

音響信号のグループ構造を自動適応させるスパース時間周波数解析*

☆ 新井慶大, 山田宏樹, 矢田部浩平 (農工大)

1 まえがき

音響信号処理において時間周波数 (T-F) 解析は必要不可欠である. T-F 解析では離散ガボール変換 (DGT) が広く用いられるが, これにより得られた T-F 表現は窓の影響で広がりを持ち, 解析の性能に影響を及ぼす. そのため, l_1 最小化をはじめとするスパース T-F 表現を得るための手法が提案されている [1]. しかし, 過剰な局在化により元信号の構造がわかりにくくなるという欠点がある. 構造を保つには既知のグループに対しスパース性を誘導する手法が考えられるが, 信号のグループ構造を事前を知ることは困難である. そこで本稿では, 信号のグループ構造を自動適応させるスパース T-F 解析手法を提案する.

2 T-F 表現のスパース最適化

T-F 表現にはさまざまな種類があるが, 簡単のため本稿では DGT により得られる T-F 表現のみを考える. 音響信号 $\mathbf{d} \in \mathbb{R}^L$ の DGT は, 窓関数 $\mathbf{w} \in \mathbb{R}^L$ を用いて次式で定義される.

$$x_{m+nM} = \sum_{l=0}^{L-1} d_l \overline{w_{l-an}} e^{-2\pi i m l / M} \quad (1)$$

ただし, $a \in \mathbb{N}$ は窓のシフト幅, $N, M \in \mathbb{N}$ はそれぞれ時間フレーム数と周波数ビン数, $n = 0, \dots, N-1$ と $m = 0, \dots, M-1$ はそれぞれ時間と周波数のインデックス, $i = \sqrt{-1}$ である. 式 (1) は行列 $\mathbf{G}_w \in \mathbb{C}^{MN \times L}$ を用いて $\mathbf{x} = \mathbf{G}_w \mathbf{d}$ と表せる. ただし, $(\mathbf{G}_w)_{m+nM, l} = \overline{w_{l-an}} e^{-2\pi i m l / M}$ である. $\mathbf{G}_w^H \mathbf{G}_w$ が正則ならば, DGT の逆変換は \mathbf{w} の標準双対窓 $\boldsymbol{\gamma}^* = (\mathbf{G}_w^H \mathbf{G}_w)^{-1} \mathbf{w}$ を用いて $\mathbf{d} = \mathbf{G}_{\boldsymbol{\gamma}^*}^H \mathbf{x}$ と表される.

DGT による T-F 表現は窓の影響で広がりを持ち, 解析の性能に影響を及ぼすため, 可能な限り広がりのない表現を得たい. そこで, スパースな T-F 表現を得るための手法が研究されている. 例えば [1] では, 信号の完全再構成性を制約とし, l_1 ノルムをペナルティ関数に用いた凸最適化問題

$$\min_{\mathbf{x} \in \mathbb{C}^{MN}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{G}_{\boldsymbol{\gamma}^*}^H \mathbf{x} = \mathbf{d} \quad (2)$$

を解くことでスパースな T-F 表現を得る. また, l_1 ノルムの代わりに混合 $l_{2,1}$ ノルムを用いると, 既知のグループに対しスパースな T-F 表現が得られる [2].

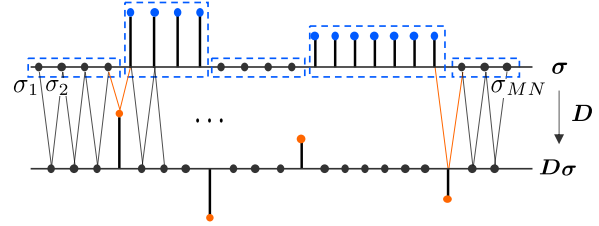


図-1 σ がグループ構造を持つ一例と, そのときの $D\sigma$ の模式図. 同じグループ内の要素は破線で囲んだ. $D\sigma$ はグループの境界でのみ大きな値を示す.

しかし, l_1 最小化では過剰な局在化により元信号の構造がわかりにくくなってしまふ. 混合 $l_{2,1}$ ノルムを用いる手法では構造を保つことはできるが, 信号のグループを事前を知る事が難しいという課題がある.

3 提案手法

本稿では, 音響信号のグループ構造を自動適応させるスパース T-F 解析手法を提案する. 上記の課題を解決するため, 事前にグループ構造を知らずとも信号のグループスパース性を誘導することが可能なペナルティ関数を考える.

3.1 重みを持つ構造を誘導するペナルティ関数

2変数の組 $(\mathbf{x}, \boldsymbol{\sigma}) \in \mathbb{C}^{MN} \times \mathbb{R}^{MN}$ に対し, 凸関数 φ を以下のように定義する.

$$\varphi(\mathbf{x}, \boldsymbol{\sigma}) = \sum_{k=1}^{MN} \phi(x_k, \sigma_k) \quad (3)$$

ただし, ϕ は $|\cdot|^2/2 + 1/2$ の perspective 関数 [3] で,

$$\phi(x_k, \sigma_k) = \begin{cases} \frac{|x_k|^2}{2\sigma_k} + \frac{\sigma_k}{2} & (\sigma_k > 0) \\ 0 & (x_k = 0 \text{ and } \sigma_k = 0) \\ \infty & (\text{otherwise}) \end{cases} \quad (4)$$

で定義される. $\boldsymbol{\sigma} > \mathbf{0}$ のとき $2\varphi = \mathbf{x}^H \boldsymbol{\Sigma}^{-1} \mathbf{x} + \|\boldsymbol{\sigma}\|_1$ ($\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\sigma})$) と表せる. 第2項から φ は $\boldsymbol{\sigma}$ のスパース性を誘導する. さらに, 第1項は \mathbf{x} に関して適応的な重み $\boldsymbol{\sigma}$ を持つ重み付き l_2 ノルムの2乗と解釈できるため, \mathbf{x} は $\boldsymbol{\sigma}$ の構造を持つように誘導される. したがって, 重み $\boldsymbol{\sigma}$ に適切な正則化を施すことで, \mathbf{x} に所望の構造を付与できると考えられる.

3.2 グループ構造を自動適応させる正則化

φ が \mathbf{x} のグループスパース性を誘導するには, 重み $\boldsymbol{\sigma}$ がグループ構造を持てばよい. $\boldsymbol{\sigma}$ にグループ構造

*Sparse time-frequency analysis by automatically adapting group structure of acoustic signals. By Keidai ARAI, Koki YAMADA, and Kohei YATABE (Tokyo University of Agriculture and Technology)

Algorithm 1 提案手法のアルゴリズム

Input: $\tau > 0, \mu > 0, \rho^{[i]} \in (0, 2) (i = 0, 1, 2, \dots),$
 $\mathbf{x}^{[0]} \in \mathbb{C}^{MN}, \boldsymbol{\sigma}^{[0]} \in \mathbb{R}^{MN}, \mathbf{u}^{[0]} \in \mathbb{C}^{MN}, \mathbf{v}^{[0]} \in \mathbb{R}^{MN-1}$

for $i = 0, 1, 2, \dots$ **do**
 $(\mathbf{x}^{[i+\frac{1}{2}]}, \boldsymbol{\sigma}^{[i+\frac{1}{2}]}) = \text{prox}_{\tau\varphi}(\mathbf{x}^{[i]} - \tau\mathbf{u}^{[i]}, \boldsymbol{\sigma}^{[i]} - \tau\mathbf{D}^T\mathbf{v}^{[i]})$
 $\tilde{\mathbf{u}}^{[i+\frac{1}{2}]} = \mathbf{u}^{[i]} + \mu(2\mathbf{x}^{[i+\frac{1}{2}]} - \mathbf{x}^{[i]})$
 $\mathbf{u}^{[i+\frac{1}{2}]} = \tilde{\mathbf{u}}^{[i+\frac{1}{2}]} - \mu P_C(\tilde{\mathbf{u}}^{[i+\frac{1}{2}]}/\mu)$
 $\tilde{\mathbf{v}}^{[i+\frac{1}{2}]} = \mathbf{v}^{[i]} + \mu\mathbf{D}(2\boldsymbol{\sigma}^{[i+\frac{1}{2}]} - \boldsymbol{\sigma}^{[i]})$
 $\mathbf{v}^{[i+\frac{1}{2}]} = \tilde{\mathbf{v}}^{[i+\frac{1}{2}]} - \mu \text{prox}_{(\lambda/\mu)\|\cdot\|_{2,1}}(\tilde{\mathbf{v}}^{[i+\frac{1}{2}]}/\mu)$
 $\mathbf{x}^{[i+1]} = \mathbf{x}^{[i]} + \rho^{[i]}(\mathbf{x}^{[i+\frac{1}{2}]} - \mathbf{x}^{[i]})$
 $\boldsymbol{\sigma}^{[i+1]} = \boldsymbol{\sigma}^{[i]} + \rho^{[i]}(\boldsymbol{\sigma}^{[i+\frac{1}{2}]} - \boldsymbol{\sigma}^{[i]})$
 $\mathbf{u}^{[i+1]} = \mathbf{u}^{[i]} + \rho^{[i]}(\mathbf{u}^{[i+\frac{1}{2}]} - \mathbf{u}^{[i]})$
 $\mathbf{v}^{[i+1]} = \mathbf{v}^{[i]} + \rho^{[i]}(\mathbf{v}^{[i+\frac{1}{2}]} - \mathbf{v}^{[i]})$
end for

を持たせるような正則化を導くために、図-1のような状況を考える。 $\boldsymbol{\sigma}$ を $M \times N$ の配列とみたときに、各要素の勾配を近似した差分作用素を \mathbf{D} とすると、図-1より $\mathbf{D}\boldsymbol{\sigma}$ はスパースになる。このとき、 $\mathbf{D}\boldsymbol{\sigma}$ のスパース性を誘導すると $\boldsymbol{\sigma}$ はグループ構造を持つ [2]。そこで、 $\boldsymbol{\sigma}$ の全変動 $\|\mathbf{D}\boldsymbol{\sigma}\|_{2,1}$ を導入し、 $\boldsymbol{\sigma}$ がグループ構造を持つように誘導する。

$\boldsymbol{\sigma}$ の全変動を φ と組み合わせることで、 $\boldsymbol{\sigma}$ を通して \mathbf{x} のグループスパース性を誘導するペナルティを実現する。このペナルティを式 (2) の ℓ_1 ノルムの代わりに用い、音響信号のグループ構造を自動適応させるスパースな T-F 表現の推定を、正則化パラメータ $\lambda > 0$ を用いて次のように定式化する。

$$\min_{(\mathbf{x}, \boldsymbol{\sigma})} \varphi(\mathbf{x}, \boldsymbol{\sigma}) + \lambda \|\mathbf{D}\boldsymbol{\sigma}\|_{2,1} \quad \text{s.t.} \quad \mathbf{G}_{\gamma^*}^H \mathbf{x} = \mathbf{d} \quad (5)$$

これを主双対分離法で解くと **Algorithm 1** が得られる。ただし、 $\text{prox}_{\tau\varphi}(\mathbf{x}, \boldsymbol{\sigma}) = (\text{prox}_{\tau\varphi}(x_k, \sigma_k))_{k=1}^{MN}$ であり、各要素の $\text{prox}_{\tau\varphi}$ は [3, Example 2.4] から求められる¹。また、 $\mathcal{C} = \{\mathbf{x} \in \mathbb{C}^{MN} \mid \mathbf{G}_{\gamma^*}^H \mathbf{x} = \mathbf{d}\}$ への射影 P_C は、DGT のフレーム作用素の性質を用いて

$$\begin{aligned} P_C(\mathbf{x}) &= \mathbf{x} - \mathbf{G}_{\gamma^*}(\mathbf{G}_{\gamma^*}^H \mathbf{G}_{\gamma^*})^{-1}(\mathbf{G}_{\gamma^*}^H \mathbf{x} - \mathbf{d}) \\ &= \mathbf{x} - \mathbf{G}_w(\mathbf{G}_{\gamma^*}^H \mathbf{x} - \mathbf{d}) \end{aligned} \quad (6)$$

のように計算できる。

$${}^1\text{prox}_{\tau\varphi}(x_k, \sigma_k) = \begin{cases} (0, 0) & (2\tau\sigma_k + |x_k|^2 \leq \tau^2) \\ (0, \sigma_k - \frac{\tau}{2}) & (x_k = 0 \text{ and } 2\sigma_k > \tau) \\ (x_k - \tau s \frac{x_k}{|x_k|}, \sigma_k + \tau \frac{s^2 - 1}{2}) & (\text{otherwise}) \end{cases}$$

ただし、 $s > 0$ は $s^3 + ((2/\tau)\sigma_k + 1)s - (2/\tau)|x_k| = 0$ の正の根で、 $p = (2/\tau)\sigma_k + 1, q = -(2/\tau)|x_k|, r = -q^2/4 - p^3/27$ とするとカルダノの公式から次式で求められる。

$$s = \begin{cases} \sqrt[3]{-\frac{q}{2} + \sqrt{-r}} + \sqrt[3]{-\frac{q}{2} - \sqrt{-r}} & (r < 0) \\ 2\sqrt[3]{-\frac{q}{2}} & (r = 0) \\ 2\sqrt[3]{\sqrt{\frac{q^2}{4} + r} \cos\left(\frac{\arctan(-2\sqrt{r}/q)}{3}\right)} & (r > 0) \end{cases}$$

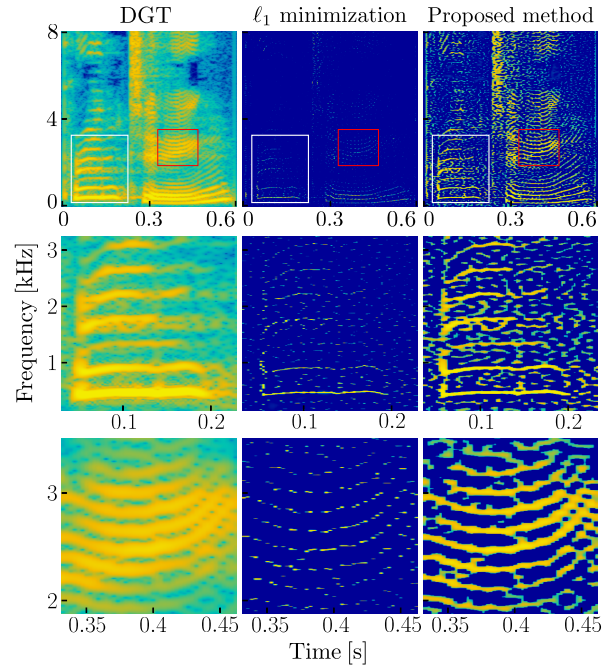


図-2 得られた音声信号の T-F 表現. 各列は左から順に DGT, ℓ_1 最小化 (式 (2)), 提案手法 (式 (5), $\lambda = 1$) により得られたもので、中段と下段はそれぞれ白枠と赤枠の拡大図である。色の範囲は 100 dB とした。

4 実験

DGT, ℓ_1 最小化 (式 (2)), 提案手法 (式 (5)) の 3 つの手法により得られた音声信号 \mathbf{d} の T-F 表現を比較した。窓関数 \mathbf{w} として窓長が 2^9 のハン窓を用い、シフト幅 $a = 2^6$, 周波数ビン数 $M = 2^{12}$ とした。Algorithm 1 の反復回数は 5000 回とし、 $\tau = 1/2, \mu = 1/5, \rho^{[i]} = 1.99, \mathbf{x}^{[0]} = \mathbf{G}_w \mathbf{d}, \boldsymbol{\sigma}^{[0]} = |\mathbf{G}_w \mathbf{d}|, \mathbf{u}^{[0]} = 0, \mathbf{v}^{[0]} = 0$ とした。

実験結果を図-2 に示す。 ℓ_1 最小化では DGT よりもスパースな T-F 表現が得られているが、過剰な局在化により元信号のグループ構造がわかりにくい。一方、提案手法では元信号のグループ構造を保ちつつスパースな T-F 表現が得られている。

5 むすび

本稿では、音響信号のグループ構造を自動適応させるスパース T-F 解析手法を提案した。今後は定式化を拡張し、さまざまな構造を T-F 表現に付与する手法を検討する。

参考文献

- [1] G. E. Pfander and H. Rauhut, “Sparsity in time-frequency representations,” *J. Fourier Anal. Appl.*, vol. 16, no. 2, pp. 233–260, 2010.
- [2] H. Kuroda and D. Kitahara, “Block-sparse recovery with optimal block partition,” *IEEE Trans. Signal Process.*, vol. 70, pp. 1506–1520, 2022.
- [3] P. L. Combettes and C. L. Müller, “Perspective maximum likelihood type estimation via proximal decomposition,” *Electron. J. Stat.*, vol. 14, no. 1, pp. 207–238, 2020.