

## 特異値分解を用いた音声の雑音除去 Noise reduction using SVD

指導教授 保谷哲也

Tetsuya Hoya

M3015 吉井 達也<sup>1</sup>

Tatsuya Yoshii

### 1. 概要

スマートフォン等が普及している昨今、音声検索等で音声認識の需要が上昇している。本論文では、音声の雑音除去を行った実験と結果について報告する。雑音除去には様々な方法があり、その一つとして、特異値分解(SVD)を用いる方法がある。本論文では、主成分分析[1]の行列の成分数と並べ替えを行い、SVDの特異値の個数を減らすことを利用して雑音除去を行うことについて述べる。

### 2. 特異値分解(SVD)

特異値分解は、18 世紀 70 年代に、Bletrami と Jordan によって、まず実正方行列について提案された。後に Autonne によって複素正方行列に、さらに Eckart と Young によって一般的な長方形行列の拡張がなされた。任意の行列  $A$  は特異値分解によって以下のように表すことができる。

行列  $A$  に対し、以下を成立させる直行行列  $U \in R^{m \times m}$  と  $V \in R^{n \times n}$  が存在する。

$$A = U \Sigma V^T \quad (1)$$

但し、

$$\Sigma = \begin{pmatrix} S & 0 \\ 0 & 0 \end{pmatrix} \quad (2)$$

であり、 $S = \text{diag}(\sigma_1, \dots, \sigma_r)$  の対角要素が降順で並ぶ。ここで、

$$\sigma_1 \geq \sigma_2 \geq \sigma_r > 0, \quad r = \text{rank}A$$

であり、 $\sigma_i (i = 1, \dots, r)$  は  $A^T A \geq 0$  より、 $A^T A$  の固有値は、 $n < m$  のとき、 $\sigma(A^T A) = \{\sigma_1^2, \dots, \sigma_n^2\} \subseteq [0, +\infty)$  となる。但し、 $\sigma_1 \geq \dots \geq \sigma_r > 0 = \sigma_{r+1} = \dots = \sigma_n$  である。ここで、 $\sigma_i$  を  $A$  の特異値という[2]。また特異値分解では、 $\Sigma \in R^{m \times n}$  の  $|m - n|$  の量の行または列の成分がすべて 0 になる。本研究では、この性質を利用し、メモリを節約するために成分がすべて 0 の行または列を省いた。

#### 2.1 行列の並べ替え

本研究では、音声を  $12000 \times 2$  (サンプル数  $\times$  チャンネル数) の行列とした。よって、特異値の数は行列の rank と同じ 2 個になる。特異値が少ない場合、後述する閾値を変化させて行う雑音除去の結果が完全に一致する場

合があるために行列の成分を並べ替えた。このような並べ替えを行うことにより、行列の rank を多くすることができる。さらに、並び替えた行列のサイズ別で雑音除去を試みる。以上の方法による雑音除去の実験結果について学術講演会当日に発表する予定である。

#### 2.2 主成分分析の成分数と SVD の特異値

画像の圧縮等に用いられる手法の一つが主成分分析である。主成分分析は行列の共分散行列の固有値を求め、その固有値の数を減らすことで行われる。これにより、行列の情報の次元数が減るので、データの低次元化がなされる。この時に用いられる固有値は、共分散行列の固有値なので非負定値であり、固有値が大きい順に並べられる。さらに、大きいものから順に一定数の固有値を用い、その固有値に対応する固有ベクトルと元の行列の内積演算により画像の圧縮等が行われる。ここで、用いられた固有値の個数は成分数と呼ばれる[1]。本研究では、以上述べたような主成分分析の成分数を減らす手法により特異値を減らした。また、閾値  $t$  を設定し、特異値を減らす条件を次のように決定した:

$$\text{sum} = \sigma_1 + \dots + \sigma_r \quad (3)$$

ここで、

$$\sigma_1 + \dots + \sigma_n \geq \text{sum} * t \quad (4)$$

である。次に、最小の  $n$  を算出し、 $n+1$  番目以降の特異値を 0 とした。本研究では、 $t$  の設定値は 0.5~0.9 の 0.1 刻みの値とした。

#### 3. 雑音除去実験

雑音除去実験を行うに当たり、音声データに関しては、本論文の執筆者が 3 つの短文 (日本語) 「おはようございます」「こんにちは」「こんばんは」を 3 回ずつ録音した。録音の条件は、44.1kHz、量子化 16 ビット、ステレオで 8kHz にダウン・サンプリングを施し、サンプル数はどれも固定で 12000 とした。次に、雑音音声に関しては、MATLAB の randn()関数を用いて  $12000 \times 2$  の正規乱数を生成したものを、各音声データに加算することにより得られたものを用いた [3]。本研究では特異値分解で雑音除去を行う過程で前述の行列の並び替えを行い、その並び変えた行列を特異値分解した。並び替えは行列の列が 4~152 の 4 刻みで設定した。列を少ないままに並び替えた理由は、特異値分解の性質

1 : 日大理工・院・数学

上, 行 $\geq$ 列の場合が行 $<$ 列の場合より演算が基本的に早いからである. 分解後, (4)式と照らし合わせ,  $\Sigma$ の  $S$  を $S_n$ として (1)式に代入する:

$$S_n = \text{diag}(\sigma_1, \dots, \sigma_n)$$

最後に, 並び替えた行列を元のサイズに戻すことで, 雑音除去を行った.

#### 4. 実験結果

実験結果の一部を図1~図7に示す. 図の入力音声は「こんにちは」の文を行列化した際に 1000 行 24 列に並び替えたものである. なお, 本論文では, 左右のチャンネルでほぼ同じ結果が得られたので, 左チャンネルのみを掲載する. 図3と図4に示されるように, 閾値 0.6 以下では, 音声部分の振幅まで極端に小さくなり, ほぼ無音になっている. また, 図3~5で示されるように, 横軸のサンプル数が 1000 サンプル単位でまとまって雑音除去がされ, その振幅が小さな値になっていることが観察された.

#### 5. 終わりに

本研究では, 特異値を減らすことによって雑音除去が

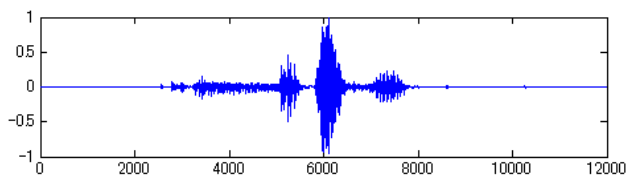


図1 入力音声

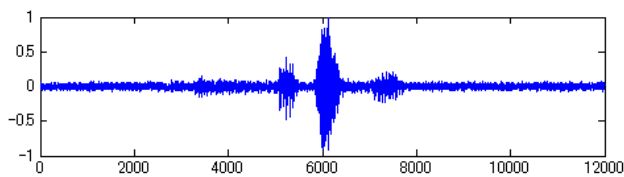


図2 雑音付加音声

可能となることが確認された. しかし, 閾値を低く設定すると音声部分の劣化が顕著になる. 学術講演会ではこの雑音除去の結果を数値的に検証した結果について発表する予定である. 今後の課題として, 並び替えを行ったときに行単位で雑音除去に差が出ることや, 音声の劣化を抑えつつ雑音除去が行える目安となる閾値について検討することなどが挙げられる. より具体的には, [3]で提案された手法におけるような, 小さいサイズの行列ごとに雑音除去を行い, 元の行列に戻し, 雑音除去を行う部分を少しずつするという手法により精度の高い雑音除去が期待される. 閾値の設定以外にも, 閾値を設けずに特異値を算出した段階で, 特異値が小さいものの比率を 0 とするのではなく, 比率を微量残し, 特異値の大きいものの比率を少し下げることなどが考えられる.

#### 6. 参考文献

- [1]羽石秀昭<<http://www.cfme.chiba-u.jp/~haneishi/class/2009/20091118.pdf>>(2013年9月現在)
- [2]張賢達, “信号処理のための線形代数”, 森北出版株式会社, 2008年1月, pp152-153.
- [3] T. Hoya, T. Tanaka, A. Cichocki, T. Murakami, G. hori, and J.A. Chambers, ”Stereophonic Noise Reduction Using a Combined Sliding Subspace Projection and Adaptive Signal Enhancement”, in IEEE Trans. Speech and Audio Processing, vol.13, no.3, May 2005, pp.309-320.

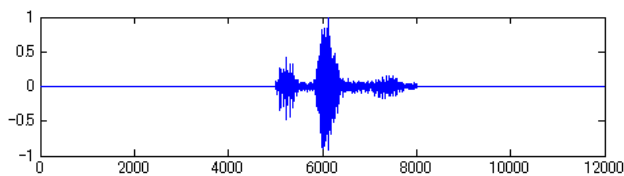


図4 雑音除去音声 t=0.6

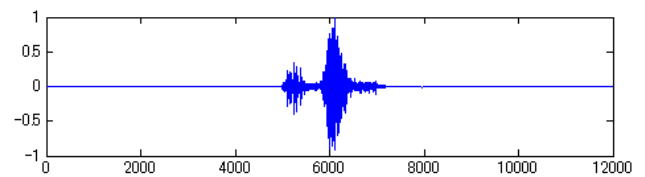


図3 雑音除去音声 t=0.5

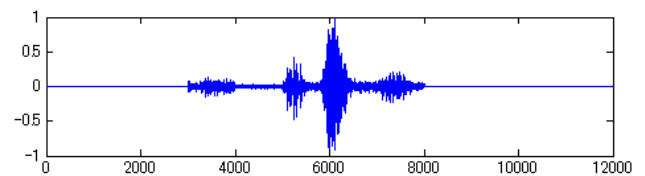


図5 雑音除去音声 t=0.7

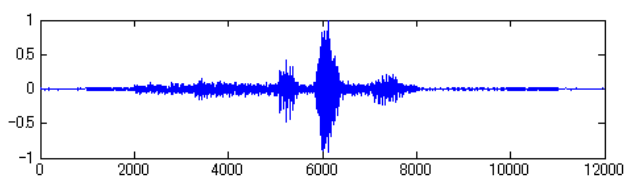


図6 雑音除去音声 t=0.8

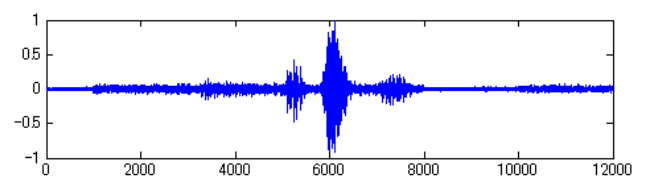


図7 雑音除去音声 t=0.9