sum_{i=1}^I f_n(x_i - (\mathbf{A}\mathbf{s})_i) \quad (n = 1, 2, 3) \quad (15)$$

が立てられる。いずれの目的関数においても \mathbf{s} の最適解は γ に依らないので、以後簡単の為 $\gamma = 1$ とする。なお、これらはいずれも \mathbf{s} に関してやはり凸であり、局所解と大域最適解は一致する。

各々のコスト関数 $f_n(z)$ ($n = 1, 2, 3$) に対し、2 章と同様の方針で、2 次関数 $h_n(z)$

$$h_1(z) = \frac{z^2}{2 \max(|b|, 2\beta)} + \frac{\max(|b|, 2\beta)}{2}$$

$$h_2(z) = \frac{z^2}{2\sqrt{b^2 + \beta^2}} + \frac{b^2 + 2\beta^2}{2\sqrt{b^2 + \beta^2}}$$

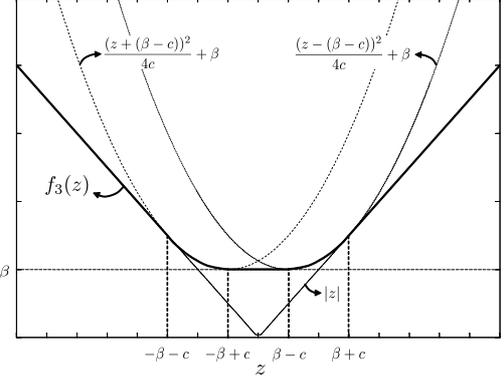
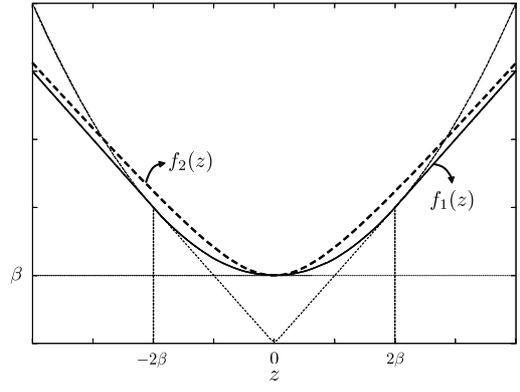


Fig. 2 Rice 符号量モデル $f_n(z)$ ($n = 1, 2, 3$)

$$h_3(z) = \begin{cases} \frac{z^2}{2|b|} + \frac{|b|}{2} & (|b| \geq 2\beta) \\ \frac{(z - (2\text{sign}(b)\beta - b))^2}{4(|b| - \beta)} + \beta & (\beta + c < |b| < 2\beta) \\ \frac{(z - \text{sign}(b)(\beta - c))^2}{4c} + \beta & (\beta - c < |b| < \beta + c) \\ \frac{(z - b)^2}{4(\beta - |b|)} + \beta & (0 \leq |b| < \beta - c) \end{cases}$$

が導出できる (導出過程は略)。 $h_n(z)$ ($n = 1, 2, 3$) はいずれも $f_n(z)$ と点 b で接し、 $f_n(z)$ より決して小さくなることのない 2 次関数の中で 2 次の係数が最小のものである。これにより適切な補助関数が設計でき、

1. \mathbf{s} の初期値設定
2. \mathbf{s} と \mathbf{B}, \mathbf{w} を収束するまで次のように交互に更新

$$\mathbf{B} \leftarrow \begin{pmatrix} b_1^{(n)}(\mathbf{s}) & & 0 \\ & \ddots & \\ 0 & & b_I^{(n)}(\mathbf{s}) \end{pmatrix}, \quad \mathbf{w} \leftarrow \begin{pmatrix} w_1^{(n)}(\mathbf{s}) \\ \vdots \\ w_I^{(n)}(\mathbf{s}) \end{pmatrix}$$

$$\mathbf{s} \leftarrow (\mathbf{A}^T \mathbf{B} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{B} (\mathbf{x} - \mathbf{w})$$

により、目的関数 $L_n(\mathbf{s})$ を最小化する回帰係数を得ることができる。行列 \mathbf{B} の対角要素 $b_i^{(n)}(\mathbf{s})$ およびベクトル \mathbf{w} の要素 $w_i^{(n)}(\mathbf{s})$ の更新則は、符号量モデルの選択に応じて

$$b_i^{(1)} = \frac{1}{2 \max(|z_i|, 2\beta)}, \quad w_i^{(1)} = 0$$

$$b_i^{(2)} = \frac{1}{2\sqrt{z_i^2 + \beta^2}}, \quad w_i^{(2)} = 0$$

$$b_i^{(3)} = \begin{cases} \frac{1}{2|z_i|} & (|z_i| \geq 2\beta) \\ \frac{1}{4(|z_i|-\beta)} & (\beta + c \leq |z_i| < 2\beta) \\ \frac{1}{4c} & (\beta - c \leq |z_i| < \beta + c) \\ \frac{1}{4(\beta-|z_i|)} & (0 \leq |z_i| < \beta - c) \end{cases}$$

$$w_i^{(3)} = \begin{cases} 0 & (|z_i| \geq 2\beta) \\ 2\text{sign}(z_i)\beta - z_i & (\beta + c \leq |z_i| < 2\beta) \\ \text{sign}(z_i)(\beta - c) & (\beta - c \leq |z_i| < \beta + c) \\ z_i & (0 \leq |z_i| < \beta - c) \end{cases}$$

と与えられる。ただし、 $z_i = x_i - (\mathbf{A}\mathbf{s})_i$ である。ここでは、 \mathbf{B} だけでなく \mathbf{w} も補助変数であることに注意が必要である。

Huber の M 推定における IRLS 法では目的関数 $L_3(\mathbf{s})$ の単調減少が保証される更新則を導くことができない。この点が、補助関数法により導出された提案法と IRLS 法の原理的な違いを明確に物語っている。

4 評価実験

提案法の効果を確認するため、音響信号を対象として圧縮性能の評価実験を行った。本実験の目的は、提案法の原理を実証することだけでなく、LPC に基づく可逆圧縮符号化において残差の符号量削減の余地がどの程度残されているのかを見極めることである。

評価実験用に、RWC 研究用音楽データベース [4] に含まれる、サンプリング周波数 44.1kHz、16 ビットで収録されたステレオの音楽ファイル (WAV 形式) 10 個 (RWC-MDB-P-2001 No.33~42) を実験データ (合計約 432MB) として用いた。LPC 分析フレーム長は 2048 サンプルとした。

ℓ_2 ノルムを規準とした LPC を Levinson-Durbin アルゴリズムにより行う従来の方式を従来法、従来法で得られた予測係数を初期値として、 $L_2(\mathbf{s})$ および $L_3(\mathbf{s})$ を規準とした LPC を提案する反復最適化アルゴリズムにより行う方式をそれぞれ提案法 1、提案法 2 と略称する。提案法 1 および提案法 2 の反復計算回数は 10 とした。なお、各方式ではいずれもチャンネル間相関に基づく圧縮は行っていない。また、LPC の予測次数 (式 (5) における J) は固定とした。

$L_2(\mathbf{s})$ は ℓ_2 ノルムに比べて Rice 符号量に則した目的関数となっているため、従来法に比べ、提案法 1 の圧縮率は向上するものと予想される。一方、 $L_3(\mathbf{s})$ は $L_2(\mathbf{s})$ に比べてより Rice 符号量に則した目的関数となっているため、提案法 1 に比べ、提案法 2 の圧縮率は若干向上するものと予想される。

従来法、提案法 1、提案法 2 それぞれについての、予測次数が 7, 15, 31 の場合の圧縮率をを図 3 に示す。圧縮率は、次のように算出した。

$$\text{圧縮率 (\%)} = \frac{\text{圧縮後のファイルサイズ}}{\text{元のファイルサイズ}} \times 100 \quad (16)$$

図 3 より、予想通り、提案法 2 による圧縮率が最も高く、次いで提案法 1 による圧縮率が高かった。以上

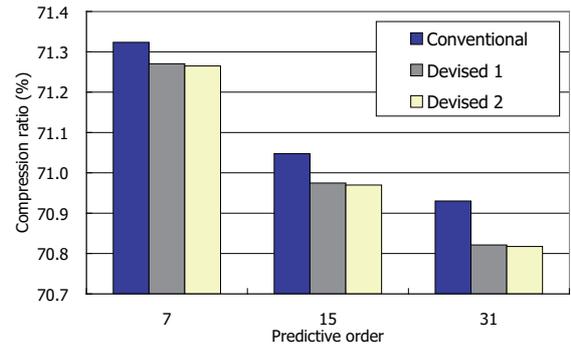


Fig. 3 ℓ_2 ノルム規準の線形予測符号化 (従来法) に対する、 $L_2(\mathbf{s})$ および $L_3(\mathbf{s})$ を規準とした線形予測符号化 (提案法 1&2) による圧縮率の比較

の実験結果から、提案法の有効性と、提案法により実現された残差符号量削減の度合いを確認することができた。

5 まとめ

本稿では、従来の LPC による可逆圧縮符号化は残差符号量を直接的に小さくするものではなかった点に着目し、残差符号量をできるだけ小さくする線形予測分析法の確立を目的とし、(1) 残差の符号化に利用される Rice 符号における振幅と符号量との関係をモデル化し、(2) Rice 符号量モデルに基づいて立てられる目的関数の最小解を効率的に求める反復アルゴリズム (補助関数法) を考案した。評価実験により、提案法の動作と提案法による圧縮性能を確認した。

提案法は、ある条件を満たす目的関数を規準とする場合には Huber の M 推定における IRLS 法と同一の反復アルゴリズムを与えることが明らかになった。しかしながら、IRLS 法では収束性が保証されない目的関数に対しても、提案法の枠組では収束性が保証された反復アルゴリズムを導けることを例示した。このことで、提案法は IRLS 法より幅広いクラスの目的関数を扱える方法論であることが示される。

参考文献

- [1] A. Charnes, W. W. Cooper, and R. O. Ferguson, "Optimal Estimation of Executive Compensation by Linear Programming," *Management Science*, Vol. 1, No. 2, pp. 138–151, 1955.
- [2] P. J. Huber, *Robust Statistics*, New York: Wiley, 1981.
- [3] V. Vapnik, *Statistical Learning Theory*, New York: Wiley, 1998.
- [4] 後藤, 橋口, 西村, 岡, "RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース," 情処論文, Vol. 45, No. 3, pp. 728–738, 2004.