HEARING BEHIND WALLS: LOCALIZING SOURCES IN THE ROOM NEXT DOOR WITH COSPARSITY

S. Kitić[†], *N. Bertin*[⋆], *R. Gribonval*[†]

[†]Inria , [⋆]IRISA - CNRS UMR 6074 - Inria

ABSTRACT

Acoustic source localization is traditionally performed using cues such as interchannel time of arrival and intensity differences to infer the geometric localization of emitting sources with respect to the receiving microphone array. However the presence of obstacles between the sources and the array makes it impossible to rely on the direct path, and more advanced techniques are needed. The huge body of work on sparse recovery suggests an approach where source localization is expressed as a linear inverse problem and the spatial sparsity of the sources is exploited. An inverse problem can be naturally expressed in the recently introduced cosparse framework, exploiting the fact that the acoustic pressure satisfies the homogeneous wave equation except in the few locations of the sources. The resulting optimization problem involves a discretized second derivative analysis operator, which is extremely sparse. In this paper, we demonstrate the performance of the cosparse approach on an extreme source localization problem, where the microphone array is installed in the room next door to the room where the emitting sources are located, somehow hearing behind a wall.

Index Terms— localization, cosparsity, sparse analysis, wave equation

1. INTRODUCTION

Acoustic source localization is a challenging problem that arises commonly in fields such as robotics [1], speech and sound enhancement [2], acoustic tomography [3] and many others. Reverberations make the problem harder to solve and it becomes particularly difficult if the sound sources are obscured by an obstacle (a wall, for instance - Figure 1).

If the domain includes an obstacle between the microphones and the sources, as presented on Figure 1, the problem is insolvable by traditional goniometric methods [4, 5]. Indeed, most of these methods are based on the time difference of arrival (TDOA) approach. It usually involves computing the cross-correlations between the recorded signals, and then using this information for computing the positions of the sources, assuming the direct propagation path. For the spatial domain proposed here, however, cross-correlation between microphones is not informative, which can be seen on Figure 2 (the highest peaks on the right graph correspond to the reflections).

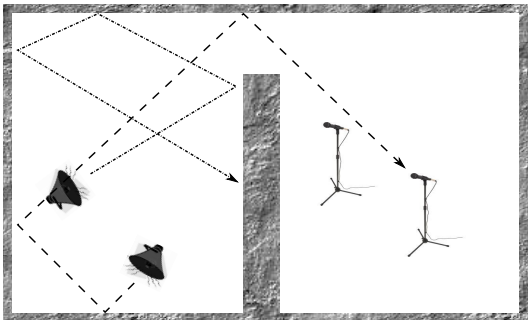

Fig. 1: Prototype "hearing behind a wall" scenario

It is possible to formulate source localization as an inverse problem. Let $\mathbf{p}_t \in \mathbb{R}^n$ denote the discretized sound pressure at the time instant $t \in [1, T]$ in a spatial domain Γ indexed by an integer in $[1, n]$. If there are m microphones distributed in Γ, the signal recorded by all of them is equal to $\mathbf{y}_t = \boldsymbol{\mathcal{M}}(\mathbf{p}_t) + \boldsymbol{\varepsilon}_t \in \mathbb{R}^m$, where $\boldsymbol{\mathcal{M}} \in \mathbb{R}^{m \times n}$ $(m < n)$ represents the subsampling system, and $\boldsymbol{\varepsilon}_t$ is the additive noise.

By concatenating these vectors for all time instances in the finite interval $[1, T]$, we formulate the following inverse problem: *find the source positions given the (possibly noisy) measurements* $\mathbf{y} \in \mathbb{R}^{mT}$ *of the acoustic pressure field* $\mathbf{p} \in \mathbb{R}^{nT}$. Formally:

$$\mathbf{y} = \mathbf{M}\mathbf{p} + \boldsymbol{\varepsilon}, \text{ where } \mathbf{M} \in \mathbb{R}^{mT \times nT} \tag{1}$$

Unfortunately, this apparently requires the estimation of the entire pressure field, which is an ill-posed problem, even in the noiseless case (since (1) has infinitely many solutions). Generally, to regularize ill-posed problems, one seeks the solutions which satisfy a certain *data model*. This is often done by encouraging solutions that embed some form of sparsity.

This work was supported in part by the European Research Council, PLEASE project (ERC-StG-2011-277906).

[†] Centre Inria Rennes - Bretagne Atlantique, France

[⋆] Rennes, France

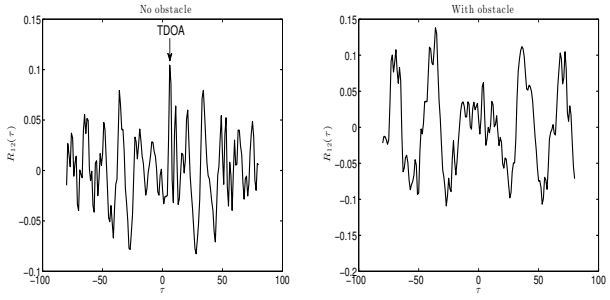

Fig. 2: The cross-correlations of the impulse responses in a bounded 2D space without (left) and with a wall (right)

Indeed, there is knowledge about the signal which can be exploited. It is known that the sound pressure obeys the acoustic wave equation:

$$\Delta p(\vec{r},t) - \frac{1}{c^2}\frac{\partial^2 p(\vec{r},t)}{\partial t^2} = \begin{cases} 0, \text{ if no source at location } \vec{r} \\ f(\vec{r},t), \text{ if source at location } \vec{r} \end{cases} \tag{2}$$

In other words, this partial differential equation is homogeneous for all regions of space not being occupied by sources. At the remaining positions, it will contain a non-zero right term $f(\vec{r},t)$ which represents the contribution of the sound source at the position $\vec{r}$ at time t. The constant c represents the sound propagation speed in the medium.

Assuming that the number of sound sources is small compared to the size of the spatial domain, one can envision two approaches. The most common is *sparse synthesis*, which would in this case mean estimating the pressure field as the linear combination of a small number of column vectors taken from the large dictionary of associated *Green's functions* [6] $\mathbf{\Psi}$ (*i.e.* $\hat{\mathbf{p}} = \mathbf{\Psi}\hat{\boldsymbol{\alpha}}$, and $\hat{\boldsymbol{\alpha}}$ is sparse). The objective would be to minimize the "ℓ_0-norm"[1] of the weights $\boldsymbol{\alpha}$ used to generate the estimate $\hat{\mathbf{p}}$ (as done in, *e.g.* [7]):

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \text{ s.t. } \|\mathbf{y} - \mathbf{M}\mathbf{\Psi}\boldsymbol{\alpha}\|_2 \leq \sigma_\varepsilon \tag{3}$$

There are two main problems with this approach, and they are both related to the dictionary $\mathbf{\Psi}$. Firstly, tailored Green's functions often need to be computed numerically, since the analytical solutions exist only for some (simple) spatial geometries. The second issue is practical: the matrix $\mathbf{\Psi}$ is usually dense and its size grows polynomially with dimensions, making the optimization problem quickly intractable in storage and computational cost.

2. COSPARSE REGULARIZATION

Recently, a different approach, based on the *sparse analysis* or *cosparse* data model has been proposed [8] to deal with the localization task. This approach is somehow more intuitive, as it naturally arises from the wave equation (2). If by $\mathbf{\Omega}$ we denote the discretized D'Alembertian operator [9], then applying $\mathbf{\Omega}$ to the vectorized acoustic pressure $\mathbf{p}$ will induce a *sparse product* $\mathbf{z} = \mathbf{\Omega}\mathbf{p}$. If the total number of zero components of this vector is l, we term the signal $\mathbf{p}$ to be l*-cosparse*. The aim of cosparse regularization is to promote solutions $\hat{\mathbf{p}}$ of (1) for which l is as large as possible. It corresponds to the minimization of ℓ_0-norm of $\mathbf{z}$, given the noisy measurements $\mathbf{y}$:

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \|\mathbf{\Omega}\mathbf{p}\|_0 \text{ s.t. } \|\mathbf{y} - \mathbf{M}\mathbf{p}\|_2 \leq \sigma_\varepsilon \tag{4}$$

Problems (3) and (4) are equivalent only in the special case when $\mathbf{\Psi} = \mathbf{\Omega}^{-1}$ *i.e.* both matrices are square and invertible [10].

For both models, ℓ_0 optimization is NP-hard [11, 12] and feasible solutions are obtained through approximations. Common approaches to approximate ℓ_0 cosparse solutions are convex relaxation and greedy methods. Convex relaxation methods substitute ℓ_0 by some convex norm [11, 13] (thus a global minimum can be obtained), while greedy algorithms use iterative schemes to approximate the *cosupport* (set of rows of $\mathbf{\Omega}$ orthogonal to the estimate) [14, 15]. Empirical results [13] lead to the conclusion that, for cosparsity applied to the wave equation, ℓ_1 minimization offers the most robust recovery performance. For the purpose of demonstrating the concept, we will focus only on the noiseless case ($\sigma_\varepsilon = 0$), thus we define the optimization problem as follows:

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \|\mathbf{\Omega}\mathbf{p}\|_1 \text{ s.t. } \mathbf{y} = \mathbf{M}\mathbf{p} \tag{5}$$

3. IMPLEMENTATION

We formulate the acoustic wave equation (2) (for the two-dimensional domain Γ) in a discretized form:

$$(\mathbf{\Omega}\mathbf{p})_{ij}^t = \begin{cases} 0, \text{ if no source at location } \vec{r} = (i,j) \\ f_{i,j}^t, \text{ if source at location } \vec{r} = (i,j) \end{cases} \tag{6}$$

To discretize the D'Alembertian operator we use the finite difference method through the leap-frog scheme [16]. For a 2D spatial domain and unit stepsizes, we can express (6) with the following causal relation:

$$\begin{aligned} p_{i,j}^{t+1} = \tilde{c}^2 \left(p_{i-1,j}^t + p_{i+1,j}^t + p_{i,j-1}^t + p_{i,j+1}^t\right) + \\ - 2\tilde{c}^2(2 - 1/\tilde{c}^2)p_{i,j}^t - p_{i,j}^{t-1} + f_{i,j}^{t+1} \end{aligned} \tag{7}$$

In case that the location (i,j) is not occupied by a source at time $t+1$, the source contribution term $f_{i,j}^{t+1}$ is equal to 0. The constant $\tilde{c}^2$ depends on the resolution of the space-time grid, and for unit stepsizes CFL condition [17] suggests it is less or equal to $1/\sqrt{2}$ in 2D domains, and $1/\sqrt{3}$ in 3D domains, to preserve the stability of the scheme.

Equation (2) is a second order partial differential equation, hence the initial values of p with its first derivative at $t = 0$ are required to ensure it has a unique solution. In accordance with

[1] $\|\mathbf{u}\|_0 := \#\{\mathbf{u}\}$, the count of non-zero elements in $\mathbf{u}$.

our assumption, we set zero values for both (approximated by $p^1_{i,j} = p^2_{i,j} = 0$), meaning that the sources may start emitting only after $t = 2$. Assuming Dirichlet boundary condition, we set the boundary values to 0 at all times. Other types of boundaries can also be implemented (e.g. Neumann reflecting boundary condition).

Finally, we obtain a full rank square matrix $\mathbf{\Omega}$ of size $N \times N$, where $N = nT = IJT$ (I, J represent the spatial resolution, T is the discretized time span of the experiment - the *acquisition time*).

The discretization (7) yields an operator $\mathbf{\Omega}$ which is extremely sparse: each row can have at most seven non zero elements. From a computational point of view this is very favorable and the benefit can be observed in the iterative update steps of the (scaled) ADMM (*Alternating Direction Method of Multipliers* [18]) algorithm used for numerically solving the optimization problem (5):

$$\begin{aligned} \hat{\mathbf{p}}_k &= \arg\min_p \|\mathbf{\Omega p} - \hat{\mathbf{z}}_{k-1} + \mathbf{u}_{k-1}\|_2^2 \text{ s.t. } \mathbf{y} = \mathbf{Mp} \\ \hat{\mathbf{z}}_k &= \mathcal{S}_{1/\rho}(\mathbf{\Omega}\hat{\mathbf{p}}_k + \mathbf{u}_{k-1}) \\ \mathbf{u}_k &= \mathbf{u}_{k-1} + \mathbf{\Omega}\hat{\mathbf{p}}_k - \hat{\mathbf{z}}_k \end{aligned} \tag{8}$$

The second step ($\mathcal{S}_{1/\rho}(\cdot)$) is just an element-wise soft thresholding, while the update of auxiliary variable $\mathbf{u}_k$ requires only vector addition. Hence, the first step is the most computationally expensive, since it imposes solving the linearly constrained linear least squares problem. However, it involves the sparse matrix $\mathbf{\Omega}$ and the subsampling matrix $\mathbf{M}$, thus the problem scales as $\mathcal{O}(IJT)$.

Since $\mathbf{\Omega}$ is a square invertible matrix, the analysis and the synthesis problems are formally equivalent and $\mathbf{\Psi}$ could theoretically be computed by taking its inverse. However, the dictionary $\mathbf{\Psi}$ is not sparse [13] and the analogous implementation of the sparse synthesis ℓ_1 ADMM optimization would require multiplying $\mathbf{\Psi}$ and $\mathbf{\Psi}^T$ in the $\hat{\mathbf{p}}_k$-update step, which would effectively scale as $\mathcal{O}\left((IJT)^2\right)$.

4. HEARING BEHIND WALLS

Experiments in two dimensions have been conducted in a simulated "split room" environment $(I = 30) \times (J = 30)$ presented on Figure 3. The number of sensors is set constant to $m = 10$ and they have been randomly distributed in the right bottom quarter of the room[2]. The acquisition time is set to $T = 400$ and the source emitting duration is set to $T_e = 10$.

For each experiment, we place s wideband sources (modeled as the white Gaussian noise emitters with the amplitude distribution $\mathcal{N}(0, 1)$) randomly in the left bottom quarter of the room. Then, for a given free space distance between the obstacle and the opposite wall ("*the door width*") w, we compute the ground truth signal $\mathbf{p}$ using the leap-frog expression (7). Finally, the numerical solution $\hat{\mathbf{p}}$ of (5) is computed using the method (8). The experiment is repeated 10 times.

[2]Thus, "the wall" does not completely divide the room.

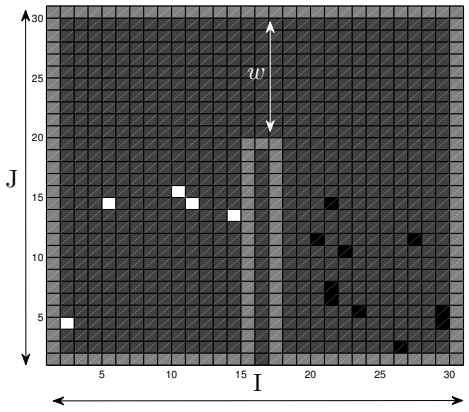

Fig. 3: Discretized "split room" in 2D (*white pixels*: sources, *black pixels*: sensors, *light gray*: "walls", *dark gray*: propagation medium)

Then, we vary the number of sources ($1 \leq s \leq m$) and the door width: from $w = 1$ (only one pixel wide) to $w = 28$ (no obstacle).

Following (6), it is indicative that non-zeros in the product $\hat{\mathbf{z}} = \mathbf{\Omega}\hat{\mathbf{p}}$ represent the potential sources: vector $\hat{\mathbf{z}}$ consists of T slices of source contributions for each spatial location. The most likely locations (i, j) are the ones having the highest magnitude sum $\hat{\mathbf{Z}}_{i,j} = \sum_{t=1}^{T} |\hat{\mathbf{z}}^t_{i,j}|$. Since the number of sources in usually not known in advance, the detection of source locations is done by applying some threshold λ. Therefore, standard *precision* P_λ and *recall* R_λ measures are used to evaluate the localization performance. If we term the number of correctly identified sources by $\bar{s}(\lambda)$ and the total number of identified sources by $\hat{s}(\lambda)$, these values are equal to $P_\lambda = \bar{s}(\lambda)/\hat{s}(\lambda)$ and $R_\lambda = \bar{s}(\lambda)/s$. In addition, we compute the empirical probability of accurate source localization given the total number of sources: $P_s = \bar{s}/s$. Here $\bar{s}$ represents the number of correctly identified sources from the set of locations (i, j) obtained by keeping s highest in magnitude sums $\hat{\mathbf{z}}_{i,j}$.

For measuring the localization performance, we maintain the *total accuracy* principle: to compute $\bar{s}(\lambda)$ and $\bar{s}$ we classify as correctly identified only those locations (i, j) which **exactly** correspond to the ground truth position of the sources.

One advantage of the proposed approach is that it is not limited to localization, but yields the acoustic pressure field estimate as a natural byproduct. Therefore, we can also evaluate the wavefield signal-to-noise ratio $\text{SNR}_{\mathbf{p}} = 20 \log_{10} \|\mathbf{p}\|_2 / \|\mathbf{p} - \hat{\mathbf{p}}\|_2$.

As an outlook to the scaling capabilities we also conduct illustrative experiment in three dimensions. Thus we focused on a single setup (fixing $s = 3$ and $w = 10$) in a simulated space of size $(I = 20) \times (J = 20) \times (K = 20)$ with duration $T = 400$, whose results were obtained by averaging the outcome of 10 consecutive experiments. The equivalent sparse

synthesis setup would require the dictionary matrix $\mathbf{\Psi}$ of the order $(IJKT)^2 \sim 10^{11}$ non-zero elements.

5. RESULTS

Figure 4 shows precision and recall graphs for the cases of $s = 4$ and $s = 10$ sources in space, and different door widths w. The presented results indicate that already a small door width ($w = 5$) is sufficient to highly accurate localize the sources, even when their number is high.

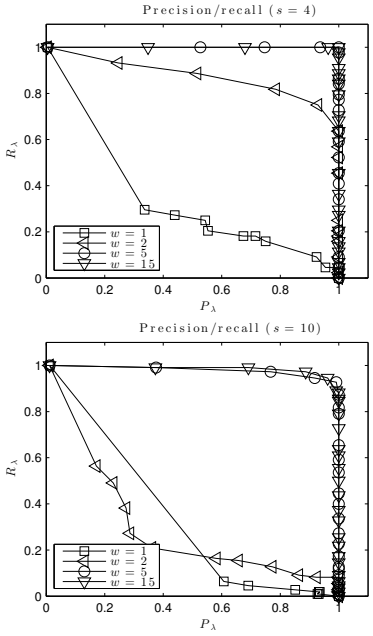

Fig. 4: Precision/recall diagrams for $s = 4$ (up) and $s = 10$ (down) sources

Figure 5 (left) presents the empirical probability P_s for varying s and w parameters. We can see that the localization probability is high, even in those cases where the door width is considerably small. As expected, the performance is lower for higher number of sources (*i.e.* lower cosparsity) and smaller door width.

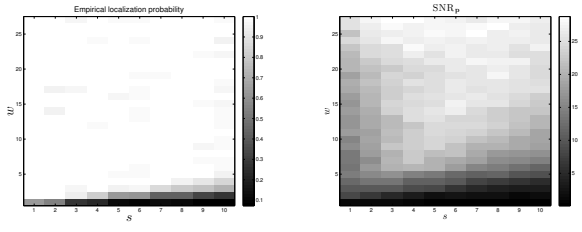

Fig. 5: Probability of accurate source localization given s (left) and wavefield SNR (right)

Figure 5 (right) depicts the estimated $\mathrm{SNR}_\mathbf{p}$, for the same range of s and w. It seems that these results are correlated with the source localization probability, although there are some surprises, namely the fact that $\mathrm{SNR}_\mathbf{p}$ is not the highest for the signals having the highest cosparsity (left side of the SNR graph).

The obtained results are in accordance with physics of propagation. The well-known Huygens-Fresnel principle suggests that there is a minimal door width $\tilde{w}$ beyond which it will be impossible to detect the sources in the other half of the room: it will always appear as if they are located at the door position. This is exactly what happens for very small values of w in our experiments.

Figure 6 is the precision/recall graph for the three-dimensional setup. For conveniently chosen range of thresholds, it was possible to accurately localize the sources in 9 out of 10 experiments. The computational time per experiment was approximately 2 to 3 times higher than needed for the 2D experiments presented before. This experiment could not have been conducted using the equivalent synthesis approach, due to its extremely high computational and storage requirements.

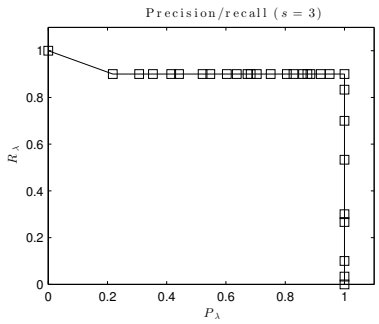

Fig. 6: Precision/recall graph for 3D problem ($w = 10$)

6. CONCLUSION

We have presented a method, based on cosparse data model, for sound source localization behind the obstacle that blocks the direct propagation path. The experimental results confirmed the assumption that sparse analysis based on the physical model of the wave propagation performs well even in complicated spatial domains and long time spans, where the equivalent sparse synthesis model is intractable. Furthermore, it is possible to scale the problem to three dimensions without significant impact on the accuracy.

Future work will be aimed towards real-world experiments and extended scenarios. One can envision cases where some physical properties are not known in advance, *e.g.* wave propagation speed, boundary type or shape. Additionally, since a signal estimate is also produced by the approach, it may be used to perform source signal separation or to deploy virtual microphones. The cosparse regularization could be one key to solve challenging inverse problems such as these.

7. REFERENCES

[1] K. Nakadai, H. G. Okuno, H. Kitano, et al., "Real-time sound source localization and separation for robot audition.," in *INTERSPEECH*, 2002.

[2] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *Signal Processing, IEEE Transactions on*, vol. 49, no. 8, pp. 1614–1626, 2001.

[3] W. Munk, P. Worcester, and C. Wunsch, *Ocean acoustic tomography*, Cambridge University Press, 2009.

[4] E. Van Lancker, *Acoustic goniometry*, Ph.D. thesis, STI, Lausanne, 2002.

[5] T. Kundu, "Acoustic source localization," *Ultrasonics*, vol. 54, no. 1, pp. 25 – 38, 2014.

[6] L. Ziomek, *Fundamentals of Acoustic Field Theory and Space-Time Signal Processing*, CRC Press, 1995.

[7] I. Dokmanic and M. Vetterli, "Room helps: Acoustic localization with finite elements," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. Ieee, 2012, pp. 2617–2620.

[8] S. Nam and R. Gribonval, "Physics-driven structured cosparse modeling for source localization," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 5397–5400.

[9] M. Bruneau, *Fundamentals of acoustics*, vol. 99, John Wiley & Sons, 2010.

[10] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," in *Inverse Problems 23*, 2007, pp. 947–968.

[11] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, "The Cosparse Analysis Model and Algorithms," *Applied and Computational Harmonic Analysis*, vol. 34, no. 1, pp. 30–56, 2013.

[12] B. K. Natarajan, "Sparse Approximate Solutions to Linear Systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, 1995.

[13] S. Kitić, N. Bertin, and R. Gribonval, "A review of cosparse signal recovery methods applied to sound source localization," in *Le XXIVe colloque Gretsi*, 2013.

[14] R. Giryes, S. Nam, M. Elad, R. Gribonval, and M. E. Davies, "Greedy-Like Algorithms for the Cosparse Analysis Model," *arXiv preprint arXiv:1207.2456*, 2013.

[15] S. Nam, M.E. Davies, M. Elad, and R. Gribonval, "Recovery of cosparse signals with greedy analysis pursuit in the presence of noise," in *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2011 4th IEEE International Workshop on*. IEEE, 2011, pp. 361–364.

[16] A.R. Mitchell and D.F. Griffiths, *The finite difference method in partial differential equations*, Wiley-Interscience publication. Wiley, 1980.

[17] G. Strang, *Applied Mathematics and Scientific Computing*, Wellesley-Cambridge Press, 2007.

[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.