マルチカーネル線形予測モデルによる音声分析

亀岡弘和, 大石康智, 持橋大地, ルルージョナトン (NTTコミュニケーション科学基礎研究所)

NTT

1. 序論

◆線形予測(Linear Prediction; LP)法による音声分析 [1]

- 声帯駆動信号を白色Gauss性雑音と仮定し，全極モデルによって表される声道伝達特性を観測信号から推定

$$\epsilon_i \longrightarrow \boxed{1/A(z)} \longrightarrow x_i = \sum_{p=1}^{P} a_p x_{i-p} + \epsilon_i$$

$$A(z) = 1 - a_1 z^{-1} \cdots - a_P z^{-P}$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2)\ (i = 1, \cdots, I)$$

問題：白色性の仮定は高い基本周波数の音声ほど適合しなくなる
(声道伝達特性の推定結果に影響を及ぼす可能性がある)

◆LP法の改良アプローチ

- 駆動音源信号がラプラス分布に従うと仮定したもの [2]
- 予め抽出した調波成分のスペクトルピークだけに対して全極スペクトルをフィッティングするもの [3]
- 声帯駆動信号を隠れマルコフモデルにより確率モデル化したもの [4]

駆動音源の周期性を陽に仮定し，声道伝達特性とともに声帯駆動信号の周期，周期性の有無についても推定できるLP法が実現できないか？

2. 従来のLP法 [1]

◆音声信号のモデル化

- P次の自己回帰過程：$x_i = \sum_{p=1}^{P} a_p x_{i-p} + \epsilon_i$

$$\boldsymbol{\epsilon} := (\epsilon_1, \cdots, \epsilon_I)^{\mathrm{T}},\quad \boldsymbol{x} := (x_1, \cdots, x_I)^{\mathrm{T}},\quad \boldsymbol{\Psi} := \begin{bmatrix} 1 & & & & O \\ -a_1 & \ddots & & & \\ \vdots & \ddots & \ddots & & \\ -a_P & & \ddots & \ddots & \\ & \ddots & & \ddots & \ddots \\ O & & -a_P & \cdots & -a_1\ 1 \end{bmatrix} \Rightarrow \boldsymbol{\Psi x} = \boldsymbol{\epsilon} \quad \cdots(1)$$

- 白色声帯駆動信号の仮定：$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$

(1)より $\boldsymbol{x}|\sigma^2, \underline{\boldsymbol{a}} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 (\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi})^{-1})$，$\boldsymbol{a} := (a_1, \cdots, a_P)^{\mathrm{T}}$

- パラメータの対数尤度：$|\boldsymbol{\Psi}| = 1$より

$$\log p(\boldsymbol{x}; \boldsymbol{a}) = -\frac{I}{2}\log 2\pi - \frac{1}{2\sigma^2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi}\boldsymbol{x} = -\frac{I}{2}\log 2\pi - \frac{1}{2\sigma^2}(\boldsymbol{x} - \boldsymbol{X}\boldsymbol{a})^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{X}\boldsymbol{a})$$

$$\boldsymbol{X} := \begin{bmatrix} 0 & \cdots & 0 \\ x_1 & \ddots & \vdots \\ \vdots & \ddots & 0 \\ \vdots & & x_1 \\ x_{I-1} & \cdots & x_{I-P} \end{bmatrix}$$

- 最尤の予測係数：

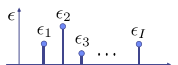
$$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\boldsymbol{a} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{x}$$

(正規方程式)

3. GPによる声帯駆動のモデル化

- 基底関数 $\phi_1(t), \cdots, \phi_H(t)$ の線形結合＋ノイズ：

$$\epsilon = \underline{f(t)} + \eta \qquad \sum_{h=1}^{H} \phi_h(t) w_h = \boldsymbol{\phi}(t)^{\mathrm{T}}\boldsymbol{w}$$



$$\boldsymbol{\epsilon} := (\epsilon_1, \cdots, \epsilon_I)^{\mathrm{T}},\quad \boldsymbol{\Phi} := (\boldsymbol{\phi}(t_1), \cdots, \boldsymbol{\phi}(t_I))^{\mathrm{T}},\quad \boldsymbol{w} := (w_1, \cdots, w_H)^{\mathrm{T}},\quad \boldsymbol{\eta} := (\eta_1, \cdots, \eta_I)^{\mathrm{T}} \Rightarrow \boldsymbol{\epsilon} = \boldsymbol{\Phi}\boldsymbol{w} + \boldsymbol{\eta} \quad \cdots(2),\ \boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \sigma_\eta^2 \boldsymbol{I})$$

- 結合係数 $\boldsymbol{w}$ の事前確率：$\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \sigma_w^2 \boldsymbol{I})$

正規分布する結合係数による基底関数の線形和で表される回帰モデルをガウシアンプロセス(GP)という

(2)より $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma_w^2 \underline{\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}} + \sigma_\eta^2 \boldsymbol{I})$，$=: \boldsymbol{K}$

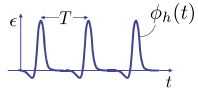
$\boldsymbol{K}$の$\{i,j\}$成分 $K_{i,j}$は，$k(t,t') := \boldsymbol{\phi}(t)^{\mathrm{T}}\boldsymbol{\phi}(t')$と定義されるカーネル関数を用いて $K_{i,j} = k(t_i, t_j)$で表される

- 白色雑音に帰着する基底関数：

$$\phi_h(t) := \begin{cases} 1 & (t = c_h) \\ 0 & (t \neq c_h) \end{cases} \longrightarrow \begin{matrix} \{t_1, \cdots, t_I\} \\ \subseteq \{c_1, \cdots, c_H\} \text{ならば} \end{matrix} \quad \boldsymbol{K} = \boldsymbol{I}$$

4. 周期カーネルの導入

- 有声音の場合，声帯駆動信号は周期性をもつことが多い
  → 周期 T の基底関数を用いて ϵ をモデル化したい



- 周期カーネルの例 [5]

$$k(t,t') := \exp\left\{-2\sin^2\left(\pi\frac{t-t'}{T}\right)\Big/\ell^2\right\}$$

  - 正定値性を満たす
  - 周期 T の基底関数を仮定したことに相当

- 本発表では以下のように周期カーネルを定義

$$\phi_h(t) := \sum_{n=1}^{N} \sin\left(2\pi n\frac{t-c_h}{T}\right) \longrightarrow k(t,t') := \sum_{h=1}^{H} \phi_h(t)\phi_h(t')$$

5. カーネル線形予測モデル

- 声帯駆動信号の周期が既知の場合のLPモデル
- (2)を(1)に代入：$\boldsymbol{\Psi x} = \underline{\boldsymbol{\Phi w}} + \underline{\boldsymbol{\eta}}$
  （$\boldsymbol{\Phi w}$：周期 T の雑音，$\boldsymbol{\eta}$：白色雑音）

$$\longrightarrow \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \sigma_w^2 \boldsymbol{\Psi}^{-1}\boldsymbol{K}\boldsymbol{\Psi} + \sigma_\eta^2 (\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi})^{-1})$$

6. マルチカーネル線形予測モデル

- 声帯駆動の周期は未知⟶カーネル自体を推定したい
- マルチカーネル学習の考え方を導入：
  - 複数の周期カーネルの線形結合でカーネル関数を定義

$$k(t,t') = \sum_{m=1}^{M} \theta_m k_m(t,t'),\quad \sum_{m=1}^{M} \theta_m = 1$$

  - 周期T_mの周期カーネル $k_m(t,t')$の優勢度 θ_mが推定すべき未知パラメータ

$$\boldsymbol{K} = \sum_{m=1}^{M} \theta_m \underline{\boldsymbol{K}_m}$$

（$k_m(t_i, t_j')$をi行j列の要素にもつ行列）

- $\boldsymbol{x}$の確率密度関数：

$$\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_w^2 \sum_{m=1}^{M} \theta_m \boldsymbol{\Psi}^{-1}\boldsymbol{K}_m\boldsymbol{\Psi}^{-\mathrm{T}} + \sigma_\eta^2 (\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi})^{-1}\right)$$

- モデルパラメータ $\Theta := \{\boldsymbol{a}, \theta, \sigma\}$の対数尤度：

$$\log p(\boldsymbol{x}; \Theta) = -\frac{1}{2}\left(I\log 2\pi + \log|\boldsymbol{\Sigma}| + \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Psi}\boldsymbol{x}\right)$$

$$\boldsymbol{\Sigma} = \sigma_w^2 \sum_{m=1}^{M} \theta_m \boldsymbol{K}_m + \sigma_\eta^2 \boldsymbol{I}$$

- $\theta := \{\theta_m\}_{m=1}^{M}$の推定を通して音声信号に含まれる優勢な基本周期を知ることができる
- $\sigma := \{\sigma_w, \sigma_\eta\}$ の推定を通して周期成分・白色成分の混在比を知ることができる

7. スパース正則化パラメータ学習

- θの事前分布：$p(\theta) = \prod_{m=1}^{M} \frac{\lambda}{2}\exp\{-\lambda|\theta_m|\}$ (ラプラス分布)

⟶ θ の分布をスパースに誘導する

- 最大事後確率パラメータの反復推定：

$$\boldsymbol{a} \leftarrow \underset{\boldsymbol{a}}{\mathrm{argmax}}\ \log p(\boldsymbol{x}; \Theta) + \log p(\theta) \quad \cdots(3)$$

$$\{\theta, \sigma\} \leftarrow \underset{\theta, \sigma}{\mathrm{argmax}}\ \log p(\boldsymbol{x}; \Theta) + \log p(\theta) \quad \cdots(4)$$

  - (3): $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{a} = \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{x}$
  - (4): 解析的に求まらないがEM (Expectation-Maximization)法で反復推定できる

完全データ $\boldsymbol{y} := (\boldsymbol{y}_1^{\mathrm{T}}, \cdots, \boldsymbol{y}_M^{\mathrm{T}}, \boldsymbol{y}_{M+1}^{\mathrm{T}})$

$\boldsymbol{y}_m \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_w^2 \theta_m \boldsymbol{\Psi}^{-1}\boldsymbol{K}_m\boldsymbol{\Psi}^{-\mathrm{T}}\right),\ m = 1, \cdots, M$

$\boldsymbol{y}_{M+1} \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_\eta^2 (\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi})^{-1}\right)$

$\boldsymbol{H} := [\boldsymbol{I}, \cdots, \boldsymbol{I}]$

$\boldsymbol{x} = \boldsymbol{H}\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$

$\boldsymbol{\Lambda} = \begin{bmatrix} \sigma_w^2\theta_1\boldsymbol{\Psi}^{-1}\boldsymbol{K}_1\boldsymbol{\Psi}^{-\mathrm{T}} & & O \\ & \ddots & \\ O & & \sigma_\eta^2(\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi})^{-1} \end{bmatrix}$

完全データ対数尤度：$\log p(\boldsymbol{y}; \Theta) \overset{c}{=} -\frac{1}{2}\left(\log|\boldsymbol{\Lambda}| + \boldsymbol{y}^{\mathrm{T}}\boldsymbol{\Lambda}^{-1}\boldsymbol{y}\right)$

$$Q(\Theta, \Theta') = \log p(\theta) - \frac{1}{2}\left(\log|\boldsymbol{\Lambda}| + \mathrm{tr}\left(\boldsymbol{\Lambda}^{-1}\underline{\mathbb{E}[\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}|\boldsymbol{x}; \Theta']}\right)\right) \longrightarrow \text{「Q関数」}$$

$$\overset{c}{=} -\sum_{m=1}^{M}\lambda|\theta_m| - \frac{I}{2}\log\left(\sigma_w^{2M}\sigma_\eta^2\theta_1\cdots\theta_M\right) - \frac{1}{2\sigma_w^2}\sum_{m=1}^{M}\frac{1}{\theta_m}\mathrm{tr}\left(\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{K}_m^{-1}\boldsymbol{\Psi}\boldsymbol{R}_m\right) - \frac{1}{2\sigma_\eta^2}\mathrm{tr}\left(\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi}\boldsymbol{R}_{M+1}\right)$$

$\underline{\mathbb{E}[\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}|\boldsymbol{x}; \Theta']} = \boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^{\mathrm{T}})^{-1}\boldsymbol{H}\boldsymbol{\Lambda} + \boldsymbol{\Lambda}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^{\mathrm{T}})^{-1}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{H}^{\mathrm{T}})^{-\mathrm{T}}\boldsymbol{H}\boldsymbol{\Lambda}^{\mathrm{T}}$

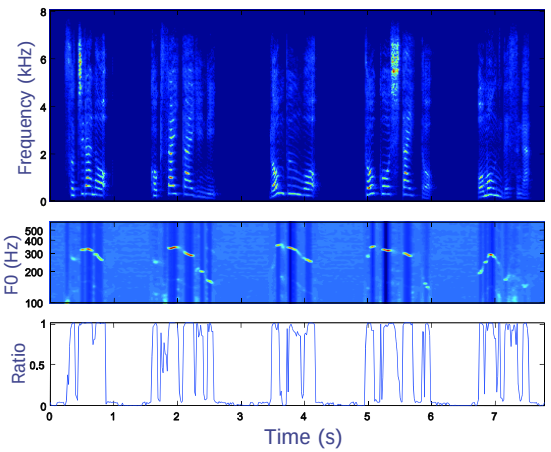
$\begin{bmatrix} \boldsymbol{R}_1 & & * \\ & \ddots & \\ * & & \boldsymbol{R}_M \\ & & & \boldsymbol{R}_{M+1} \end{bmatrix}$

⟶ Mステップ：$\theta_m \leftarrow 2\sigma_w^2\lambda\theta_m^2 + I\sigma_w^2\theta_m - \mathrm{tr}\left(\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{K}_m^{-1}\boldsymbol{\Psi}\boldsymbol{R}_m\right) = 0$ の正の根

$$\sigma_w^2 \leftarrow \frac{1}{IM}\sum_{m=1}^{M}\frac{1}{\theta_m}\mathrm{tr}\left(\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{K}_m^{-1}\boldsymbol{\Psi}\boldsymbol{R}_m\right) \quad \sigma_\eta^2 \leftarrow \frac{1}{I}\mathrm{tr}\left(\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi}\boldsymbol{R}_{M+1}\right)$$

8. 動作実験

- 分析条件
  - 音声データ：ATR音声データベースBセットの女性話者音声
  - 標本化周波数：16kHz
  - フレーム分割：64ms長のHanning窓，シフト幅32ms
- 提案法の条件
  - 反復回数：600
  - 予測次数 P ：13
  - 周期カーネルの設定：100Hz～600Hzを6cent間隔で離散化した計518個の基本周波数値 $1/T_m$ をもとに構成
- 実験結果
  - （上）観測信号のスペクトログラム
  - （中）各時刻における θ の分布
  - （下）各時刻における周期成分と白色成分のエネルギー比



参考文献

[1] F. Itakura and S. Saito, "Analysis synthesis telephony based upon the maximum likelihood method," In Proc. 6th Int'l Cong. Acoust. (ICA'68), C-5-5, pp. C17–20, 1968.
[2] C.-H. Lee, "On robust linear prediction of speech," IEEE Trans. Acoust., Speech & Signal Process., **36**(5), pp. 642–650, 1988.
[3] A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling," IEEE trans. Signal Process., **39**(2), pp. 411–423, 1991.
[4] 佐宗, 田中, "HMMによる音源のモデリングと高基本周波数に頑健な声道特性抽出," 信学論D-II, **J84-D-II**(9), 1960–1969, 2001.
[5] C.E. Rasmussen and C.K.I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, Mass, USA, 2006.