LETTER

Multiple Matrix Rank Minimization Approach to Audio Declipping

Ryohei SASAKI[†a)], *Nonmember*, Katsumi KONISHI[††], Tomohiro TAKAHASHI[†], *and* Toshihiro FURUKAWA[†], *Members*

SUMMARY This letter deals with an audio declipping problem and proposes a multiple matrix rank minimization approach. We assume that short-time audio signals satisfy the autoregressive (AR) model and formulate the declipping problem as a multiple matrix rank minimization problem. To solve this problem, an iterative algorithm is provided based on the iterative partial matrix shrinkage (IPMS) algorithm. Numerical examples show its efficiency.

key words: *audio declipping, signal resortoration, switched AR model, sparse optimization, compressed sensing*

1. Introduction

Distortion of the speech and music signals is caused by various factors, for example, impulse noise represented by a click sound, clipping noise caused by thresholding and truncating a signal whose amplitude is over the maximum quantization width, packet loss in the IP telephone and so on. This letter deals with an audio declipping problem that is a signal degradation process in which an undistorted audio waveform is truncated whenever the audio signal exceeds the maximum input range of a digital acquisition system. This distortion seriously deteriorates the sound quality.

Several approaches have been proposed for audio inpainting and declippng, and they are roughly classified into three approaches, 1) patch based approach, 2) dictionary based approach, and 3) autoregressive (AR) model based approach. The patch based approach was proposed in [1], where missing signals are divided into some blocks and substituted by neighborhood known signals. This approach recovers signals for audio inpainting, however, does not work for audio declipping. In the audio declipping problem, signals are declipped in all intervals, and hence there is no proper reference signal to substitute. In [2], [3], dictionary based declipping algorithms were provided. Though these algorithms work well, their performances depend on dictionaries given in advance such as the discrete cosine transform dictionary and a Gabor dictionary. The AR model based approach is proposed in [4], [6], [7]. In this approach, audio signals are assumed to be modeled by the AR model and restored based on the model. Because the performance of this approach depends on the estimated model order, matrix rank minimization algorithms for audio declipping problem have been proposed in [6] and [7] based on the fact that the model order is equal to the rank of Hankel matrix generated from audio signals. Although it is known that audio signals are not steady-state in practice and hardly modeled by time-invariant AR model, they can be approximated by the AR model in a short time, and therefore these algorithms divide audio signals into multiple short time frames and restore signals in each frame. However, they achieve bad performance when the property of audio signals changes in a frame, which contradicts a single AR model assumption. Their quality depends on a frame length, and the best frame length is not constant but time-variant.

In order to improve the quality of matrix rank minimization based audio declipping algorithms, this letter takes a multiple AR model approach, where audio signals are assumed to be modeled by switched models consisting of multiple AR models. This model enables us to restore audio signals well because the model provides the best length of subframes even when the property of signals changes in each short time frame. Based on this model, the audio declipping problem is formulated as a multiple matrix rank minimization problem, and a new audio declipping algorithm is proposed by modifying the iterative partial matrix shrinkage (IPMS) algorithm [9]. Numerical examples show that the proposed algorithm has a good performance for mixed speech signals.

2. Problem Formulation

This section introduces an audio declipping problem and formulates it as a multiple matrix rank minimization problem.

Let us assume undistorted signals $\{x_t\}_{t=1}^{M+N-1}$ are provided by the following AR model,

$$x_t = \sum_{k=1}^{r} a_k x_{t-k} + v_t, \tag{1}$$

where a_k and v_t denote the AR coefficient and noise. Then consider the audio declipping problem which is a problem of restoring undistorted signals x_t from the observed signals y_t given as follows,

Manuscript received June 9, 2017.
Manuscript revised October 31, 2017.
Manuscript publicized December 6, 2017.

[†]The authors are with the Graduate School of Engineering, Tokyo University of Science, Tokyo, 125–8585 Japan.
[††]The author is with the Department of Computer Science, Kogakuin University, Tokyo, 163–8677 Japan.
a) E-mail: 4417703@ed.tus.ac.jp

DOI: 10.1587/transinf.2017EDL8129

$$y_t = g_C(x_t) = \begin{cases} C & \text{if} \quad x_t > C \\ -C & \text{if} \quad x_t < -C \\ x_t & \text{else} \end{cases}, \tag{2}$$

where C is a given constant called a clipping gain. In this problem, the AR coefficients a_k and the model order r are unknown. To formulate the audio declipping problem, we define Hankel matrix $X \in \mathbb{R}^{M\times N}$ of $\{x_t\}_{t=1}^{M+N-1}$ by

$$X_{i,j} = x_{i+j-1}, \tag{3}$$

where $M > N$, and $X_{i,j}$ denotes the (i, j) entry of X. If there are no noise, it holds that $\mathbf{rank}(X) = r$ since x_t satisfies (1). Based on the low-rankness of X, we formulate the audio declipping problem as the following matrix rank minimization problem,

$$\begin{array}{ll} \text{Minimize} & \mathbf{rank}(X) \\ \text{subject to} & X \in \mathcal{H} \cap \mathcal{I} \subset \mathbb{R}^{M\times N} \end{array}, \tag{4}$$

where $\mathcal{H} \subset \mathbb{R}^{M\times N}$ denotes the set of Hankel matrices defined in (3), and $\mathcal{I}$ denotes the set of matrices defined by

$$\mathcal{I} = \left\{ X \in \mathbb{R}^{M\times N} \middle\| \begin{array}{ll} X_{i,j} \geq C & \text{for } (i,j) \in \Pi^+ \\ X_{i,j} \leq -C & \text{for } (i,j) \in \Pi^- \\ X_{i,j} = Y_{i,j} & \text{for } (i,j) \notin \Pi^+ \cup \Pi^- \end{array} \right\}. \tag{5}$$

Π^+ and Π^- denote the index sets of positive and negative clipped signals in X corresponding to the clipping rules $x_t > C$ and $x_t < -C$, respectively. Because X is not exactly low-rank matrix even if x_t satisfies (1) due to noise and because the matrix rank minimization problem is NP hard in general, the objective function of (4) is often relaxed by the nuclear norm of X, and we can restore the signals by solving the relaxed problem if (1) holds.

While the assumption described by (1) is tight, the audio signals in a short time frame can be approximated by the AR model, and therefore some low-rank approaches restore clipped signals by dividing signals into multiple short time frames and declipping each frames [6], [7]. However, the quality of declipped signals in these approaches depends on a frame length, and the best frame length is not constant but time-variant in the assumption that the signals are modeled by a single AR model in each frame. In order to achieve high quality of signal declipping, this letter assumes that the audio signals in each frame can be described by multiple AR models and formulates a signal declipping problems as the following multiple matrix rank minimization problem,

$$\begin{array}{ll} \text{Minimize} & \sum_{i=1}^{L} \mathbf{rank}(D_i X) \\ \text{subject to} & \sum_{i=1}^{L} D_i = I,\ D_i \in \mathcal{D}, X \in \mathcal{H} \cap \mathcal{I}, \end{array} \tag{6}$$

where L is a given constant, $\mathcal{D}$ denotes a set of diagonal matrices whose elements are 0 or 1, and I denotes the identity matrix with a proper size. This problem recovers X by dividing its row vectors into L groups and minimizing a rank of each matrix. If all entries of X are known, this problem is equal to the generalized principle component analysis (GPCA) [8].

3. Main Results

This section provides an algorithm for the multiple matrix rank minimization problem and applies it to the audio declipping. Prior to the discussion of the proposed method, in 3.1, we will introduce the iterative partially matrix shrinkage (IPMS) algorithm in [9], in which the proposed method is based. After that, in 3.2, we formulate the relaxed problem for (6) and provide an IPMS based algorithm.

3.1 Iterative Partially Matrix Shrinkage

Let us consider the following problem,

$$\text{Minimize } \mathbf{rank}(X) \ \text{ subject to } X \in \mathcal{H} \cap \mathcal{I}. \tag{7}$$

Several algorithms have been proposed in order to solve this problem, and one of the most widely used algorithms is the nuclear norm minimization approach, which gives a low rank solution by solving the following problem,

$$\text{Minimize } \|X\|_* \ \text{ subject to } X \in \mathcal{H} \cap \mathcal{I}, \tag{8}$$

where $\|X\|_*$ denotes the nuclear norm of X, that is, the sum of its singular values. Instead of minimizing the nuclear norm, the IPMS algorithm provides a low rank solution by approximately solving the following the sum of non-dominant singular values minimization problem,

$$\text{Minimize } \|X\|_{*,r} \ \text{ subject to } X \in \mathcal{H} \cap \mathcal{I}, \tag{9}$$

where $\|X\|_{*,r}$ is the sum of non-dominant singular values of X defined by using its ith biggest singular value $\sigma_i(X)$ as $\|X\|_{*,r} = \sum_{i=r+1}^{end} \sigma_i(X)$. The IPMS algorithm solves (9) by iterating the following update schemes,

Step 1 $Z \leftarrow \mathcal{T}_{r,\lambda}(X)$.
Step 2 $X \leftarrow \mathcal{P}_{\mathcal{H}}(\mathcal{P}_{\mathcal{I}}(Z))$.

where $\mathcal{T}_{r,\lambda}(X)$ denotes the partial soft thresholding operator which replaces the ith singular values of X with $\max(\sigma_i - \lambda, 0)$ for $i \geq r+1$, $\mathcal{P}_{\mathcal{I}}$ and $\mathcal{P}_{\mathcal{H}}$ denote orthogonal projections onto $\mathcal{I}$ and $\mathcal{H}$ defined respectively by

$$\{\mathcal{P}_{\mathcal{I}}(X)\}_{i,j} = \begin{cases} C & \text{if } (i,j) \in \Pi^+ \wedge X_{i,j} < C \\ -C & \text{if } (i,j) \in \Pi^- \wedge X_{i,j} > -C \\ Y_{i,j} & \text{if } (i,j) \notin \Pi^+ \cup \Pi^- \\ X_{i,j} & \text{otherwise} \end{cases},$$

$$\{\mathcal{P}_{\mathcal{H}}(X)\}_{i,j} = \frac{\sum_{k+l=i+j} X_{k,l}}{\min(N, i+j-1, M+N-i-j+1)},$$

and $\{\cdot\}_{i,j}$ denotes the (i, j) entry of a matrix. Since (9) requires the value of r regarding with a matrix rank, the IPMS estimates a matrix rank r during iterations by using the scheme,

$$r \leftarrow \operatorname*{argmin}_{\hat{r}} \sigma_{\hat{r}} \text{ s.t. } \sigma_{\hat{r}} \geq \alpha\sigma_1,$$

where $\alpha < 1$ is a given constant. The details of the IPMS algorithm are written in [9], and its performance is also shown in [10] comparing with other algorithms.

3.2 Multiple Matrix Rank Minimization Algorithm

Next we focus on a multiple matrix rank minimization problem (6). The nuclear norm relaxation cannot work to solve (6) because it holds that

$$\sum_{i=1}^{L} \|D_i X\|_* \geq \left\| \sum_{i=1}^{L} D_i X \right\|_* = \|X\|_*,$$

that is, a solution is always equal to the nuclear norm relaxation problem for a single matrix rank minimization, and the formulation using D_i does not work better than that without using D_i. Hence this letter applies the relaxation used in the IPMS algorithm and provides the following problem,

$$\begin{array}{ll} \text{Minimize} & \sum_{i=1}^{L} \|D_i X\|_{*,r_i} \\ \text{subject to} & \sum_{i=1}^{L} D_i = I, D_i \in \mathcal{D}, X \in \mathcal{H} \cap \mathcal{I}, \end{array} \tag{10}$$

which is a multiple matrix version of (9) with the constraint for declipping problem. In order to obtain an approximate solution, this letter relaxes this problem as follows,

$$\begin{array}{ll} \text{Minimize} & \sum_{i=1}^{L} \left\{ \gamma_i \|Z_i\|_{*,r_i} + \frac{1}{2}\|Z_i - D_i X\|_F^2 \right\} \\ \text{subject to} & \sum_{i=1}^{L} D_i = I, D_i \in \mathcal{D}, X \in \mathcal{H} \cap \mathcal{I}, \end{array} \tag{11}$$

and provides the following scheme which updates Z_i, D_i and X alternately,

Step 1 $\quad Z_i \leftarrow \mathcal{T}_{r_i,\gamma_i}(D_i X) \quad$ for $i = 1, 2, \ldots, L.$

Step 2 $\quad \{D_i\}_{i=1}^{L} \leftarrow \underset{\hat{D}_1,\ldots,\hat{D}_L \in \mathcal{D}}{\text{argmin}} \sum_{i=1}^{L} \|Z_i - \hat{D}_i X\|_F^2$ subject to $\sum_{i=1}^{K} \hat{D}_i = I.$

Step 3 $\quad X \leftarrow \underset{\hat{X}}{\text{argmin}} \sum_{i=1}^{L} \|Z_i - D_i \hat{X}\|_F^2$ subject to $\hat{X} \in \mathcal{H} \cap \mathcal{I}$.

In **Step 2**, the jth diagonal element $(d^{(i)})_j$ of D_i can be obtained analytically by

$$(d^{(i)})_j = \frac{1}{L}\left(1 - \sum_{k=1}^{L} \frac{\langle \boldsymbol{z}_j^{(k)}, \boldsymbol{x}_j \rangle}{\langle \boldsymbol{x}_j, \boldsymbol{x}_j \rangle}\right) + \frac{\langle \boldsymbol{z}_j^{(i)}, \boldsymbol{x}_j \rangle}{\langle \boldsymbol{x}_j, \boldsymbol{x}_j \rangle}, \tag{12}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product, and $\boldsymbol{x}_j$ and $\boldsymbol{z}_j^{(i)}$ denote jth row vector of X and Z_i, respectively. The above update scheme gives $(d^{(i)})_j$ in $[0, 1]$ not $\{0, 1\}$. To obtain $(d^{(i)})_j$ which is nearly eqaul to 0 or 1, we introduce a heuristic soft shrinkage technique used in the sparse optimization. Let d_j to be a vector in $\mathbb{R}^L$ whose ith element is $(d^{(i)})_j$. Then d_j should be a sparse vector, and therefore this letter proposes an additional update scheme to make d_j sparse as follows,

$$(d^{(i)})_j \leftarrow \max\left(0, \frac{1}{L}\left(1 - \sum_{k=1}^{L} \frac{\langle \boldsymbol{z}_j^{(k)}, \boldsymbol{x}_j \rangle}{\langle \boldsymbol{x}_j, \boldsymbol{x}_j \rangle}\right) + \frac{\langle \boldsymbol{z}_j^{(i)}, \boldsymbol{x}_j \rangle}{\langle \boldsymbol{x}_j, \boldsymbol{x}_j \rangle} - \tau\right). \tag{13}$$

To satisfy the constraint $\sum_{i=1}^{L} D_i = I$, we use a projection defined as $(d^{(i)})_j \leftarrow (d^{(i)})_j / \sum_{k=1}^{L} (d^{(k)})_j$ after obtaining D_i by (13). Note that the update scheme (13) does not always give $(d^{(i)})_j$ in $\{0, 1\}$ while empirical results show that the number of binary elements is increased by (13). Finally, a multiple rank minimization based signal declipping algorithm is obtained as shown in Algorithm 1. Algorithm 1 requires initial values of D_i except to satisfy all $D_i = \frac{1}{L} I$ since there is no update when the initial values of all entries $(d^{(i)})_j$ satisfy $(d^{(i)})_j = 1/L$. Because audio signals do not often switch their AR model, random values are not suitable for initial values of D_i. This paper gives the initial values as the following window function, which changes smoothly according to i and j.

Algorithm 1: Audio declipping algorithm based on IPMS

Input : $Y, \mathcal{H} \cap \mathcal{I}, \{D_i\}_{i=1}^{L}, \alpha, \alpha_{min}, \eta_\alpha, \lambda, \tau, \epsilon, t_{max}$

$t \leftarrow 0$
$X \leftarrow Y$
repeat
  $X^{old} \leftarrow X$
  $\alpha \leftarrow \max(\alpha/\eta_\alpha, \alpha_{min})$
  for $i = 1$ to L **do**
    $[U, \sigma_1, \sigma_2, \cdots, \sigma_N, V] \leftarrow \mathbf{svd}(D_i X)$
    $r_i \leftarrow \underset{\hat{r}}{\text{argmin}}\, \sigma_{\hat{r}}$ s.t. $\sigma_{\hat{r}} \geq \alpha \sigma_1$
    $Z_i \leftarrow \mathcal{T}_{r_i, \lambda\sigma_{r_i}}(D_i X)$
  end
  for $(i, j) \in \{1, 2, \cdots, L\} \times \{1, 2, \cdots, N\}$ **do**
    $(d^{(i)})_j \leftarrow \max\left(0, \frac{1}{L}\left(1 - \sum_{k=1}^{L} \frac{\langle \boldsymbol{z}_j^{(k)}, \boldsymbol{x}_j \rangle}{\langle \boldsymbol{x}_j, \boldsymbol{x}_j \rangle}\right) + \frac{\langle \boldsymbol{z}_j^{(i)}, \boldsymbol{x}_j \rangle}{\langle \boldsymbol{x}_j, \boldsymbol{x}_j \rangle} - \tau\right)$
  end
  $(d^{(i)})_j \leftarrow (d^{(i)})_j / \sum_{k=1}^{L} (d^{(k)})_j$ for all i, j
  $X \leftarrow \left(\sum_{k=1}^{L} D_i^2\right)^{-1} \left(\sum_{k=1}^{L} D_i Z_i\right)$
  $X \leftarrow \mathcal{P}_{\mathcal{H} \cap \mathcal{I}}(X)$
  $t \leftarrow t + 1$
until $\|X - X^{old}\|_F / \|X\|_F < \epsilon$ *or* $t_{max} < t$;
Output: X

$$(d^{(i)})_j = \exp\left\{ -\left(\frac{j}{M} - \frac{(i-1)}{L-1} \right)^2 \Big/ 2\sigma^2 \right\},$$

where $\sigma = 10/L$, and then each initial value of $\{D_i\}_{i=1}^{L}$ is obtained by normalizing $(d^{(i)})_j$ such that $\sum_{i=1}^{L} (d^{(i)})_j = 1$.

4. Numerical Examples

This section demonstrates numerical examples to show the declipping quality of Algorithm 1 comparing with the rank minimization approach using null space based alternating optimization (NSAO) [7], the orthogonal matching pursuit (OMP) based algorithm [2] and persistent empirical Wiener (PEW) algorithm [3]. In all examples, signals were range-normalized as $\max(|y_k|) = 1$, and we compare the performance using the clipping level $C = 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$. In Algorithm 1, each signal

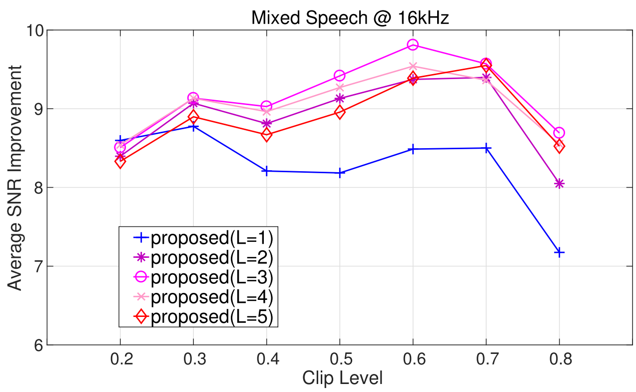

Fig. 1 Average SNR of Algorithm 1 with $L = 1, 2, 3, 4, 5$ for 10 mixed speech signals.

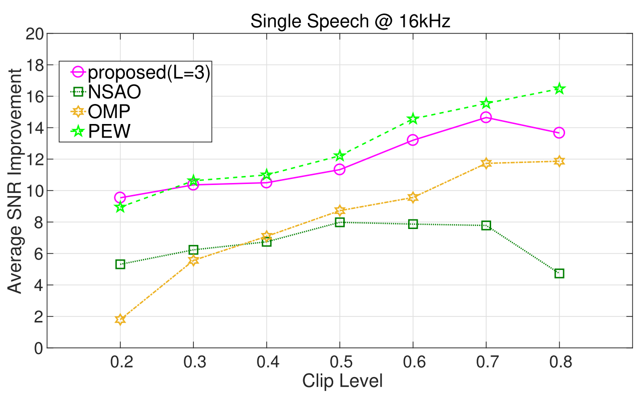

Fig. 2 Average SNR for 10 single speech signals for clipping levels $C = 0.2, 0.3, \cdots, 0.8$.

is separated into short-time frames (1024 samples) and restored sequentially using 75% overlap with the previous frame, and we use $N = 128$, $\alpha = 1$, $\alpha_{min} = 5.0 \times 10^{-3}$, $\eta_\alpha = 1.01$, $\lambda = 1.0 \times 10^{-1}$, $\tau = 1.0 \times 10^{-1}$, $\epsilon = 10^{-5}$ and $t_{max} = 3 \times 10^3$. We utilize the speech data sets sampled at 16kHz consisting of 5 different audio containing both male and female voice, which are available at the web site†, and the algorithms are evaluated by the signal-to-noise ratio (SNR).

First we examined Algorithm 1 with $L = 1, 2, 3, 4, 5$. Figure 1 shows the results of declipping 10 mixed speech signals. We can see that multiple matrix rank minimization achieves better performance than single matrix rank minimization ($L = 1$) for almost all clipping levels and that the best performance is achieved at $L = 3$. Next Algorithm 1 with $L = 3$ was compared with other algorithms, and Figs. 2 and 3 show the results of single speech signals and mixed speech signals, respectively. As can be seen, the proposed algorithm has a high accuracy for both the single and mixed signals as compared to NSAO based algorithm and OMP algorithm. Though PEW algorithm recovers single speech signals the best of all, the proposed algorithm achieves the best declipping for mixed speech signals of $0.2 \leq C \leq 0.7$. These results indicate that the proposed algorithm works well for declipping mixed signals. Because mixed speech signals have less sparsity and because PEW algorithm utilizes the sparsity of signals in the time-frequency its performance for mixed speech signals is worse than that for single speech signals.

†http://www.small-project.eu/software-data/AudioInpaintingToolbox.zip/view.html

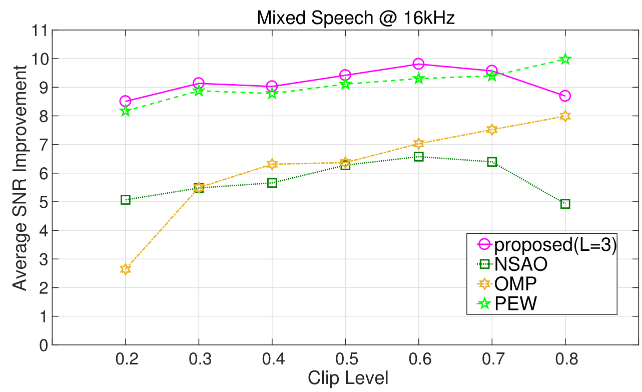

Fig. 3 Average SNR for 10 mixed speech signals for clipping levels $C = 0.2, 0.3, \cdots, 0.8$.

5. Conclusion

This letter proposed a new AR model and matrix rank minimization approach to audio declipping problem. Assuming that audio signals are modeled by multiple AR models, the problem is formulated as a multiple matrix rank minimization problem. To solve this problem, an iterative algorithm was proposed based on the IPMS algorithm. Numerical examples show that the proposed algorithm can recover the mixed speech signal efficiently.

References

[1] J. Tang, "Evaluation of double sided periodic substitution (DSPS) method for recovering missing speech in packet voice communications," Proc. 10th Annual International Phoenix Conference on Computers and Communications, pp.454–458, 1991.

[2] A. Adler, V. Emiya, M.G. Jafori, M. Elad, R. Gribonval, and M.D. Plimbley, "A constrained matching pursuit approach to audio declipping," Proc. IEEE ICASSP, pp.329–332, 2011.

[3] K. Siedenburg, M. Kowalski, and M. Dorfler, "Audio declipping with social sparsity," Proc. IEEE ICASSP, pp.1577–1581, 2014.

[4] A. Janssen, R. Veldhuis, and L. Vries, "Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes," IEEE Trans. Acoust. Speech Signal Process., vol.34, no.2, pp.317–330, 1986.

[5] M. Verhaegen and P. Dewilde, "Subspace Model Identification Part I: The Output-error State Space Model Identification Class of Algorithms," Int. J. Control, vol.56, no.5, pp.1187–1210, 1992.

[6] T. Takahashi, K. Konishi, and T. Furukawa, "Hankel structured matrix rank minimization approach to signal declipping," Proc. EUSIPCO, pp.1–5, 2013.

[7] T. Takahashi, K. Uruma, K. konishi, and T. Furukawa, "Block Adaptive Algorithm for Signal Declipping Based on Null Space Alternating Optimization," IEICE Trans. Inf. & Syst., vol.E98-D, no.1, pp.206–209, 2015.

[8] R. Vidal, Y. Ma, and S. Sastry, "Generalized principal component

analysis (GPCA)," IEEE Trans. Pattern Anal. Mach. Intell., vol.27, no.12, pp.1945–1959, 2005.

[9] K. Konishi, K. Uruma, T. Takahashi and T. Furukawa, "Iterative partial matrix shrinkage algorithm for matrix rank minimization," Signal Processing, vol.100, pp.124–131, 2014.

[10] D. Lazzaro, "A nonconvex approach to low-rank matrix completion using convex optimization," Numerical Linear Algebra with Applications, vol.23, no.5, pp.801–824, 2016.