

# Seismic interferometry for sparse data: SVD-enhanced Green's function estimation

Gabriela Melo<sup>1</sup>, Alison Malcolm<sup>2</sup>, and Thomas Gallot<sup>3</sup>

Earth Resources Laboratory - Earth, Atmospheric, and Planetary Sciences Department - MIT

Email: <sup>1</sup>gmelo@mit.edu, <sup>2</sup>amalcolm@mit.edu, <sup>3</sup>gallot@mit.edu

**Abstract**—Seismic interferometry (SI) is a technique used to estimate the Green's function (GF) between two receivers, as if there were a source at one of the receiver locations. The GF obtained in this way, the interferometric GF (IGF), is estimated here by crosscorrelating the signals from two receivers for many sources and averaging these crosscorrelations over sources. However, in many applications, the conditions needed to recover the exact GF are not met and thus the estimated IGF is inaccurate. For such cases, we improve the IGF by summing lower-rank approximations of the crosscorrelations obtained via the Singular Value Decomposition (SVD), instead of averaging the original crosscorrelations. SVD allows us to enhance low-rank, coherent signals; these are the signals needed to reconstruct the GF. We apply this method to a field dataset where seismic signals from active sources are transformed to simulate passive seismic recordings. In this data set we find that filtering with SVD allows for IGF recovery in cases where standard SI does not.

## I. INTRODUCTION

Seismic interferometry (SI), first suggested by [1], can be used to estimate the Green's function (GF) between two receivers, as if there were a source at one of the receiver locations, by crosscorrelating the seismic signals recorded at the two receivers and summing these crosscorrelations over many sources. The sources can be deterministic (e.g., [2]) or uncorrelated noise (e.g., [3]). A recent tutorial on the basic principles, theory, and a few examples can be found in [4].

To accurately construct the IGF requires that the receivers be surrounded by a closed surface of sources. It is generally accepted that the sources located on or near rays that pass through both receivers give the primary contribution to the IGF [5]. As this result is derived by approximating the interferometric integral (or summation in the discrete case) with the method of stationary phase, we call these sources stationary sources, and those outside this zone non-stationary sources. When the source coverage is incomplete, the crosscorrelations from non-stationary sources do not cancel completely resulting in artifacts in the final IGF. [6] provides a summary of methods for alleviating this problem.

Our approach starts with constructing a crosscorrelogram matrix (collection of crosscorrelated signals). The dimensions of this matrix are time-lags (from crosscorrelations) and sources (or time windows in the passive case). Thus, by stacking the crosscorrelogram along the source (time window) dimension, we obtain the IGF. We follow the suggestion in [7], of analyzing and preprocessing the crosscorrelogram before stacking. In the crosscorrelogram, energy from the stationary sources (stationary energy) provide the main contribution to the GF and energy from the non-stationary sources (non-stationary energy) should ideally cancel. Stationary energy is characterized by coherency, small wavenumber, and nearly in-phase events along the source/time window dimension. Non-stationary energy, by contrast, is characterized by incoherency,

larger wavenumber, and out-of-phase events along the source dimension. It is by separating the stationary and non-stationary energy, using SVD, in the crosscorrelogram that we obtain more accurate IGFs for incomplete source distributions. This idea of filtering stationary energy in the crosscorrelograms using SVD is similar to the approach used in [8] to increase the signal to noise ratio (SNR) and filter linear events. A detailed derivation of this method and a few applications to active-source data can be found in [9]. Here we extend the application of this methodology to a passive-source case.

We start by discussing the methodology in the following section. We then show a simple acoustic synthetic example (active sources) to illustrate the SVD method, followed by an application to a field dataset in which we simulate passive data.

## II. METHOD

Here we present the main equations of SI and the SVD method; more details can be found in [9] and references therein. We follow the derivation of SI in [10].

Let  $\mathbf{x}_i$  for  $1 \leq i \leq N$  be the location of sources,  $\omega$  the angular frequency, and  $\hat{G}(\mathbf{x}_A, \mathbf{x}_i, \omega)$  and  $\hat{G}(\mathbf{x}_B, \mathbf{x}_i, \omega)$  the Fourier transforms of the signal for a source located at  $\mathbf{x}_i$  and recorded at receivers  $A$  and  $B$ , respectively. The IGF between the receivers can be obtained by summation of the crosscorrelated signals for all sources

$$\hat{G}(\mathbf{x}_A, \mathbf{x}_B, \omega) + \hat{G}^*(\mathbf{x}_A, \mathbf{x}_B, \omega) \approx \sum_{i=1}^N \hat{G}^*(\mathbf{x}_A, \mathbf{x}_i, \omega) \hat{G}(\mathbf{x}_B, \mathbf{x}_i, \omega), \quad (1)$$

where  $\hat{G}(\mathbf{x}_A, \mathbf{x}_B, \omega)$  is the frequency-domain GF for a receiver at  $\mathbf{x}_A$  and a source at  $\mathbf{x}_B$ ; the  $*$  represents complex conjugation. In the time domain, SI recovers the sum of the causal and acausal GF; here we will focus only on the causal portion of the signal. The absolute amplitudes of the IGF are lost due to simplifications made to obtain equation 1. The phases, however, are preserved, making (1) suitable for most applications of SI.

Now, assume there are  $M$  time samples and let  $\tau_l$  for  $1 \leq l \leq 2M - 1$  be the time lags from crosscorrelations in the time domain. We consider the crosscorrelogram as the matrix  $\mathbf{C} = \mathbf{C}(\mathbf{x}_i, \tau_l)$ , where each row is the crosscorrelation of the two signals (recorded at the two receivers) for each source (for the passive field data case presented below, the sources are replaced by time windows). Thus, the crosscorrelogram  $\mathbf{C}$  can be written as

$$\mathbf{C}(\mathbf{x}_i, \tau_l) = \int G(\mathbf{x}_A, \mathbf{x}_i, t + \tau_l) G(\mathbf{x}_B, \mathbf{x}_i, t) dt, \quad (2)$$

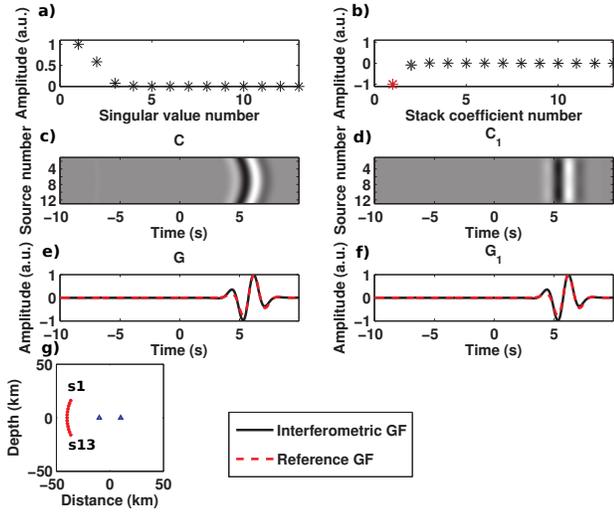


Fig. 1. (a) singular values,  $\sigma_k$ ; (b) stack coefficients,  $s_k$  (the coefficient used to construct  $C_1$  appears in red); (c) original crosscorrelogram,  $C$ ; (d) rank-1 crosscorrelogram,  $C_1$ ; (e) standard IGF,  $G$ ; (f) IGF,  $G_1$ , obtained from  $C_1$ ; (g) source-receiver geometry with 13 evenly distributed sources (red stars) around the stationary zone to the left of the receivers (blue triangles); The GFs in (e) and (f) are similar. Even though (a) shows that there are two significant singular values to represent  $C$ , (b) shows that  $G$  can be well represented with only one stack coefficient.

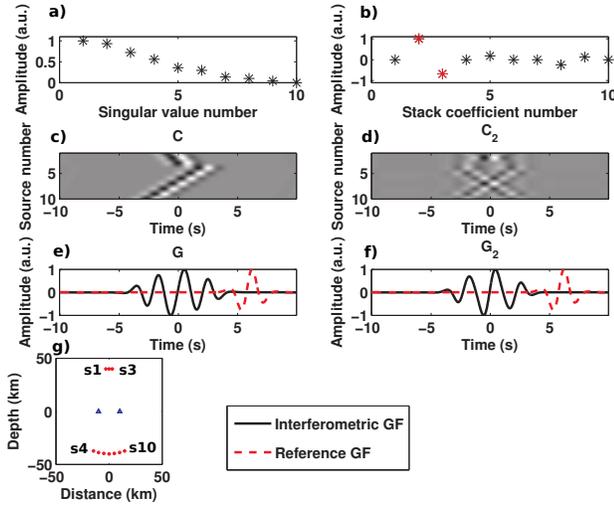


Fig. 2. Figures (a)-(g) are similar to those in Fig. 1. The stack coefficient spectrum indicates two significant coefficients, 2 and 3, contributing to the GF, however, as seen in (f), they do not resemble the GF, as expected.

where  $C$  is an  $N \times 2M - 1$  matrix. The IGF is then obtained by stacking  $C$  over the source dimension,

$$G = G(\mathbf{x}_B, \mathbf{x}_A, t) + G(\mathbf{x}_B, \mathbf{x}_A, -t) = \sum_i C(\mathbf{x}_i, \tau_i). \quad (3)$$

The SVD decomposition of the crosscorrelogram is  $C = U\Sigma V^t$ , where  $U$  and  $V$  are the left and right singular vectors, respectively, and  $\Sigma$  is the diagonal matrix whose elements are the singular values of  $C$ . Now we construct  $\Sigma_j$  by keeping  $j$  singular values of  $\Sigma$  and obtain a lower-rank approximation  $C_j = C_j(\mathbf{x}_i, \tau_i) = U\Sigma_j V^t$ . Stacking the rows of  $C$  gives the standard IGF,  $G$ , and stacking the rows of the approximation

$C_j$  gives the modified IGF,  $G_j$ ,

$$G_j = \sum_i C_j(\mathbf{x}_i, \tau_i). \quad (4)$$

According to the SVD-based crosscorrelogram decomposition, we note that  $G_j$  can be viewed as a weighted sum of the left singular vectors (rows of matrix  $V$ ). Let  $\mathbf{e}$  be a vector of dimensions  $1 \times N$ , whose elements are all equal to 1.

Then, the IGF can be written in matrix notation as

$$G = \mathbf{e}C = \mathbf{e}U\Sigma V^t = \mathbf{s}V^t, \quad (5)$$

where  $\mathbf{s} = \mathbf{e}U\Sigma$  are the coefficients of the weighted sum of the singular vectors in  $V$ . We refer to these coefficients as stack coefficients. Let  $u_{ik}$  correspond to the elements of matrix  $U$ ,  $\mathbf{v}_k$  correspond to the  $k$ -th row of matrix  $V$ , and  $\sigma_k$  be the singular values. Thus, (3) can be rewritten as

$$G = \sum_k \sigma_k \left( \sum_i u_{ik} \right) \mathbf{v}_k = \sum_k s_k \mathbf{v}_k, \quad 1 \leq k \leq N. \quad (6)$$

Next we apply this method to an active-source synthetic dataset and a field dataset with simulated passive sources.

### III. SYNTHETIC ACOUSTIC EXAMPLE

We now use a synthetic acoustic homogeneous model to illustrate the SVD-based method. Consider a medium with no reflectors and with constant velocity and density, where the GF consists of the direct wave only. We study three source distributions: (i) stationary sources only, (ii) non-stationary sources only, and (iii) both stationary and non-stationary sources. In all three cases there are gaps in the source distribution and, for comparison, all the GFs are normalized to have a peak amplitude of one.

First, we consider the case where the sources are only in the stationary-phase zone, as in Fig. 1(g). The energy from these sources contributes constructively to the GF. The spectra in Fig. 1(a)-(b) show that while there are two significant singular values to represent  $C$  (Fig. 1(c)) only one stack coefficient, the first one, should be required to well-approximate the GF. Fig. 1(d) shows  $C_1$  constructed using only the first stack coefficient. Fig. 1(e) and (f), show that the GF obtained from  $C$  and  $C_1$  are quite similar. This is a case where standard interferometry works well and the SVD technique is not necessary, although it is not detrimental.

In case (ii) we take only non-stationary sources (Fig. 2(g)). Ideally, (i.e., assuming full source coverage) all of the non-stationary energy should cancel during the summation over sources. However, if there are gaps in the source distribution, residual energy will remain because of the imperfect cancellation of the non-stationary energy. Even though here we cannot recover the GF, we present this case with the goal of observing how the crosscorrelogram (as well as singular value and stack coefficient spectrum) differs from case (i). First we note that the crosscorrelogram in Fig. 2(c) does not contain the nearly-in-phase energy present in Fig. 1(c). The singular value spectrum in Fig. 2(a) shows a smooth decay, i.e., there is no obvious truncation point of significant singular values as seen in Fig. 1(a). We construct a rank-2 crosscorrelogram approximation (from the strongest stack coefficient, 2 and 3),

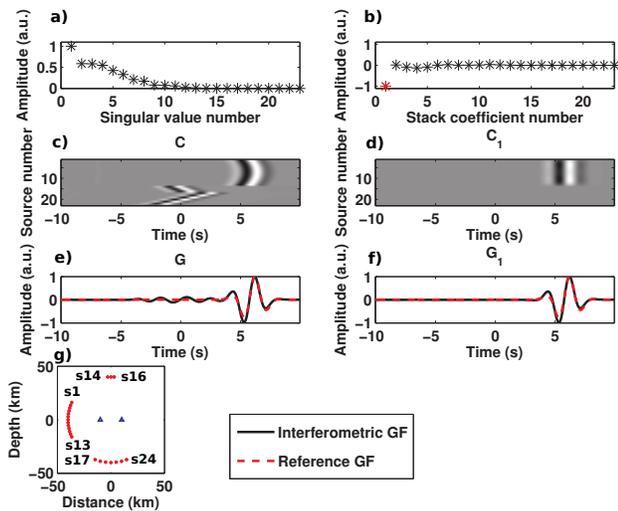


Fig. 3. Figures (a)-(f) are similar to those in Fig. 1. In (b) it is clear that the IGF is well represented by the first stack coefficient. In (f) the fluctuations are reduced and the GF is clearer than in (e).

$C_2$ , in Fig. 2(d). As expected,  $C_2$  does not enhance any linearity and does not even resemble  $C$ , Fig. 2(c).

Case (iii) mixes the two previous cases. Fig. 3(g) shows sources in the stationary and non-stationary zones, but with gaps in the source coverage. The crosscorrelogram, Fig. 3(c), thus has energy contributing to the GF and energy that should cancel out completely; however, because of the gaps, it does not. In the singular value spectrum, Fig. 3(a), we observe a mixture of the two previous cases, a break after the first singular value followed by smooth decay. Fig. 3(b) again indicates that the IGF is well represented with only the first singular vector. Fig. 3(d) shows  $C_1$ , constructed using only the first stacking coefficient, which corresponds to the stationary energy. This rank-1 approximation thus suppresses the residual energy caused by the imperfect cancellation of non-stationary energy, and  $G_1$  is more accurate than  $G$  as seen in Fig 3(e)-(f).

Even though in this case we used deterministic sources, similar considerations regarding stationary and non-stationary energy can be made for the passive case, as we show in the next example.

#### IV. FIELD DATASET

We now apply the SVD technique to a portion of a high-resolution field dataset acquired in the Middle East. The full data were obtained in a  $1 \times 1$ -km area covered with 1600 geophones located on a  $25 \times 25$ -m cell grid. Seismic vibrators were located in a similar grid shifted with respect to the receiver grid by half a cell (12.5 m) in both directions. The subset of sources and receivers (448 sources and two receivers) used in this work is shown in Fig. 4.

In the field, waves from each source were recorded separately at each receiver for 4 s; recorded signals have a bandwidth of 725 Hz, and the signals are dominated by surface waves. In order to simulate a passive seismic experiment for each receiver, we first generate random activation times for each source and then create one single long-time signal that contains all 4 s responses of each individual source, similar to

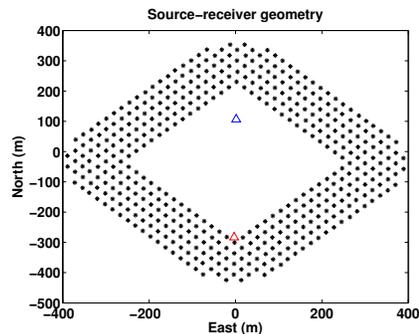


Fig. 4. Stars and triangles represent source and receivers locations, respectively. The virtual source corresponds to the red triangle.

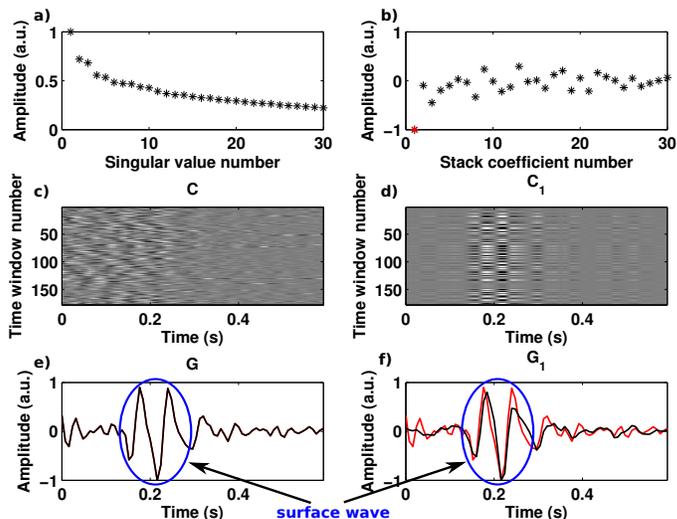


Fig. 5. Figures (a)-(f) are similar to those in Fig. 1. In (f), the red line corresponds to the reference GF (obtained using 175 time windows) and the black line to the IGF. In (b) we see that  $G$  can be well represented with only one stack coefficient.

what was done in [11]. The activation time of each source is taken from a uniform distribution between zero and 1300 s, leading to signal superpositions and a minimum delay of 0.2 s between consecutive sources. In this way, noise records are simulated and temporal and spatial information about sources is lost.

We now proceed to obtain IGFs for the pair of receivers highlighted in Fig. 4. The goal is to construct the IGFs using different lengths of the noise signal (from 100% down to about 28% of the signal), using both standard SI and the SVD method, to show that the IGF can be obtained via the SVD method in cases where standard SI fails to converge. First, the signal at each receiver is separated into 8 s long time windows, with a total of 175 windows.

Next, we generate a reference GF (that will be used for keeping track of the convergence of both standard SI and the SVD method), by estimating the IGF between the receivers using all 175 windows. The reference GF is shown in Fig. 5(e). The singular value spectra, Fig. 5(a), shows a smooth decay except for the small jump from the first to the second singular value. The stack coefficient spectra, Fig. 5(b), shows one most significant coefficient, the first one. The rank-1 cross-correlo-

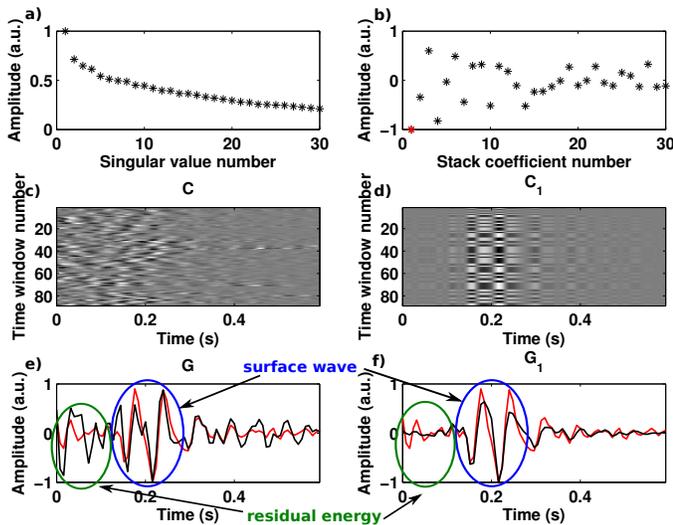


Fig. 6. Figures (a)-(f) are similar to those in Fig. 1. Here we use 88 time windows to construct the IGFs. In (e) and (f), the red line corresponds to the reference GF and the black line to the IGF. The strong residual energy (due to uncanceled noise) makes the convergence of  $G$  be very poor. In particular, in this case it would be very difficult to identify the surface wave. Here, even though (b) shows that the energy in  $G$  is spread across many stack coefficients, it still can be reasonably represented with only one.

gram, Fig. 5(d), shows the zone of strongest coherent energy (in the vicinity of 0.2 s) in the original crosscorrelogram in Fig. 5(c). This energy corresponds primarily to the surface wave, as can also be seen in Fig. 5(f).

Now, we decrease the number of time windows used to construct the IGFs. As the number of time windows decreases, the GF obtained with SVD remains stable while the GF obtained through standard SI deteriorates. Fig. 6 shows results after we decrease the number of time windows by 50%. As seen in Fig. 6(e)-(f), the convergence of the standard IGF is poor while the GF from the SVD method remains stable. We then further reduce the number of time windows by approximately 43%. The results are shown in Fig. 7. Again, while  $G$  in Fig. 7(e) converges poorly,  $G_1$  remains stable. This illustrates how the SVD method may require fewer sources to converge to the IGFs as compared to standard SI. Note that crosscorrelograms from passive data should present linear features (unlikely crosscorrelograms from deterministic data which has a curved structure as seen, for example, in Fig.1) indicating that the first singular vector should contain most of the relevant information. Thus, for this field dataset, the IGFs were constructed from rank-1 crosscorrelograms obtained by preserving the first singular value only.

## V. CONCLUSION

We have shown how using lower-rank approximations to crosscorrelograms, obtained through SVD, is a promising approach to improve IGF estimates. The SVD approach preserves stationary energy in the crosscorrelogram, which is the energy that contributes most to GF recovery, and helps to attenuate the non-stationary energy, which contributes primarily to artifacts in the IGF. We presented a synthetic and a field data example where this technique allows the recovery of IGFs in situations where standard SI fails.

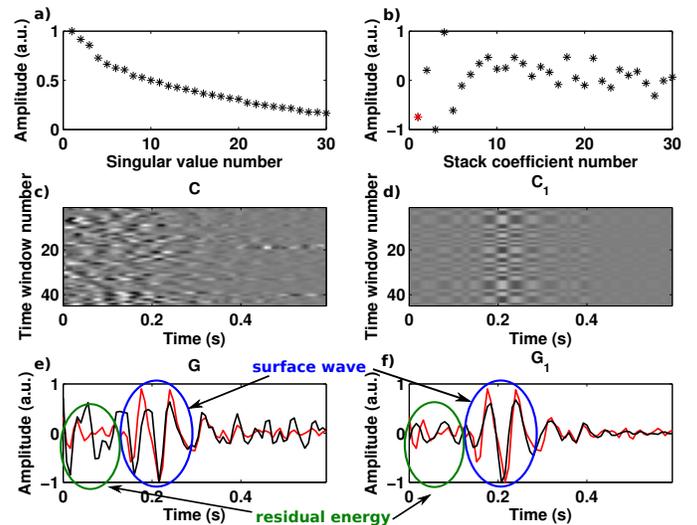


Fig. 7. Figures (a)-(f) are similar to those in Fig. 1. In (e) and (f), the red line corresponds to the reference GF and the black line to the IGF. Here we use 50 time windows. The convergence of  $G$  is again very poor (also making it difficult to identify the surface wave) and its energy is also spread across many stack coefficients, but it can still be reasonably represented with only the first coefficient.

## ACKNOWLEDGMENT

This work is supported by grants from the Department of Energy (DOE), the National Science Foundation (grant number 1115406), and the founding members consortium at Earth Resources Laboratory (ERL).

## REFERENCES

- [1] J[] F[] Claerbout, "Synthesis of a layered medium from its acoustic transmission response," *Geophysics*, vol. 33, pp. 264–269, 1968.
- [2] T[] G[] Schuster, J[] Yu, J[] Sheng, and J[] Rickett, "Interferometric/daylight seismic imaging," *Geophysics Journal International*, vol. 157, pp. 838–852, 2004.
- [3] R[] L[] Weaver, "Information from seismic noise," *Science*, vol. 307, pp. 1568–1569, 2005.
- [4] K. Wapenaar, D. Draganov, R. Snieder, X. Campman, and A. Verdel, "Tutorial on seismic interferometry: Part 1 Basic principles and applications," *GEOPHYSICS*, vol. 75, no. 5, pp. 75A195–75A209, 2010.
- [5] R[] Snieder, "Extracting the Green's function from correlation of coda waves: A derivation based on stationary phase," *Phys. Rev. E*, vol. 69, pp. 046610, 2004.
- [6] Simon King and Andrew Curtis, "Suppressing nonphysical reflections in Green's function estimates using source-receiver interferometry," *Geophysics*, vol. 77, no. 1, pp. Q15–Q25, 2012.
- [7] Oleg V. Poliannikov and Mark E. Willis, "Interferometric correlogram-space analysis," *Geophysics*, vol. 76, no. 1, pp. SA9–SA17, 2011.
- [8] Tadeusz J. Ulrych, Mauricio D. Sacchi, and J. Michael Graul, "Signal and noise separation: Art and science," *Geophysics*, vol. 64, no. 5, pp. 1648–1656, 1999.
- [9] Gabriela Melo, Alison Malcolm, Dylan Mikesell, and Kasper van Wijk, "Using SVD for improved interferometric Green's function retrieval," *Geophysical Journal International*, 2013.
- [10] K[] Wapenaar and J[] Fokkema, "Green's function representations for seismic interferometry," *Geophysics*, vol. 71, no. 4, pp. S133–S146, 2006.
- [11] Thomas Gallot, Stefan Catheline, Philippe Roux, and Michel Campillo, "A passive inverse filter for Green's function retrieval," *The Journal of the Acoustical Society of America*, vol. 131, no. 1, pp. EL21–EL27, 2012.