A COMPUTATIONALLY EFFICIENT ALGORITHM FOR HIGH QUALITY SEPARATION OF SIMULTANEOUS SOURCES IN SEISMOLOGY

Aboulnasr Hassanien, Sergiy A. Vorobyov

University of Alberta
Dept. Electrical and Computer Engineering
Edmonton, AB, T6G 2V4, Canada
hassanie@ualberta.ca
vorobyov@ece.ualberta.ca

Mauricio Saachi, Mostafa Naghizadeh

University of Alberta
Dept. of Physics
Edmonton, AB, T6G 2E1, Canada
msacchi@ualberta.ca
mostafan@ualberta.ca

ABSTRACT

We consider the problem of separating simultaneous source blended data in applied seismology. Cross-source interference that masks the desired signal in the common source domain can be translated into incoherent noise by rearranging the data in the common receiver domain. We show that applying a virtual blending/deblending process to the data in the common receiver domain enables obtaining an additional noisy version of the data. By measuring the local similarities and dissimilarities between the two noisy versions of the data, it is possible to discriminate between corrupt and non-corrupt data points. Corrupt data points can be replaced by a weighted sum (e.g., averaging) of neighboring non-corrupt data points. The proposed method is applied directly in the time-space domain, i.e., no computationally expensive data transformation is needed. Moreover, it can be straightforwardly extended to higher-dimensional data scenarios. Simulation results are given to validate the effectiveness of the proposed method.

***Index Terms*—** Simultaneous data acquisition, signal separation, seismic data.

1. INTRODUCTION

In conventional seismic data acquisition, sources need to be separated by sufficient time to avoid overlap between the responses of different sources. However, this makes the acquisition process time-consuming and economically-expensive. Recently, the emerging concept of simultaneous source blended data acquisition has received a considerable attention [1]–[3]. Simultaneous data acquisition is used to either reduce the cost by reducing the time interval between successive sources or to increase the quality of the data by increasing the number of sources within a specific survey interval. The use of simultaneous sources with small random time delays was presented in [1]. The concept was extended to the incoherent shooting in [2], [3]. However, the benefits of simultaneous source data acquisition comes at the price of cross-source interference.

The problem of cross-source interference removal in blended data acquisition has been addressed in [4]–[6]. It has been shown that the problem of cross-source interference mitigation can be treated as a de-noising one. In particular, it has been shown that the cross-source interference, which appears as a coherent interference in the common source domain, can be translated into a non-coherent noise by rearranging the data in the common receiver domain.

In this paper, we show that by applying the blending/deblending process to the data in the common receiver domain an additional noisy version of the desired signal can be obtained. Due to the random nature of blending, the resulting noise component is incoherent. We propose a method for mitigating the incoherent noise by processing the two noisy versions of the data jointly. In particular, we use the local similarities/dissimilarities between the corresponding data at each time-space point to discriminate between noiseless and noisy points. Noisy points are replaced with a weighted sum (e.g., averaging) of neighboring noise-free points. The proposed method is applicable directly in the time-space domain and can be straightforwardly extended to higher-dimensional data. Simulation examples are used to demonstrate the effectiveness of the proposed method.

2. RECEIVED DATA MODEL

We assume N_s seismic sources and N_r seismic receivers. Let $\mathbf{X}_n$ be the matrix of dimension $K \times N_r$ which corresponds to the data recorded by the N_r receivers due to the n-th, $n = 1, \ldots, N_s$ source in the absence of cross-source interference, where K is the number of samples in the time dimension. The overall survey time required for conventional shooting is KN_s which can be too long especially when the number of sources is too large. To save survey time, sources are fired simultaneously. Assume that K_T is the total number

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

of samples gathered by each receiver due to N_s simultaneous sources, where $K_T \ll KN_s$. Let t_n be the time instant at which the n-th source is fired where t_n is selected randomly from the set $[1,\ 2,\ \ldots,\ K_T - K]$.

The $K_T \times Nr$ received blended data can be modeled as

$$\mathbf{X}_b = \sum_{n=1}^{N_s} \mathbf{U}_n \tag{1}$$

where the $K_T \times N_r$ matrix $\mathbf{U}_n$ associated with the n-th source is given by

$$\mathbf{U}_n = \left[\mathbf{0}^T_{t_n-1,N_r},\ \mathbf{X}^T_n,\ \mathbf{0}^T_{K_T-K-t_n,N_r}\right]^T. \tag{2}$$

In (2), $\mathbf{0}_{n,k}$ is a matrix of dimension $n \times k$ with all entries equal to zero and $(\cdot)^T$ stands for the transpose. Given the recorded blended data $\mathbf{X}_b$ and the time delays $t_1, \ldots, t_{N_s}$, our goal is to obtain the data $\mathbf{X}_1, \ldots, \mathbf{X}_{N_s}$ after removing the cross-source interference.

3. PROPOSED SOURCE SEPARATION METHOD

The direct way for obtaining the data associated with the n-th source from the received blended data $\mathbf{X}_b$ is the so-called pseudo-deblending which can be modeled as

$$\begin{aligned}\mathbf{D}_n &= \mathbf{P}^T_n \mathbf{X}_b \\ &= \mathbf{P}^T_n \mathbf{U}_n + \sum_{m=1, m\neq n}^{N_s} \mathbf{P}^T_n \mathbf{U}_m \\ &= \mathbf{X}_n + \sum_{m=1, m\neq n}^{N_s} \mathbf{C}_{nm}, \quad n = 1, \ldots, N_s \end{aligned} \tag{3}$$

where $\mathbf{P}_n = [\mathbf{0}^T_{t_n-1,K},\ \mathbf{I}_K,\ \mathbf{0}^T_{K_T-K-t_n,K}]^T$, $\mathbf{I}_K$ is the identity matrix of size K, and $\mathbf{C}_{nm} \triangleq \mathbf{P}^T_n \mathbf{U}_m$ is the coherent cross-source interference that contaminates $\mathbf{X}_n$ due to overlap with $\mathbf{X}_m$. The overall cross-source interference $\mathbf{C}_n \triangleq \sum_{m=1,m\neq n}^{N_s} \mathbf{C}_{nm}$ occurs when the response of the n-th source overlaps with one or more responses of other sources. It is worth noting that the shorter the overall simultaneous shooting time K_T the larger the cross-source interference and vice versa.

Let the $K \times N_s$ matrices $\tilde{\mathbf{D}}_m,\ m = 1, \ldots, N_r$ be the common receiver data associated with $\mathbf{D}_n,\ n = 1, \ldots, N_s$. The common receiver data matrices can be obtained as follows

$$\tilde{\mathbf{D}}_m = [\mathbf{d}_{m,1},\ \ldots,\ \mathbf{d}_{m,N_s}], \quad m = 1, \ldots, N_r \tag{4}$$

where $\mathbf{d}_{m,n}$ is the m-th column of $\mathbf{D}_n$. Using the principle of superposition, the common receiver data can be represented as

$$\tilde{\mathbf{D}}_m = \tilde{\mathbf{X}}_m + \tilde{\mathbf{C}}_m, \quad m = 1, \ldots, N_r \tag{5}$$

where

$$\tilde{\mathbf{X}}_m = [\mathbf{x}_{m,1},\ \ldots,\ \mathbf{x}_{m,N_s}], \quad m = 1, \ldots, N_r \tag{6}$$

$$\tilde{\mathbf{C}}_m = [\mathbf{c}_{m,1},\ \ldots,\ \mathbf{c}_{m,N_s}], \quad m = 1, \ldots, N_r. \tag{7}$$

In (6) and (7), $\mathbf{x}_{m,n}$ and $\mathbf{c}_{m,n}$ correspond to the m-th columns of $\mathbf{X}_n$ and $\mathbf{C}_n$, respectively.

It is worth noting that $\tilde{\mathbf{C}}_m, m = 1, \ldots, N_r$ represent incoherent[1] noise [4]. Also, it is worth mentioning that for the case when the array of sources and the array of receivers are identical, the reciprocity principle can be applied. In this case, the cross-source interference-free signal in the common source domain translates into itself in the common receiver domain, i.e.,

$$\mathbf{X}_n = \tilde{\mathbf{X}}_n, \quad n = 1, \ldots, N_r. \tag{8}$$

Note that in (8), it is assumed that the number of sources and the number of receivers are equal. In the sequel, we assume that the reciprocity property holds.

Substituting (8) in (5) yields

$$\tilde{\mathbf{D}}_n = \mathbf{X}_n + \tilde{\mathbf{C}}_n, \quad n = 1, \ldots, N_r \tag{9}$$

By blending the common receiver domain data (9), we obtain

$$\mathbf{Y}_b = \sum_{n=1}^{N_r} \mathbf{V}_n \tag{10}$$

where the $K_T \times N_r$ matrix $\mathbf{U}_n$ associated with the n-th source is given by

$$\mathbf{V}_n = \left[\mathbf{0}^T_{t_n-1,N_s},\ \tilde{\mathbf{D}}^T_n,\ \mathbf{0}^T_{K_T-K-t_n,N_s}\right]^T. \tag{11}$$

Substituting (9) in (11), $\mathbf{V}_n$ can be decomposed into two components, that is,

$$\mathbf{V}_n = \mathbf{U}_n + \mathbf{W}_n, \quad n = 1, \ldots, N_r \tag{12}$$

where $\mathbf{W}_n = [\mathbf{0}^T_{t_n-1,N_s},\ \tilde{\mathbf{C}}^T_n,\ \mathbf{0}^T_{K_T-K-t_n,N_s}]^T$. Substituting (12) in (10), we obtain

$$\begin{aligned}\mathbf{Y}_b &= \sum_{n=1}^{N_r} (\mathbf{U}_n + \mathbf{W}_n) \\ &= \mathbf{X}_b + \sum_{n=1}^{N_r} \mathbf{W}_n \\ &= \mathbf{X}_b + \tilde{\mathbf{C}}_b \end{aligned} \tag{13}$$

where $\tilde{\mathbf{C}}_b \triangleq \sum_{n=1}^{N_r} \mathbf{W}_n$ is the blended component associated with the incoherent noise $\tilde{\mathbf{C}}_n$. Note that (1) is used to obtain

[1] Note that the amplitudes of $\mathbf{X}_n$, $\mathbf{C}_n$, and $\tilde{\mathbf{C}}_m$ have the same statistical distribution. However due to the random shuffling of $\tilde{\mathbf{C}}_m$ in the space-time domain, the correlation between $\tilde{\mathbf{C}}_m$ and $\mathbf{X}_n$ is zero $\forall\ m, n$ which is the reason we refer to $\tilde{\mathbf{C}}_m$ as incoherent noise.

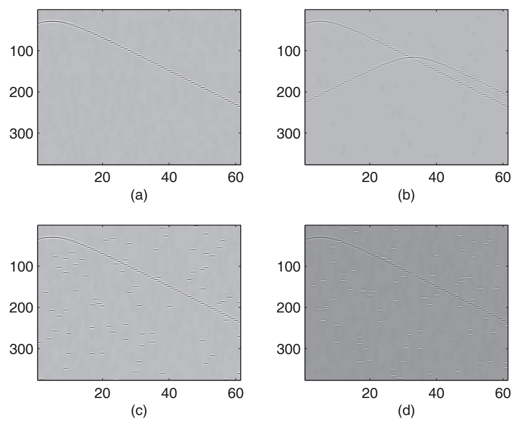

Fig. 1. (a) Top left: Original shot. (b) Top right: Original shot plus coherent interference (3). (c) Bottom left: Original shot plus incoherent interference (5). (d) Bottom right: Original shot plus incoherent interference (15).

the second line in (13) from the first line. It is also important to stress that $\mathbf{X}_b$ corresponds to the original received blended data. Therefore, the blended version of the incoherent noise can be obtained by subtracting $\mathbf{X}_b$ from (13). Re-applying the pseudo-deblending principle to $\tilde{\mathbf{C}}_b$, we obtain the $K \times N_s$ matrices

$$\begin{aligned} \mathbf{F}_n &= \mathbf{P}_n^T \tilde{\mathbf{C}}_b \\ &= \mathbf{P}_n^T \mathbf{W}_n + \sum_{m=1, m\neq n}^{N_r} \mathbf{P}_n^T \mathbf{W}_m \\ &= \tilde{\mathbf{C}}_n + \mathbf{G}_n, \qquad n = 1, \dots, N_r \end{aligned} \tag{14}$$

where $\mathbf{G}_n \triangleq \sum_{m=1,m\neq n}^{N_r} \mathbf{P}_n^T \mathbf{W}_m$ is an incoherent noise component which contaminates $\tilde{\mathbf{C}}_n$. Subtracting (14) from (5), we obtain

$$\begin{aligned} \mathbf{Z}_n &= \tilde{\mathbf{D}}_n - \mathbf{F}_n \\ &= \mathbf{X}_n - \mathbf{G}_n, \qquad n = 1, \dots, N_r. \end{aligned} \tag{15}$$

Unlike (3), which represents the desired signal $\mathbf{X}_n$ contaminated by coherent cross-source interference, both (5) and (15) represent the desired signal contaminated by incoherent noise. This observation is illustrated in Fig. 1. The top left of that figure shows the interference free signal while the top right shows the signal contaminated with coherent cross-source interference. The bottom left shows the corresponding signal contaminated with incoherent interference in the common receiver domain which corresponds to (5). The bottom right shows the signal contaminated with incoherent noise which corresponds to (15).

To attenuate the incoherent noise, we compare (5) and (15) at every point in the time-space domain. If the corresponding data at a certain time-space point are similar, that point is retained as good data and no filtering is required. On the other hand, if the corresponding data at a certain time-space point are dissimilar, then this point is determined to be corrupt with noise and local filtering is required to obtain an estimate of that data. Mathematically, the output data can be expressed as

$$\mathbf{X}_n(t,s) = \begin{cases} \tilde{\mathbf{D}}_n(t,s) & \text{if } \|\tilde{\mathbf{D}}_n(t,s) - \mathbf{Z}_n(t,s)\| \leq \zeta \\ \hat{\mathbf{X}}_n(t,s) & \text{if } \|\tilde{\mathbf{D}}_n(t,s) - \mathbf{Z}_n(t,s)\| > \zeta \end{cases} \tag{16}$$

where $\hat{\mathbf{X}}_n(t,s)$ is an estimate of $\mathbf{X}_n(t,s)$ that can be obtained using a weighted sum (e.g. averaging) of neighboring noise-free data points and ζ is a threshold of user choice. It is worth noting that choosing large ζ means that more points are accepted as noise free and less points are estimated using local filtering and vice versa. A zero-tolerance policy can be adopted by choosing $\zeta = 0$ which means that only actual noiseless points are accepted while all other points go through local filtering. More sophisticated filtering techniques than the moving averaging filter can be used, but even with moving averaging filter the achieved performance improvement due to the introduction of an additional noisy version of the desired data as can be seen from the next section.

4. SIMULATION RESULTS

We consider a synthetic example of a survey system equipped with 57 sources spaced 25 m apart and 57 receivers also spaced 25 m apart. The received data at the output of each receiver is sampled with sampling interval 0.04 sec. For conventional shooting, the sources are activated sequentially where the time required to record the response of each source is 376 samples (approximately 15 sec). Fig. 2 shows the response of the 20-th source in the absence of cross-source interference. We refer to this shot as original shot.

To reduce the overall time of the survey, the responses of different sources are allowed to overlap. The total recording time is reduced by factor 3 and the sequence of the sources and the time instants at which they are fired are selected randomly. Fig. 3 shows the combined output in the common source domain of the 20-th source (corrected to its zero time) and the cross-source interference. It can be observed from this figure that the cross-source interference introduces ambiguity to the graph making it difficult to interpret the data. Fig. 4 shows the same data of Fig. 3 but in the common receiver domain. It can be observed from this figure that the coherent interference in the common source domain translates into incoherent noise in the common receiver domain. Finally, Fig. 5 shows the filtered output of the proposed method. It is clear from the figure that the incoherent noise is effectively removed. Other examples will be given in the journal paper.

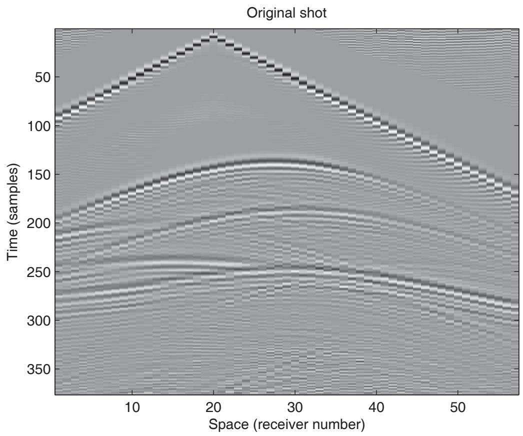

Fig. 2. Original shot.

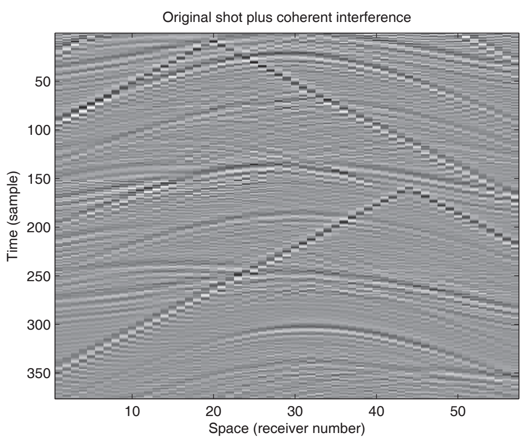

Fig. 3. Pseudo-deblended shot in common source domain.

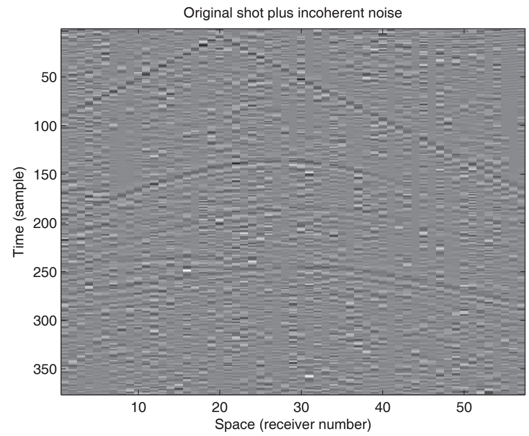

Fig. 4. Pseudo-deblended shot in common receiver domain.

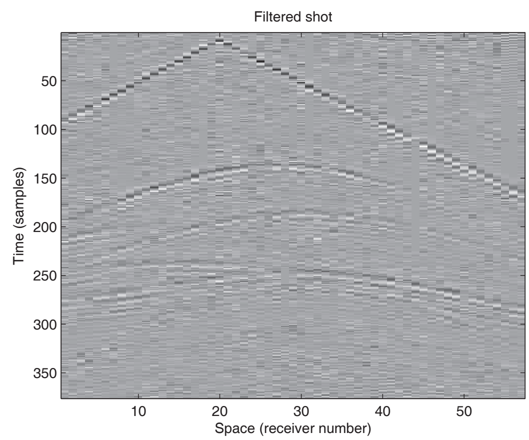

Fig. 5. Filtered shot.

5. CONCLUSIONS

The problem of separating simultaneous source blended data in seismology is considered. Cross-source interference that masks the desired signal in the common source domain has been shown to translate into incoherent noise in the common receiver domain. A virtual blending/deblending process has been applied to the common receiver domain data which enables obtaining an additional noisy version of the data. A method for discriminating between corrupt and non-corrupt data points by measuring the local similarities and dissimilarities between the two noisy versions of the data has been proposed. Corrupt points are replaced with a weighted sum (e.g., averaging) of neighboring non-corrupt data points. The proposed method does not require computationally expensive data transformation and it can be straightforwardly extended to higher-dimensional data scenarios. Simulation results have been given to validate the effectiveness of the proposed method.

6. REFERENCES

[1] G. Hampson, J. Stefani, and F. Herkenhof, "Acquisition using simultaneous sources," *The Leading Edge*, no. 7, pp. 918–923, July 2008.

[2] A. J. Berkhout, "Changing the mindset in seismic data acquisition," *The Leading Edge*, 27, no. 7, pp. 924–938, July 2008.

[3] A. J. Berkhout, G. Blacquiere, and E. Verschuur, "From simultaneous shooting to blended acquisition," in *Proc. 78th Annual Int. Meeting, SEG, Expanded Abstracts*, Las Vegas, NV, Nov. 2008, pp. 2831–2838.

[4] A. Mahdad,P. Doulgeris, and G. Blacquiere, "Separation of blended data by iterative estimation and subtraction of blending interference noise," *Geophysics*, vol. 76, no. 3, pp. Q9–Q17, May-June 2011.

[5] I. Moore, B. Dragoset, T. Ommundsen, D. Wilson, C. Ward, and D. Eke, "Simultaneous source separation using dithered sources," in *Proc. 78th Annual Int. Meeting, SEG, Expanded Abstracts*, Las Vegas, NV, Nov. 2008, pp. 2806-2809.

[6] P. Akerberg, G. Hampson, H. Richett, H. Martin, and J. Cole, Simultaneous source separation by sparse radon transform," in *Proc. 78th Annual Int. Meeting, SEG, Expanded Abstracts*, Las Vegas, NV, Nov. 2008, pp. 2801-2804.