BAYESIAN SPARSE SENSING OF THE JAPANESE 2011 EARTHQUAKE

Peter Gerstoft[1]*, Christoph F. Mecklenbräuker*[2]*, Huajian Yao*[3]

[1] Scripps Institution of Oceanography, La Jolla, CA 92093-0238, USA,
[2] Institute of Telecommunications, Vienna University of Technology, 1040 Vienna, Austria,
[3] Earth and Space Sciences, University of Science and Technology of China, Hefei, China

ABSTRACT

Sparse sensing is a technique for finding sparse signal representations to underdetermined linear measurement equations. We use sparse sensing to locate seismic sources during the rupture of the 2011 Mw9.0 earthquake in Japan from teleseismic P waves recorded by a seismic sensor array of stations in the United States. The location estimates of the seismic sources are obtained by minimizing the square of ℓ_2-norm of the difference between the observed and modeled waveforms penalized by the ℓ_1-norm of the seismic source vector. The resulting minimization problem is convex and can be solved efficiently using LASSO type optimization. The potential to track the rupture sequentially is demonstrated.

1. INTRODUCTION

Fortunately, large earthquake are sparse in time and location. However, large earthquakes do evolve in both time and space, often lasting 100s of seconds and rupturing over 100s of kilometer[1]. For subduction earthquakes, there is evidence that radiation of frequencies depends on distance to the plate interface [2, 3, 4], with the higher frequencies being further away from the trench and deeper below the surface.

The earthquake rupture process is complicated, but it tends to evolve sequentially with sparse locations that typically propagate at a speed of approximately 80% of the shear speed. In this contribution, we motivate a sparse sequential model and carry out a simple test to demonstrate its utility.

Let $\boldsymbol{y}_t$ be the measurement vector (i.e., seismic recording at the sensor array) at step t and $\boldsymbol{x}_t$ represents the state vector (for example, location of a source emitting signal $\boldsymbol{y}_t$), where $t = 1, \ldots, T$. Our goal is to sequentially estimate $\boldsymbol{x}_t$ as data measurements $\boldsymbol{y}_t$ become available.

The *state equation* (1) and *measurement equation* (2) define a state-space model:

$$\boldsymbol{x}_t = \mathbf{f}_t(\boldsymbol{x}_{t-1}, \boldsymbol{v}_t) \tag{1}$$

$$\boldsymbol{y}_t = \mathbf{h}_t(\boldsymbol{x}_t, \boldsymbol{n}_t) \tag{2}$$

The *state equation*, Eq. (1), describes the evolution or transition of $\boldsymbol{x}_t$ with t and assumes that states follow a first order Markov process.

Function $\mathbf{f}_t$ and $\mathbf{h}_t$ are known functions. Variable $\boldsymbol{v}_k$ is the process or state noise and $\boldsymbol{w}_k$ is the measurement noise.

Such system has been solved using sequential filtering, e.g., a Kalman type or a Particle filtering approach [5, 6]. These often lead to a *wide* posterior probability distribution of the state vector estimate $\hat{\boldsymbol{x}}_t$. We here utilize that the state vector $\boldsymbol{x}_t$ is expected to be sparse. The sparseness is enforced by using a Laplace-like prior model.

The sequential sparse sampling approach has received much attention recently [7, 8, 9]. For each time step t such approaches require solving a LASSO-type optimization problem [10, 11, 12, 13].

2. SEISMIC RECORDING MODEL

Let $(\theta_1, \ldots, \theta_M)^{\mathrm{T}}$ be a vector of M *potential* earthquake source locations on a suitably chosen grid on the earth's surface. Further, let $\boldsymbol{x}_t(\omega) = (x_{1t}(\omega), \ldots, x_{Mt}(\omega))^{\mathrm{T}}$ be the associated complex-valued source vector at time t and frequency ω. We observe time-sampled seismic waveforms on an array of N sensors which are stacked in a vector. After a short-time Fourier transform, we obtain the following linear model which relates the Fourier transformed sensor array output $\boldsymbol{y}_t(\omega)$ to the source vector,

$$\boldsymbol{y}_t(\omega) = \boldsymbol{A}(\omega)\boldsymbol{x}_t(\omega) + \boldsymbol{n}_t(\omega)\ . \tag{3}$$

The mth column of the transfer matrix $\boldsymbol{A}(\omega)$ is the steering vector $\boldsymbol{a}_m(\omega)$ for potential source location θ_m. We model the (n, m)-element of $\boldsymbol{A}(\omega)$ by $\mathrm{e}^{-\mathrm{j}\omega\tau_{nm}}$. Here τ_{nm} is the traveltime from location θ_m to the nth seismic sensor,

$$\tau_{nm} = \int_{\mathcal{C}_{nm}} \frac{\mathrm{d}s}{c_{\mathrm{p}}(s)}\ , \tag{4}$$

where we integrate along the curved ray $\mathcal{C}_{nm}$ which connects location θ_m to sensor n, and $c_{\mathrm{p}}(s)$ is the phase velocity of compressional ("P") waves along the ray in the earth's mantle. Travel times are calculated from the one-dimensional IASP91 earth reference model [14]. The additive noise vector $\boldsymbol{n}_t(\omega)$ is assumed to be spatially uncorrelated and follows the zero-mean complex normal distribution with diagonal covariance $\nu(\omega)\boldsymbol{I}$. Further, the noise is assumed to be uncorrelated across frequency.

3. BAYESIAN FORMULATION

First, we select a set of frequencies $\{\omega_1, \ldots, \omega_J\}$ at which we observe measurement data $\boldsymbol{y}_t(\omega_j)$ according to the model (3). We stack all these measurement data into an NJ-dimensional multi-frequency observation $\boldsymbol{y}_t$ and likewise for the source $\boldsymbol{x}_t$ and noise $\boldsymbol{n}_t$. For the linear model (3), we arrive at the following conditional probability density for the stacked multi-frequency observations given the multi-frequency source vector $\boldsymbol{x}_t$,

$$p(\boldsymbol{y}_t|\boldsymbol{x}_t) = \prod_{j=1}^{J} \frac{\exp\left(-\frac{1}{\nu_j}\|\boldsymbol{y}_t(\omega_j) - \boldsymbol{A}(\omega_j)\boldsymbol{x}_t(\omega_j)\|_2^2\right)}{(\pi\nu_j)^N} \tag{5}$$

where we have used the shorthand notation $\nu_j = \nu(\omega_j)$ and $\|\cdot\|_p$ is the p-norm. In our setting, the number of candidate source locations M is larger than the number of sensors N, i.e. $N < M$ and the linear model (3) are underdetermined. To reconstruct a physically meaningful source vector, we exploit its sparsity through the introduction of a prior probability density on $\boldsymbol{x}$ which promotes sparsity. A widely used sparseness prior is the Laplace probability density [15, 16] which puts a higher probability mass than the normal distribution both for *large* and *small* absolute deviations . We assume the multivariate complex Laplace-like probability density [17],

$$p(\boldsymbol{x}_t) = \prod_{j=1}^{J} \left(\frac{\lambda_{jt}}{\sqrt{2\pi}}\right)^{2M} \exp\left(-\lambda_{jt}\|\boldsymbol{x}_t(\omega_j)\|_1\right) . \tag{6}$$

The parameters λ_j are positive scale parameters which are suitably chosen later. For the posterior probability density function $\Lambda_t = p(\boldsymbol{x}_t|\boldsymbol{Y}_t)$ base on the history of all multi-frequency observations $\boldsymbol{Y}_t = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_t]$, we follow Bayes' rule and apply the logarithm. We arrive at the following cost function to be minimized,

$$\begin{aligned}\Lambda_t = & -\ln p(\boldsymbol{x}_t|\boldsymbol{Y}_t) = -\ln p(\boldsymbol{Y}_t|\boldsymbol{x}_t) - \ln p(\boldsymbol{x}_t) + \ln p(\boldsymbol{Y}_t) \\ = & \sum_{j=1}^{J}\left(\frac{\|\boldsymbol{y}_t(\omega_j) - \boldsymbol{A}(\omega_j)\boldsymbol{x}_t(\omega_j)\|_2^2}{\nu_j} + \lambda_{jt}\|\boldsymbol{x}_t(\omega_j)\|_1\right. \\ & \left. + N\ln\nu_j - 2M\ln\lambda_{jt}\right) + \text{const.}\end{aligned} \tag{7}$$

In (7) the term $\ln p(\boldsymbol{Y}_t)$ is absorbed into the additive constant which will be neglected in the following.

4. NON-SEQUENTIAL COMPRESSIVE SENSING

In this section, we neglect the sequential time dependence and formulate the sparse recovery problem as follows: We observe waveforms on an array of N seismic stations in the frequency domain,

$$\boldsymbol{y}(\omega) = \boldsymbol{A}(\omega)\boldsymbol{x}(\omega) + \boldsymbol{n}(\omega) . \tag{8}$$

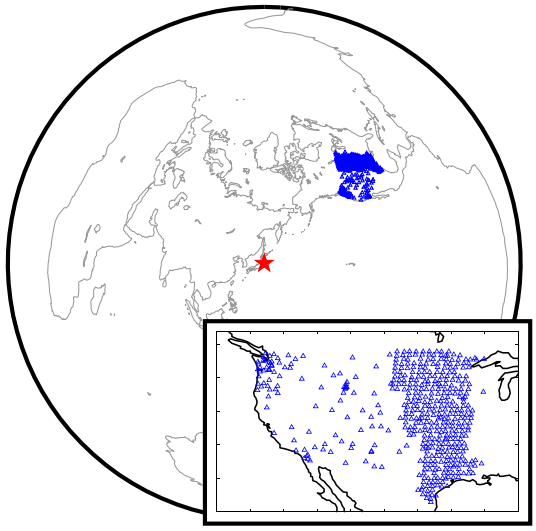

Fig. 1. Location of the 2011 Tohoku-Oki earhthquake (red star $\star$) and array geometry of 409 seismic stations in the central and western US [2].

The number of potential source locations M is substantially larger than the number of seismic stations N, i.e. $N \ll M$. The additive Gaussian noise $\boldsymbol{n}(\omega)$ is zero-mean and circular.

The linear system (8) is underdetermined. Therefore there is a large-dimensional subspace of potential source vectors $\boldsymbol{x}(\omega)$ that result in the same observed waveform $\boldsymbol{y}(\omega)$. To recover a physically meaningful source vector, we exploit our knowledge of its sparsity: We restrict the source vector to have only K non-zero ("active") entries and $K \ll M$. The recovery of such sparse $\boldsymbol{x}(\omega)$ can be formulated as

$$\hat{\boldsymbol{x}}(\omega) = \arg\min \|\boldsymbol{x}\|_0 \quad \text{s. t.} \quad \|\boldsymbol{A}(\omega)\boldsymbol{x} - \boldsymbol{y}(\omega)\|_2 < \epsilon , \tag{9}$$

where $\|\boldsymbol{x}\|_0$ counts the number of nonzero entries in $\boldsymbol{x}$. The additive Gaussian noise motivates an ℓ_2-norm interpretation of the constraint with ϵ being the noise floor. Unfortunately, this problem is hard to solve and unstable in the presence of noise [18].

Practical recovery algorithms rely on the restricted isometry property (RIP). The RIP conditions are satisfied due to the pseudo-randomly varying phases $\mathrm{e}^{-\mathrm{j}\omega\tau_{nm}}$ [19]. Following [11], we replace (9) by the following convex problem

$$\hat{\boldsymbol{x}}(\omega) = \arg\min \|\boldsymbol{x}\|_1 \quad \text{s. t.} \quad \|\boldsymbol{A}(\omega)\boldsymbol{x} - \boldsymbol{y}(\omega)\|_2 < \epsilon . \tag{10}$$

Finally, the Lagrange multiplier λ^{-1} is introduced and the following second-order cone problem is obtained

$$\begin{aligned}\hat{\boldsymbol{x}}(\omega) &= \arg\min\left(\|\boldsymbol{y}(\omega) - \boldsymbol{A}(\omega)\boldsymbol{x}\|_2 - \epsilon + \lambda\|\boldsymbol{x}\|_1\right) \\ &= \arg\min\left(\|\boldsymbol{y}(\omega) - \boldsymbol{A}(\omega)\boldsymbol{x}\|_2 + \lambda\|\boldsymbol{x}\|_1\right) .\end{aligned} \tag{11}$$

We use the package CVX [20, 21] for specifying and solving this convex optimization problem.

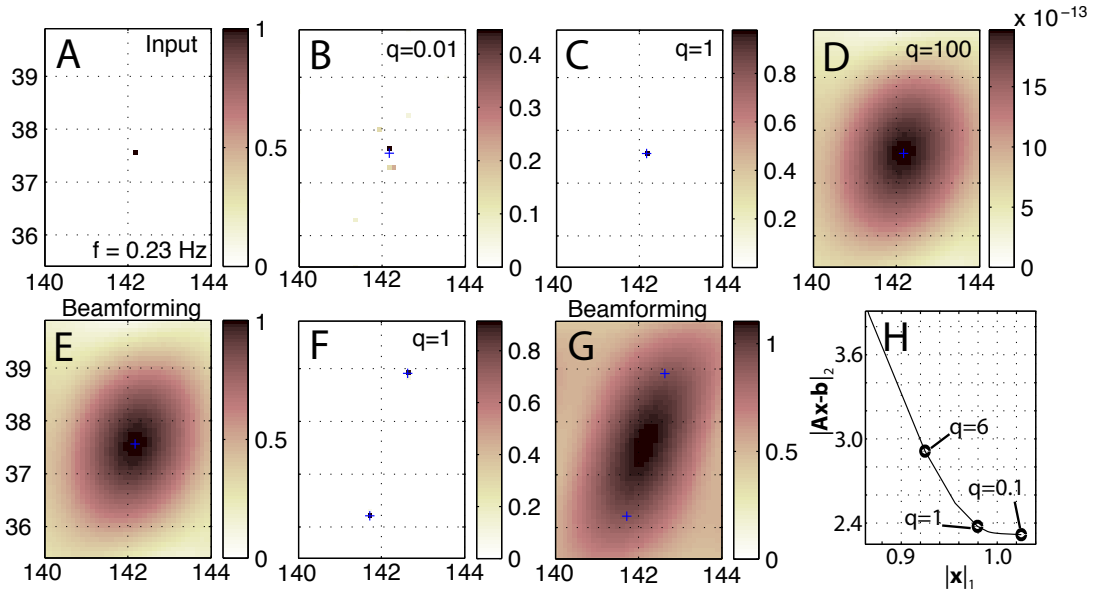

Fig. 2. Synthetic tests of source location using compressive sensing (CS) and beamforming at $f = 0.23$ Hz: (A) input amplitudes of one source model; (B-D) recovery of (A) using CS with different damping parameters ($\lambda = q\lambda_0$, with $\lambda_0 = r\sqrt{N}$ and $r = 0.1$) as indicated (see text for details); (E) beamforming output for the one source in (A); (F) CS recovery of the two sources with locations at the blue crosses; (G) beamforming output for the two-source model in (F); (H) L-curve between data misfit and model norm for one source model in (A) using different damping values.

The choice of Lagrange multiplier λ is important to obtain physically meaningful solutions to (11), Using (8) with just a single source without any noise gives $\|\boldsymbol{y}\|_2 = \|\mathbf{A}\boldsymbol{x}\|_2 = \|\mathbf{A}\|_2\|\boldsymbol{x}\|_2 = \sqrt{N}\|\mathbf{x}\|_1$. We define the residual $r = \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2/\|\boldsymbol{y}\|_2$, which depends on the noise level. This indicates the following choice of Langrange multiplier $\lambda \sim r\sqrt{N}$ for balancing the ratio between data misfit ($\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2$) and model constraint ($\|\boldsymbol{x}\|_1$) in the misfit function (11). Using the array in Fig. 3, Fig. 2 shows synthetic tests of the performance for different λ values at $f = 0.23$ Hz with a randomly generated source amplitude and phase. We add 10% random noise ($r = 0.1$) to the synthetic spectrum data. We use $\lambda = r\sqrt{N}$ as it is around the "knee" in the L-curve in Fig. 2H between the data misfit and model constraints.

First, we apply conventional beamforming to estimate the location of a single hypothetical source. For one source (same source spectrum as in Fig. 2A), the beamformer peak corresponds to the input source location (Fig. 2E), but with poor resolution. In contrast, CS picks the true source location if proper damping is used (Fig. 2C). For the synthetic model with two sources (Fig. 2F; 10% noise), the conventional beamformer (Fig. 2G) cannot resolve the sources. However, CS (Fig. 2F) recovers the source locations exactly.

5. RECONSTRUCTION PERFORMANCE FOR SEISMIC ARRAY

Numerical simulations were carried out with synthetically generated data for the geometry of the USArray [22] with $N = 409$ wideband seismic sensor stations. The USArray geometry is shown in Fig. 3. It is obvious that the geometry is two-dimensional and highly irregular. Instead of a single point source (as indicated by the red star $\star$ at the center of Fig. 3), we assume a sparse source distribution. From the simulation results, we numerically estimate the Root Mean Square (RMS) location error for the source vector.

We create a regular grid of 40×40 *potential* source locations covering the latitudes from $35°$N to $40°$N and longitudes from $140°$E to $145°$E. The chosen grid size results in $M = 1600$. We synthetically generated a sparse source vector $\boldsymbol{X}^\star$ with $K = 3$ non-zero ("active") elements which represent three simultaneous seismic sources. The associated locations $\boldsymbol{r}_1^*, \boldsymbol{r}_2^*, \boldsymbol{r}_3^*$ are evolving in time. At the initial time $t = 0$ they start at the center location (o) and the three sources are propagating north, stationary, and south. The seismic sources have equal power of 1 and uniformly distributed random phases. The remaining $M - K = 1597$ elements are set to zero. The traveltimes τ_{mn} which enter into $\boldsymbol{A}$ are calculated according to (4), cf. [14, 2].

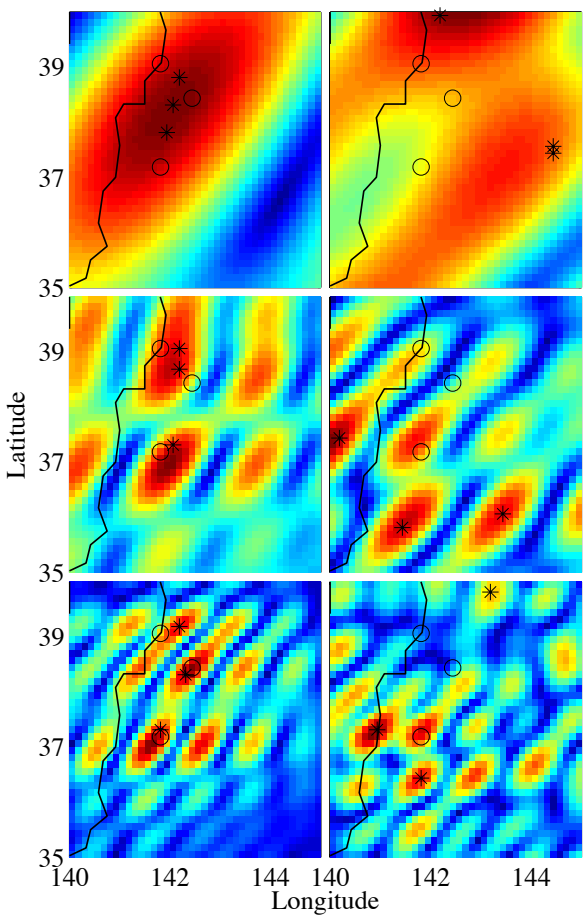

Fig. 3. Beampattern and sparse locations (*) for 3 equal power sources (o) at a SNR of 40 (left) and 0 (right) dB and frequency 0.16 (top) 0.55 (middle) 0.94 (bottom) Hz. The beampattern are just based on the current data, where as the sparse location use sequential processing based on a rupture propagating from the middle source.

We recursively estimate source location using the approach similar to Panahi and Viberg [9]. The resulting location estimate at $t = 5$ is shown in Fig. 3 for 3 frequencies 0.16, 0.55, and $0.94\,\text{Hz}$ and SNR of 40 and $0\,\text{dB}$. At low frequencies the conventional beamformer is unable to resolve the 3 sources. The horizontal wavelength is here approx. $11\,\text{km/s}/0.11 = 100\,\text{km}$ or 1°, indicating the order of magnitude of the resolution. The direction to the array is to the north-east. For the low SNR cases, the 3 estimated sources are often not at the peak in the beamformer output, and the beamformer output peaks do not correspond to the input locations. For low SNR the locations are very biased.

Figure 4 shows the estimated RMS location error from a

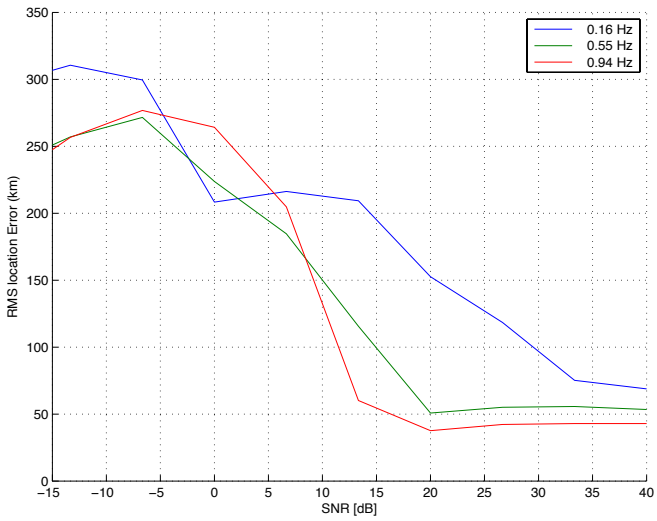

Fig. 4. RMS location error versus SNR for selected frequencies of interest. The USArray [22] has significantly lower RMS location error for higher frequencies than for lower ones.

series of Monte Carlo simulation runs,

$$\text{RMS location error} = \left(\mathrm{E} \sum_{m \in \mathcal{M}_t} \|\boldsymbol{r}_m - \boldsymbol{r}_m^\star\|_2^2 \right)^{\frac{1}{2}} , \quad (12)$$

versus Signal-to-Noise Ratio (SNR) in dB,

$$\text{SNR} = -10 \log_{10} \nu \, , \quad (13)$$

where $\mathcal{M}_t$ is the *active set* at time step t,

$$\mathcal{M}_t = \{m | m\text{th element of } \boldsymbol{x}_t \neq 0\} \, . \quad (14)$$

From the results shown in Fig. 4, we conclude that the RMS location error for sparse source distributions which emit high frequency P waves can be reconstructed more accurately than low frequency sources. An SNR of approximately 10 dB is required to estimate the sparse source locations satisfactorily.

6. CONCLUSION

We formulated the earthquake source location as sparse sequential estimation problem. Numerical simulations with synthetic data indicate that sparse source reconstruction is applicable to seismic array data. Source distributions which emit high frequency P waves can be reconstructed more precisely than low frequency sources.

7. REFERENCES

[1] C.J. Ammon, et al. (2005) Rupture process of the 2004 Sumatra-Andaman earthquake. Science 308:11331139 doi:10.1126/science.1112260.

[2] H. Yao, P. Gerstoft, P.M. Shearer, C.F. Mecklenbräuker: Compressive sensing of Tohoku-Oki M9.0 Earthquake: Frequency-dependent Rupture Modes, Geophys. Res. Lett., 38, L20310, doi:10.1029/2011GL049223.

[3] K.D. Koper AR. Hutko, T. Lay, C.J. Ammon, H Kanamori (2011) Frequency- dependent rupture process of the 2011 M(w) 9.0 Tohoku Earthquake: comparison of short-period P wave backprojection images and broadband seismic rupture models. Earth Planets Space 63:599602.

[4] L. Meng, A. Inbal, and J.-P. Ampuero (2011), A window into the complexity of the dynamic rupture of the 2011 Mw 9 Tohoku-Oki earthquake. Geophys. Res. Lett., doi:10.1029/2011GL048118.

[5] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Boston, MA: Artech House, 2004.

[6] C. Yardim, Zoi-Heleni Michalopoulou, and P. Gerstoft (2011), An overview of sequential Bayesian filtering in ocean acoustics, IEEE Oceanic Eng 36(1), 73-91, DOI:10.1109/JOE.2010.2098810.

[7] D. Angelosante, J. A. Bazerque, G. B. Giannakis: Online Adaptive Estimation of Sparse Signals: Where RLS Meets the ℓ_1-Norm, IEEE Trans. Sig. Proc., vol. 58, no. 7, pp. 3436–3447, DOI:10.1109/TSP.2010.2046897, 2010.

[8] D. M. Malioutov, S. Sanghavi, A. S. Willsky: Sequential Compressed Sensing, IEEE J. Sel. Top. Sig. Proc, vol. 4, pp. 435–444, DOI:10.1109/JSTSP.2009.203821, 2010.

[9] A. Panahi, M. Viberg: Fast LASSO Based DOA Tracking, in Proc. IEEE CAMSAP 2011, San Juan, Puerto Rico, Dec. 2011.

[10] R. Tibshirani: Regression Shrinkage and Selection via the Lasso, J. R. Statist. Soc. B, vol. 58, no. 1, pp. 267–288, 1996.

[11] D. M. Malioutov, C Müjdat, A. S. Willsky: A sparse signal reconstruction perspective for source localization with sensor arrays,

[12] A. Panahi, M. Viberg: Fast Candidate Points Selection in the LASSO Path, IEEE Sig. Proc. Lett., Vol. 19, No. 2, pp. 79–82, Feb. 2012.

[13] C. F. Mecklenbräuker, P. Gerstoft and H, Yao, Bayesian Sparse Wideband Source Reconstruction of Japanese 2011 Earthquake, in Proc. IEEE CAMSAP 2011, San Juan, Puerto Rico, Dec. 2011.

[14] B. L. N. Kennett, E. R. Engdahl: Traveltimes for global earthquake location and phase identification, Geophysical Journal International, vol. 105, no. 2, pp. 429–465, May 1991.

[15] S. Ji, Y. Xue, L. Carin, Bayesian Compressive Sensing, IEEE Trans. Sig. Proc., vol. 56, no. 6, pp. 2346–2356, Jun. 2008.

[16] J. M. Bernardo and A. F. M. Smith, Bayesian Theory. New York: Wiley, 1994.

[17] Z. He, S. Xie, S. Ding, A. Cichocki: Convolutive Blind Source Separation in the Frequency Domain Based on Sparse Representation, IEEE Trans. Audio, Speech, and Language Proc., vol. 15, no. 5, Jul. 2007.

[18] R. G. Baraniuk, Compressive sensing, IEEE Signal Processing Mag., vol. 24, no. 4, pp. 118120, Jul. 2007, 124.

[19] D. Donoho: Compressed sensing, IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[20] S. P. Boyd, L. Vandenberghe, *Convex optimization*, Chapters 1–7 (Cambridge University Press, 2004).

[21] CVX Research, Inc. CVX: Matlab software for disciplined convex programming, version 2.0 beta, http://cvxr.com/cvx, Sep. 2012.

[22] USArray `http://www.usarray.org/`