M-ESTIMATE ROBUST PCA FOR SEISMIC NOISE ATTENUATION

Hojjat Akhondi-Asl, James D. B. Nelson

Department of Statistical Science, University College London, UK
{h.akhondi-asl, j.nelson}@ucl.ac.uk

ABSTRACT

The robust principal component analysis (PCA) method has shown very promising results in seismic ambient noise attenuation when dealing with outliers in the data. However, the model assumes a general Gaussian distribution plus sparse outliers for the noise. In seismic data however, the noise standard variation could vary from one place to another leading to a more heavy-tailed noise distribution. In this paper, we present a new method which solves a convex minimisation problem of the robust PCA method with an M-estimate penalty function. Our empirical results show that the proposed method can outperform the robust PCA method.

Index Terms— Robust PCA, M-Estimate, Seismic Noise Attenuation, Convex Optimisation, ADMM

1. INTRODUCTION

Robust statistical methods combined with recent advancements in convex optimisation techniques have a great potential to rejuvenate some of the conventional noise attenuation techniques [1, 2, 3, 4, 5]. In particular, seismology has very recently recieved interest from robust statistical methods with convex optimisation [6, 5]. In a seismic acquisition system, a controlled seismic source excites the earth and generates acoustic waves that travel through the Earth layers. Seismic waves are recorded by the receivers (usually in the order of thousands) where the recorded reflected and refracted waves are used to study oil and gas reservoirs, minerals, faults or other scientific investigations [7].

One of the major challenges in seismic acquisition is how to maximise the signal-to-noise-ratio (SNR), given a fixed number of sources used and receivers deployed. Ideally one wants to keep the incoming reflections (signal) and disregard any other unwanted signal (noise). Noise in the seismic data are generally divided into two subcategories, coherent noise and incoherent noise. Coherent noise are unwanted signals that are locally coherent (correlated) from trace to trace. Examples are ground-roll, multiples and air waves. Incoherent noise or ambient noise are unwanted signals that are random and not correlated. Some examples are wind, land/marine animal activity and rain. Conventionally, random noise in seismic data are assumed to have a Gaussian distribution, with SNR related to $\sqrt{N}$ where N is the number of sensors.

Assuming noise with Gaussian distribution, rank reduction based methods work pretty well for denoising data [8, 9, 10, 11]. Rank reduction methods work on the assumption that within a small spatio-temporal seismic window, seismic reflections are linear events with different dips (gradients). In these methods [8, 9, 10, 11], data are first transformed to the Fourier domain in the temporal direction (also known as Fx transform [7]) and then for every frequency slice, rank reduction is applied on the Hankel matrix to remove incoherent signals. Cadzow's algorithm [12], popular in array processing, can also be applied for further noise reduction [9]. While these methods perform well under Gaussian noise assumption, the results are highly degraded when there are outliers in the data. This is because principal components analysis, which is the principal tool for rank reduction methods, is only optimal when the noise is Gaussian. Any outlier in data would seriously degrade the results. Due to erratic noise in the seismic data and variable standard deviations, experiments have shown that the Gaussian distribution does not model the ambient noise well. This has led to many recent robust denoising techniques, based on the rank reduction method [13, 5].

In [13, 5], M-estimate penalty functions (Bisquare and Huber respectively) were proposed to deal with the outliers. M-estimate function are functions [14] that combine quadratic cost function with another cost function that is more robust to outliers. These methods involve a rank minimization process (non-convex), usually utilizing the iterative-least-squares-minimization (IRLS) method. More recently, Chen et al. [5] proposed using the Robust PCA [1] method for seismic ambient noise attenuation. Robust PCA method involves a nuclear norm minimization (convex) which has shown its potential in many fields [15, 1, 2, 3, 4]. However, the model still assumes a general Gaussian noise model with sparse outliers.

In this paper we present a novel method that combines the robust PCA framework with M-estimation penalty function. We show that the corresponding objective can be phrased as a convex optimisation problem. We show that the flexibility on the penalty function gives rise to results with enhanced

robustness to heavy-tailed noise distribution. The rest of the paper is as follows. In Section 2 necessary background on seismic parametric denoising is presented. In Section 3 we first describe the robust PCA method and then present the proposed algorithm to combine robust PCA with an M-estimation penalty function. In Section 4, we present synthetic and field data results and demonstrate the superiority of our method over the robust PCA for seismic ambient noise removal. In Section 5, we offer conclusions and ideas for possible future directions.

2. BACKGROUND

Within a small spatio-temporal seismic window, linear events in time-offset domain can be described as [7, 5]:

$$d(t, x) = a(t - px), \tag{1}$$

where t is an index for time samples recorded in the sensors, x is the index for spatial samples, a is the reflected wavelet, and p is the ray parameter (dip or slope of the event). If we we apply the Fourier transform in the temporal domain (known as Fx transform in the seismic community) and assuming we have regular spatial intervals (spacing between sensors), that is $x = j\Delta x$, we obtain:

$$D(\omega, j) = A(\omega)e^{-i\omega pj\Delta x}. \tag{2}$$

In the case of multiple events, we obtain a sum of exponentials with frequency-varying amplitudes:

$$D(\omega, j) = \sum_{k=1}^{K} A_k(\omega)e^{-i\omega p_k j\Delta x}. \tag{3}$$

Under noise-less conditions, the Hankel matrix[1] of sum of exponentials with regular sampling and frequency-constant amplitudes has rank equal to K where K is the number of events. Under noisy conditions however, the Hankel matrix will be full rank. Assuming noise follows a Gaussian distribution, we can obtain an optimal solution by keeping only the first K principal components (or singular values) of the Hankel matrix. Given the model, we can achieve this property by applying the rank reduction process for every frequency slice, that is:

$$\hat{D}_{\omega_i}(j) = \mathcal{A}(\mathcal{R}_{\mathcal{K}}(\mathcal{H}(D_{\omega_i}(j)))), \tag{4}$$

where $D_{\omega_i}(j) = \sum_{k=1}^{K} \tilde{A}_k e^{-i\omega p_k j\Delta x}$, and $\tilde{A}_k = A_k(\omega_i)$ and i is an index to cover the frequency range of the data. Here $\mathcal{H}(D_{\omega_i}(j))$ computes the Hankel matrix of the vector of components of the frequency slice, $\mathcal{R}_{\mathcal{K}}$ computes the K rank approximation $U_K \times \Sigma_K \times V_K^H$, where Σ_K is the diagonal matrix containing the singular values up to K, and $\mathcal{A}$ computes the Cadzow's algorithm which is simply anti-diagonal averaging of the Hankel matrix [12]. Although both mathematically elegant and popular with practitioners, this method

---

[1]Hankel matrix is directly related to Toeplitz matrix. The Hankel matrix of a vector $h_0, h_1, \ldots, h_n$ is $H_{i,j} = h_{i+j-2}$

is extremely sensitive to outliers. Even one outlier has a massive effect on the performance of the method.

3. M-ESTIMATE ROBUST PCA

Robust PCA method, recently introduced to the community by Candes et al. [1], is a robust technique which involves nuclear norm minimization. One of the benefits of this method is the replacement of non-convex low rank minimisation with the convex problem of nuclear norm minimisation, therefore allowing usage of convex optimisation methods [16]. Nuclear norm, denoted by $||L||_*$ measures the sum of the singular values of L, that is $||L||_* = \sum_i \sigma_i(L)$, where σ_i are the singular values of L. When applied to seismic data in the Fx domain, the technique can be described as follows:

Let us denote H to be the Hankel matrix of a frequency slice of a small spatio-temporal seismic window. The Hankel matrix H can be modelled as:

$$H = L + S + N, \tag{5}$$

where L is the low rank matrix that includes all the events, S is a sparse matrix that contains the outliers and N contains the Gaussian noise. H can be minimised with respect to L, S and N in a convex optimisation methodology, that is:

$$\min_{L,S} ||L||_* + \lambda_1 ||S||_1 + ||H - L - S||_F^2. \tag{6}$$

Here $||S||_1$ is used to impose sparsity on S to capture the outliers and $||H - L - S||_F^2$ minimises the difference matrix with quadratic cost function, optimal when noise is Gaussian distributed. The assumption that matrix $N = H - L - S$ is Gaussian distributed can be challenged, specially when dealing with seismic data. For example, noise variance could vary from one section to another or the noise distribution of N is close to heavy-tailed distribution. To alleviate this deficiency, we propose to use M-estimators combined with robust PCA to tackle these issues. There are two ways to approach this problem, either the minimisation can be done using:

$$\min_{L,S} ||L||_* + \lambda_1 ||S||_1 + \rho(H - L - S), \tag{7}$$

or using:

$$\min_{L,S} ||L||_* + \rho(H - L), \tag{8}$$

where $\rho(X)$ in both cases is an M-estimator penalty function. Our empirical results show that the former approach is more robust and less sensitive to parameter change. Moreover, equation 7 is a natural extension of equation 6, where in the case $\rho(X)$ is Gaussian likelihood, equation 7 collapses down to 6. In this paper we consider the Huber function [17] as the the M-estimator but other cost functions can also be considered. Huber function is generally described as as:

$$\rho_\gamma(X) = \begin{cases} \frac{1}{2}||X||_2^2 & |X| < \gamma \\ \gamma||X||_1 - \frac{\gamma^2}{2} & |X| \geq \gamma. \end{cases}$$

The minimisation problem can be solved with any convex optimisation technique [16]. One technique which is popular for its simplicity and its fast convergence is the alternating direction method of multipliers (ADMM) [18]. The ADMM method, an augmented Lagrangian method, can solve unconstrained/constrained convex minimisation problems. Given the minimisation problem above, we can re-write the problem as a constrained problem as follows:

$$\min_{L,S,Z} \quad ||L||_* + \lambda_1||S||_1 + \frac{1}{\mu}\rho(Z),$$
$$\text{subject to} \quad H - L - S = Z. \tag{9}$$

Th augmented Lagrangian function of the above problem is therefore:

$$\mathcal{L}(L,S,Z,Y) = ||L||_* + \lambda_1||S||_1 + \frac{1}{\mu}\rho(Z) +$$
$$Y^T(H-L-S-Z) + \frac{\beta}{2}||H-L-S-Z||_F^2, \tag{10}$$

where Y is the called the Lagrangian multiplier. Minimizing with respect to L, S and Z yields the ADMM's variable updates for this problem:

$$Z^{K+1} = \min_Z \frac{1}{\mu}\rho(Z) - Y^K Z + \frac{\beta}{2}||H - L^K - S^K - Z||_F^2 .$$
$$L^{K+1} = \min_L ||L||_* - Y^K L + \frac{\beta}{2}||H - L - S^K - Z^K||_F^2 .$$
$$S^{K+1} = \min_S \lambda_1||S||_1 - Y^K S + \frac{\beta}{2}||H - L^K - S - Z^K||_F^2 . \tag{11}$$

By applying the minimisations, we reach the state update equations, described in Algorithm I. Here, $\mathcal{S}$ denotes element-wise soft-thresholding and $\mathcal{D}$ denotes soft-thresholding on the singular values. It is important to mentioned that, given the separable structure of the minimisation model, Algorithm I can be also implemented in a parallel framework.

Algorithm 1 ADMM steps for M-estimate Robust PCA

1: **Initializations:** $Z = 0$, $L = 0$, $S = 0$; $Y = 0$; $k = 0$, iter = 250.
2: **while** not converged & $k \leq$ iter **do:**
3: $\quad T^K = \frac{1}{\beta}Y^K + H - L^K - S^K$
4: $\quad Z^{K+1} = \begin{cases} \frac{\mu\beta}{1+\mu\beta} \times T^K_{i,j} & |T^K_{i,j}| < \gamma \\ \mathcal{S}_{\gamma + \frac{\gamma}{\mu\beta}}(T^K_{i,j}) & |T^K_{i,j}| \geq \gamma \end{cases}$
5: $\quad L^{K+1} = \mathcal{D}_{\frac{1}{\beta}}(\frac{1}{\beta}Y^K + H - S^K - Z^K)$
6: $\quad S^{K+1} = \mathcal{S}_{\frac{\lambda_1}{\beta}}(\frac{1}{\beta}Y^K + H - L^K - Z^K)$
7: $\quad Y^{K+1} = Y^K + \beta(H - L^K - S^K - Z^K)$
8: $\quad$ END LOOP if:
9: $\quad \frac{||L^{K+1} - L^K||_F^2}{||L^K||_F^2} < \epsilon$ &
10: $\quad \frac{||S^{K+1} - S^K||_F^2}{||S^K||_F^2} < \epsilon$
11: $\quad k = k + 1$

3.1. Parameter Selection

In [1, 2] it is shown that $\lambda_1 = 1/\sqrt{\max(M,N)}$ can guarantee good recovery results, where M and N are the dimensions of the matrix. In [1, 2], it is also recommended to choose $\mu = \frac{1}{10}\sqrt{\min(M,N) + \sqrt{8\min(M,N)}}\sigma$, where σ is the standard deviation of the Gaussian noise and $\beta = \eta\frac{M \times N}{||H||_1}$, where η is related to the percentage of outliers in the data. The stopping criterion parameter ϵ depends on the data and can be tunable (generally 10^{-5} works fine for most cases). The Huber parameter γ is tunable, depending on the data.

4. RESULTS

4.1. Synthetic Data

In this section we compare our proposed method with the robust PCA and also the standard PCA on a synthetic data. The synthetic data we have produced (Figure 1a) consists of three separate events with different dips generated using Ricker wavelet with max frequency at around 100Hz. The data is about 0.4s long with sampling interval of 0.001s. We take 50 traces which describes our spatial window. We add three major outliers to the data and add Gaussian noise, with variable standard deviation, to each trace (Figure 1b). As can be observed from denoised data in Figure 1c and the difference in 1d, the standard PCA heavily degrades the results largely due to the outliers in the data. Figure 1e and 1f show the results we obtain for robust PCA. The results have dramatically improved compared to the standard PCA. Figures 1g and 1h show the results for M-estimate Robust PCA. Comparing Robust PCA and M-estimate robust PCA, we can observe that both methods dramatically improve the results, however, robust PCA method suffers from signal leakage (all colour-scales are the same). The M-estimate robust PCA outperforms the robust PCA method by almost 2dB in this example. SNR was calculated as $10\log_{10}\frac{||G||_2^2}{||G-\hat{G}||_2^2}$, where G is the clean signal and $\hat{G}$ is the denoised signal. All parameters are selected according Section 3.1, the Huber parameter was set at 5 and η was set at 0.15.

4.2. Field Data

We acquired some post-stack 2D marine seismic data from U.S. Geological Survey [19], available for anyone to use. The sampling interval for this acquisition was 0.004s. We applied both the robust PCA and M-estimate robust PCA methods on small overlapping spatio-temporal windows of 240ms by 60 traces in the spatial direction. The overlapping size between the windows was chosen to be $\frac{1}{2}$ with cosine tapered windows. The chosen section of the data has two visible outliers. Figure 2 shows the results for both scenarios. As can be observed both methods deal well with the outliers in the data.

However, the M-estimate robust PCA method has less signal leakage when compared to the robust PCA method. All parameters are selected according Section 3.1. The noise estimate used for the μ parameter was calculated based on a noise snapshot captured from the data. The Huber parameter was set at 0.5 and η was set at 0.15. From the two sets of examples shown, we can see that the M-estimate robust PCA can outperform the robust PCA method and add further robustness to the denoised results.
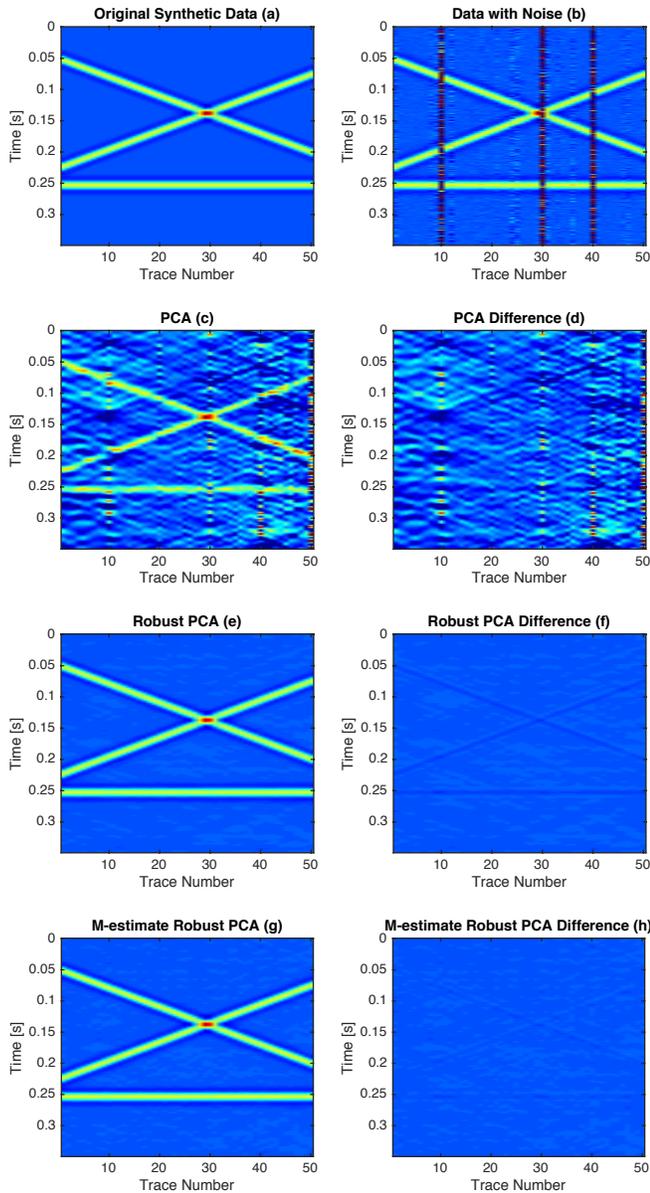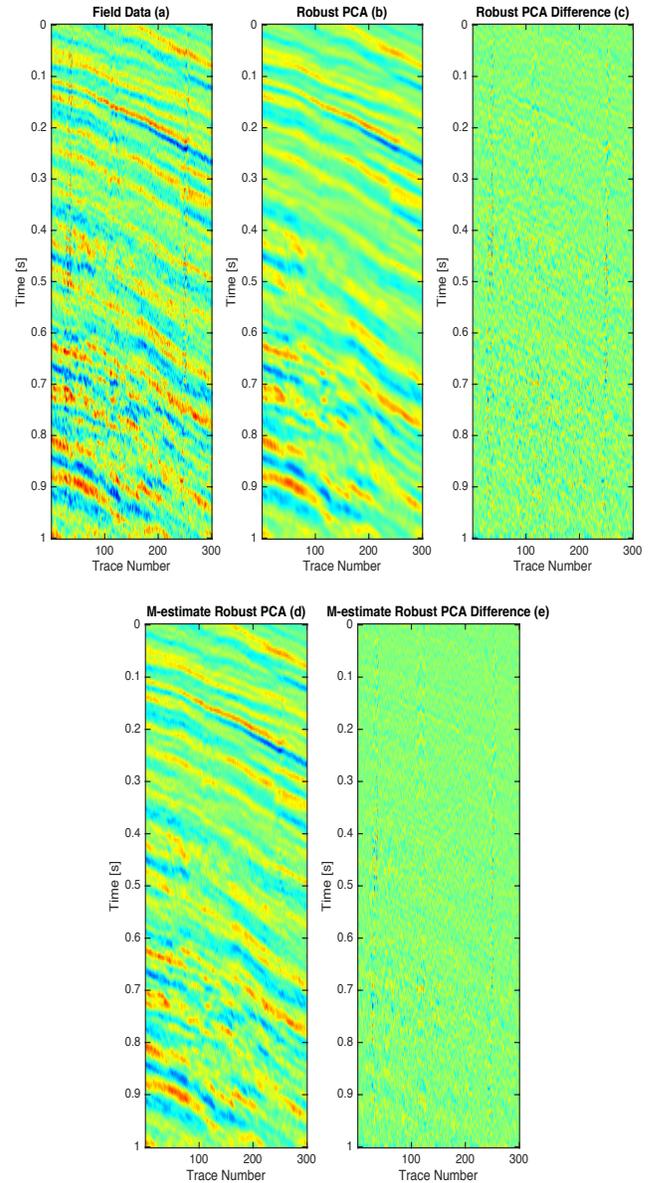

Fig. 2. Comparison of different denoising methods on sample field data. (a) Original Data. (b) Robust PCA. (c) Robust PCA result's difference with the original data. (d) M-estimate robust PCA. (e) M-estimate robust PCA result's difference with the original data. All figures have the same colour-scales.

5. CONCLUSIONS

In this paper we proposed a novel seismic noise attenuation method that can be more robust to non-Gaussian noise distribution, compared to the state-of-the-art methods. We showed that the proposed M-estimate robust PCA method can outperform the robust PCA method. One of the advantages of this method is that it can be readily turned into a robust PCA method. One possible future work is to devise a close-form equation for the γ parameter that guarantees good recovery.

Fig. 1. Comparison of different denoising methods on synthetic data. (a) Original Synthetic Data. (b) Data with noise. (c) Standard PCA. (d) Standard PCA result's difference with the original data. (e) Robust PCA. (f) Robust PCA result's difference with the original data. (g) M-estimate robust PCA. (h) M-estimate robust PCA result's difference with the original data.

6. REFERENCES

[1] E.J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 11, 2011.

[2] M. Tao and X. Yuan, "Recovering low-rank and sparse components of matrices from incomplete and noisy observations," *SIAM Journal on Optimization*, vol. 21, no. 1, pp. 57–81, 2011.

[3] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Advances in neural information processing systems*, 2009, pp. 2080–2088.

[4] H. Xu, C. Caramanis, and S. Sanghavi, "Robust pca via outlier pursuit," in *Advances in Neural Information Processing Systems*, 2010, pp. 2496–2504.

[5] K. Chen, "Robust matrix rank reduction methods for seismic data processing," M.S. thesis, University of Alberta, 2013.

[6] N. Kreimer and M.D. Sacchi, "Nuclear norm minimization and tensor completion in exploration seismology," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 4275–4279.

[7] R.E. Sheriff and L.P. Geldart, *Exploration seismology*, Cambridge university press, 1995.

[8] S.R. Trickett, "F-xy eigenimage noise suppression," *Geophysics*, vol. 68, no. 2, pp. 751–759, 2003.

[9] S.R. Trickett, "F-xy cadzow noise suppression," in *2008 SEG Annual Meeting*. Society of Exploration Geophysicists, 2008.

[10] V. Oropeza and M.D. Sacchi, "Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis," *Geophysics*, vol. 76, no. 3, pp. V25–V32, 2011.

[11] K. Chen and M.D. Sacchi, "Robust reduced-rank filtering for erratic seismic noise attenuation," *Geophysics*, vol. 80, no. 1, pp. V1–V11, 2014.

[12] J.A. Cadzow, "Signal enhancement - a composite property mapping algorithm," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 36, no. 1, pp. 49–62, 1988.

[13] S.R. Trickett, L. Burroughs, and A. Milton, "Robust rank-reduction filtering for erratic noise," in *82nd Annual International Meeting, SEG, Expanded Abstracts*, 2012, pp. 1–5.

[14] R. Maronna, D. Martin, and V. Yohai, *Robust statistics*, John Wiley & Sons, Chichester. ISBN, 2006.

[15] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 1, pp. 171–184, 2013.

[16] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge university press, 2004.

[17] P.J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[18] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[19] US Geological Survey, http://www.usgs.gov/.