

Fast LASSO Based DOA Tracking

Ashkan Panahi and Mats Viberg
 Signal Processing Group,
 Department of Signals and Systems,
 Chalmers University of Technology, 412 96 Gothenburg, Sweden
 {ashkanp, Viberg}@chalmers.se

Abstract—In this paper, we propose a sequential, fast DOA tracking technique using the measurements of a uniform linear sensor array in the far field of a set of narrow band sources. Our approach is based on sparse approximation technique LASSO (Least Absolute Shrinkage and Selection Operator), which has recently gained considerable interest for DOA and other estimation problems. Considering the LASSO optimization as a Bayesian estimation, we first define a class of prior distributions suitable for the sparse representation of the model and discuss its relation to the priors over DOAs and waveforms. Inspired by the Kalman filtering method, we introduce a nonlinear sequential filter on this family of distributions. We derive the filter for a simple random walk motion model of the DOAs. The method consists of consecutive implementation of weighted LASSO optimizations using each new measurement and updating the LASSO weights for the next step.

I. INTRODUCTION

The Direction of Arrival (DOA) tracking problem has been a central research topic for a long time due to its wide range of application. The problem considered here is to estimate a sequence of DOAs using a model consisting of the measurement and time evolution rules. As an estimation problem, one may obviously apply the classical estimation rules such as the ML method to obtain the DOA estimates. However, this is generally a complex method. Furthermore, one may be interested in a sequential estimating technique to be able to follow time-varying scenarios.

The DOA tracking problem is closely related to the simpler problem of DOA estimation in which it is assumed that there is no DOA evolution. One obvious tracking technique in a general case is to use a DOA estimator for each individual measurement and try to correct them by the movement model. From a statistical point of view, the primal DOA estimates play the role of prior knowledge for the estimation over the evolution model. Such a method also needs a long computational time because of the data association problem. Still, a good tracking technique can be achieved by generalizing the one-snapshot DOA estimation to multiple snapshots. A widely used methodology is to follow the ideas of Kalman filtering. The idea is to represent the knowledge of previous measurements by a conditional distribution, which serves as a prior for the DOA estimation at the current time. The DOA estimation technique is then generalized to admit the prior as a Bayesian estimator. In [1], a subspace method is introduced as a sequential estimator.

In the context of DOA estimation, the Least Absolute and Shrinkage Operator (LASSO) sparse regression technique [2] has recently been utilized by reformulating the measurement model in a sparse framework [3]. This representation resolves

the permutation problem. In [4] the LASSO is introduced as a Bayesian DOA estimation by a Laplacian prior. The LASSO based DOA estimator consists of representing the discretized measurement model as a noisy under-determined linear regression. It is solved by the LASSO technique, which is a linear least square problem regularized by the ℓ_1 norm of the parameter vector as a measure of sparsity. In [5], the possibility of re-weighting the LASSO technique for a better estimation is discussed.

Following the Kalman methodology, we try to generalize the LASSO technique to admit an adaptive prior through weighting the regularization parameters. Clearly the calculation time increases linearly with the number of snapshots. We express the past information at the current time by a weight updating rule. This can be done by applying the Bayes rule for the sparse parameter vector with a Laplacian distribution. However, since the time evolution model is represented by the DOA parameters, we introduce a rule to rearrange the distribution in different representation domains.

II. SYSTEM MODEL

Suppose a uniform linear array of m sensors in the far field of a set of n sources, each sending a narrow band signal with complex envelope samples $s_i(t)$, $i = 1, 2, \dots, n$, and $t = 1, 2, \dots$. The sensors are separated by d in wavelengths. The sources are also characterized by their position angle $\theta_i(t)$ with respect to the array axis at the t^{th} sampling time. Then, neglecting the Doppler effect, constant phase shift, delay and scaling, the measured signal vector $\mathbf{x}(t) = [x_1(t) \ x_2(t) \ \dots \ x_m(t)]^T$ ¹ at time t by the array can be written as

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta}(t))\mathbf{s}(t) + \mathbf{n}(t), \quad (1)$$

where $\boldsymbol{\theta}(t) = [\theta_1(t) \ \theta_2(t) \ \dots \ \theta_n(t)]^T$, $\mathbf{s}(t) = [s_1(t) \ s_2(t) \ \dots \ s_n(t)]^T$, $\mathbf{n}(t)$ is the zero-mean, circularly symmetric ($\mathcal{E}[\mathbf{n}(t)\mathbf{n}^T(t)] = \mathbf{0}$) Gaussian measurement noise vector with covariance matrix $\mathcal{E}[\mathbf{n}(t)\mathbf{n}^H(t)] = \sigma^2\mathbf{I}$ and $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1) \ \mathbf{a}(\theta_2) \ \dots \ \mathbf{a}(\theta_n)]$ is the matrix of steering vectors at the DOA vector $\boldsymbol{\theta}$ with $\mathbf{a}(\theta) = [1 \ e^{j2\pi d \cos \theta} \ e^{j4\pi d \cos \theta} \ \dots \ e^{j2\pi(m-1)d \cos \theta}]^T$ as the steering vector. We also define $\phi = 2\pi d \cos \theta$ as the electrical angle.

The DOA sequence is a process which follows an evolution model. Normally, the process is Markov. This means that for

¹We denote the transpose, conjugate transpose (Hermitian), and statistical expectation operators by $(\cdot)^T$, $(\cdot)^H$, and $\mathcal{E}[\cdot]$ respectively.

each sequence t_1, t_2, \dots, t_T of increasing sample times, we have

$$p(\boldsymbol{\theta}(1), \boldsymbol{\theta}(2), \dots, \boldsymbol{\theta}(T)) = p(\boldsymbol{\theta}(1)) \prod_{i=2}^T p(\boldsymbol{\theta}(i)|\boldsymbol{\theta}(i-1)), \quad (2)$$

where $p(\cdot)$ denotes the PDF (Probability Density Function) of the argument. The problem is to estimate a sequence of DOAs $\hat{\boldsymbol{\theta}}(t)$ given a sequence of observations $\mathbf{x}(t)$. One may try to solve the problem using statistical methods [6]. Practically, this is hard to perform. Note that the model $p(\boldsymbol{\theta}(t)|\boldsymbol{\theta}(t-1))$ is related to a continuous evolution. This implies that the model is represented by a measure of closeness over the ordered set $\boldsymbol{\theta}$. This makes the problem complicated because of an unknown permutation in the representation of $\boldsymbol{\theta}$, which brings up the computationally costly problem of data association. To avoid this problem, we introduce the sparse representation which is used with the LASSO based DOA estimation method [7].

III. SPARSE REPRESENTATION

Like [7], we discretize the problem using a finite grid $\boldsymbol{\theta}^G = \{\theta_1^g, \theta_2^g, \dots, \theta_N^g\}$ with a sufficiently large number of elements N . Assuming $\theta_{i_1}^g(t), \theta_{i_2}^g(t), \dots, \theta_{i_n}^g(t)$ as the closest points in the grid to the true DOAs $\theta_1, \theta_2, \dots, \theta_n$, and defining

$$s_k^g(t) = \begin{cases} s_l(t) & k = i_l(t) \\ 0 & \text{otherwise} \end{cases}, \quad (3)$$

we observe from (1) that

$$\mathbf{x}(t) \approx \mathbf{A}^g \mathbf{s}^g(t) + \mathbf{n}(t), \quad (4)$$

where $\mathbf{A}^g = \mathbf{A}(\boldsymbol{\theta}^G) = [\mathbf{a}(\theta_1^g) \ \mathbf{a}(\theta_2^g) \ \dots \ \mathbf{a}(\theta_N^g)]$ and $\mathbf{s}^g(t) = [s_1^g(t) \ s_2^g(t) \ \dots \ s_N^g(t)]^T$. The approximation improves as $\boldsymbol{\theta}^G$ becomes denser. The problem can be solved for one snapshot without the evolution model by minimizing the following function [3], known as the LASSO method

$$\mathcal{L}(\mathbf{s}^g) = \frac{1}{2} \|\mathbf{x} - \mathbf{A}^g \mathbf{s}^g\|_2^2 + \lambda \|\mathbf{s}^g\|_1, \quad (5)$$

where $\|\mathbf{s}^g\|_1 = \sum_{i=1}^N |s_i^g|$ is the ℓ_1 norm of \mathbf{s}^g . We neglect the time arguments since there is only one snapshot. In [5] it is shown that the performance of such a method might be improved by weighting the second term, so that the minimization is over

$$\mathcal{L}_w(\mathbf{s}^g) = \frac{1}{2} \|\mathbf{x} - \mathbf{A}^g \mathbf{s}^g\|_2^2 + \sum_{i=1}^N \lambda_i |s_i^g|. \quad (6)$$

The optimizations such as (5) and (6) are convex and can be solved by relatively fast convex optimization techniques [8].

A. Weighted LASSO as a Bayesian Estimator

In [7], the LASSO technique has also been interpreted as a Maximum A posteriori Probability estimator with a Laplacian prior. However, there are ambiguities in this definition, since the sparse points are unlikely to happen in such a Laplacian model. In this section we try to illustrate the situation.

Note that increasing λ_i in (6) increases the relative importance of the absolute value of the i^{th} source, and thus decreases the chance of this source to be nonzero. We conclude that there exists a relation between the weights λ_i and the prior knowledge about the DOAs. Now, we start by a natural way of imposing prior considerations in the sparse representation, by assuming independent Bernoulli priors with very small probability of occurrence p_i for each DOA θ_i^g in $\boldsymbol{\theta}^G$. This method is further discussed in [9]. For such a prior, the negative logarithm of the likelihood function can be written as

$$-\ln p(\mathbf{x}|\mathbf{s}^g) = \frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{A}^g \mathbf{s}^g\|_2^2 - \sum_{i=1}^N \gamma(s_i) \ln \frac{p_i}{1-p_i} + c, \quad (7)$$

where c is a constant and

$$\gamma(s) = \begin{cases} 1 & s \neq 0 \\ 0 & s = 0 \end{cases}, \quad (8)$$

is the indicator function. Comparing (7) and (6) we observe that the absolute value $|s|$ plays the role of a term proportional to the indicator function, while

$$-\ln \frac{p_i}{1-p_i} \simeq \lambda \lambda_i. \quad (9)$$

Where λ is suitable normalizing factor. We use this relation later to exchange the Laplacian priors with Bernoulli ones whenever it is necessary.

IV. KALMAN FILTERING

The idea of Kalman filtering applied to sparse regression problems [10] is to concentrate the DOA information from past observations into a PDF (in our case p_i), which can be imposed to the estimation by the current observation and a Bayesian method. For our case and for the Bernoulli interpretation we can introduce

$$p_i(t|t) = p(\theta_i(t)|\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(t)) \\ p_i(t|t-1) = p(\theta_i(t)|\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(t-1)). \quad (10)$$

The updating steps can be found using the Bayes rule. Note that having $p_i(t|t-1)$ at hand we can find the corresponding Laplacian prior using (9). Then,

$$p(\mathbf{s}^g(t)|\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(t)) \propto \\ p(\mathbf{x}(t)|\mathbf{s}^g(t))p(\mathbf{s}^g(t)|\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(t-1)). \quad (11)$$

Using (9) again, we can find $p_i(t|t)$. For the next step of computing $p_i(t+1|t)$ from $p_i(t|t)$ we need a motion model. We assume a simple random walk model, in which the probability of a source at angle θ_j to move to θ_i is given by $p(\theta_i|\theta_j)$. Using the Markov property in (2), the first order expansion of $p_i(t+1|t)$ can be written as

$$p_i(t+1|t) \approx K \sum_{j=1}^N p(\theta_i|\theta_j) p_j(t|t) \quad (12)$$

where K is a constant. The approximation means that we neglect the probability that two or more sources gather to the same point at the next time, which is practically acceptable.

A. A LASSO based Sequential Bayesian Estimator

Generalizing the LASSO idea at each time, we assume a Laplacian prior of the form

$$-\ln p(\mathbf{s}^g(t)|\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(t-1)) = \sum_{i=1}^N \mu_i(t) |s_i^g| + c, \quad (13)$$

where c is a constant, and estimate directions using (6). The idea is to find a proper prior in this family for the next time using (10), (11), and (12). However, this turns out to be difficult without reasonable approximations. Using (11) we have

$$\begin{aligned} & -\frac{1}{\sigma^2} \ln p(\mathbf{s}^g(t)|\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(t)) \\ &= \frac{1}{2} \|\mathbf{x}(t) - \mathbf{A}^g \mathbf{s}^g(t)\|_2^2 + \sum_{i=1}^N \lambda_i(t) |s_i^g| + c', \end{aligned} \quad (14)$$

where $\lambda_i(t) = \sigma^2 \mu_i(t)$ and c' is another constant. Running the LASSO optimization we find out the MAP estimates $\hat{\mathbf{s}}^g(t)$. The negative log-likelihood function can then be expanded about this point. Introducing $I = \{i_1, i_2, \dots, i_n\}$ as the set of all active indexes with nonzero $\hat{s}_i^g(t)$, we get after straightforward calculations

$$\begin{aligned} & -\frac{1}{\sigma^2} \ln p(\mathbf{s}^g(t)|\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(t)) \\ & \approx \mathcal{L}_w(\hat{\mathbf{s}}^g(t)) \\ & + \sum_{i \notin I} \lambda_i(t) |s_i^g(t)| + \text{Re}(s_i^g(t)^* \mathbf{a}^H(\theta_i^g) \hat{\mathbf{n}}(t)) + c' \\ & = \sum_{i \notin I} |s_i^g(t)| (\lambda_i + \nu_i(t) \cos(\phi_i - \rho_i(t))) + c'', \end{aligned} \quad (15)$$

where $\mathcal{L}_w(\cdot)$ is given in (6), $\hat{\mathbf{n}}(t) = \mathbf{x}(t) - \mathbf{A}^g \hat{\mathbf{s}}^g(t)$, $\nu_i(t) e^{j\rho_i(t)} = \mathbf{a}^H(\theta_i^g) \hat{\mathbf{n}}(t)$, and $s_i^g(t) = |s_i^g(t)| e^{j\phi_i}$. However, (15) is not of the Laplacian form, which can be transformed to a Bernoulli one over the DOAs. This means that the measurement not only gives information about the DOAs, but it also provides information about the waveforms which should be integrated out. It is straightforward to generalize (9) for this new case by replacing the absolute function with the normalized indicator function. We skip some computational steps that eventually leads to

$$\begin{aligned} \frac{p_i(t|t)}{1 - p_i(t|t)} &= \int_0^{2\pi} e^{-\lambda(\lambda_i + \nu_i \cos(\phi - \rho))} d\phi \\ &= e^{-\lambda \lambda_i} I_0(\lambda \nu_i) \approx e^{-\lambda(\lambda_i(t) - \nu_i(t))}, \end{aligned} \quad (16)$$

where λ is a constant and $I_0(\cdot)$ is the zeroth order modified Bessel function [11]. The approximation comes from the fact that the Bernoulli probabilities are very small. Thus, λ is very large and the Bessel function reaches its asymptotic value. Note that because $p_i(t|t)$ is small we can also write

$$p_i(t|t) \approx e^{-\lambda(\lambda_i - \nu_i)}, \quad (17)$$

which completes the measurement update state. Next, we can use (12) to compute $p_i(t+1|t)$.

B. DOA Tracking For a Simple Motion Model

Now we simplify (12) for the simple but important case of a random walk with finite lags l . Suppose $p(\theta^l|\theta)$ is given by

$$p(\theta_k|\theta_i) = \begin{cases} \alpha_{k-i} & |k-i| \leq l \\ 0 & \text{otherwise} \end{cases}, \quad (18)$$

with $\sum_{i=-l}^l \alpha_i = 1$. Then, from (9), (17), and (12), considering small values for the Bernoulli probabilities we get

$$\begin{aligned} \lambda_i(t+1) &\propto -\ln \left[K \sum_{k=-l}^l \alpha_k e^{-\lambda(\lambda_{i+k}(t) - \nu_{i+k}(t))} \right] \\ &\simeq \max_{k \in \{-l, \dots, l\}} -\ln(K \alpha_k) + \lambda(\lambda_{i+k}(t) - \nu_{i+k}(t)), \end{aligned} \quad (19)$$

where we used the fact that λ is large in the approximation. Note that to achieve the right weights, the proportionality constant in (19) should be of the order of λ , which makes the first term negligible in (19). Thus, we get

$$\lambda_i(t+1) \approx \lambda' \max_{k \in \{-l, \dots, l\}} (\lambda_{i+k}(t) - \nu_{i+k}(t)), \quad (20)$$

We can now summarize the proposed method. Given a sequence of weights $\lambda_i(t)$ at time instant t , and receiving the measurement vector $\mathbf{x}(t)$ we can implement (6) to get the DOA estimates at time t . The weights may need to be multiplied by a constant in order to get the true number of sources. This is almost similar to the problem of regularization parameter selection in [7], which can also be solved by greedy algorithms [12] when the number of sources is known or model order selection methods [13] otherwise. By this method we actually solve for the λ' in (20). At this point we get new weights $\lambda' \lambda_i(t)$. Then, we can form $\nu_i(t)$ in (15) and update the weights by

$$\lambda_i(t+1) = \max_{k \in \{-l, \dots, l\}} (\lambda' \lambda_{i+k}(t) - \nu_{i+k}(t)). \quad (21)$$

we are now ready for the next measurement etc.

V. NUMERICAL RESULTS

We implemented the proposed technique by a greedy weighted LASSO algorithm at each step. We examine our method in two cases. First, we assume a scenario in which the sources do not move and the SNR is low. Second, we try a moving scenario with higher SNRs. We compare the results to the separate DOA estimation method (i.e. without using a tracking technique).

Figure 1 shows the weights for the low-noise fixed sources case with $l = 3$ in (18). As can be seen, the weights corresponding to the true directions keep small values while the other weights start to increase. This shows a kind of increasing certainty about the DOAs, which makes a different choice unlikely. Practically, at each step, λ' gets smaller since the weights increase. This affects the performance of the greedy algorithm. To avoid such a situation we introduce a normalization step after (21). Note that it does not affect the estimates but the robustness of the greedy method. Although in such a case the combination method in [3] may give better results, the proposed method's complexity grows linearly with the number of snapshots while that of the method in [3]

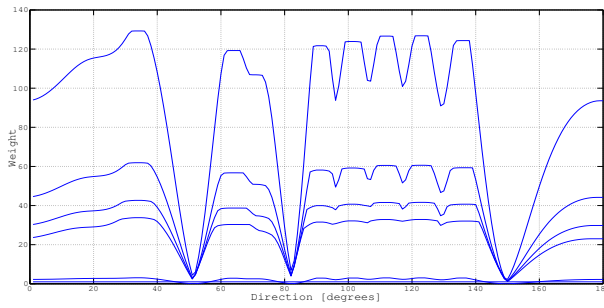


Fig. 1. The LASSO weights for SNR=10dB and 3 sources at fixed DOAs 50, 80, and 150 degrees. The weights grow larger by time.

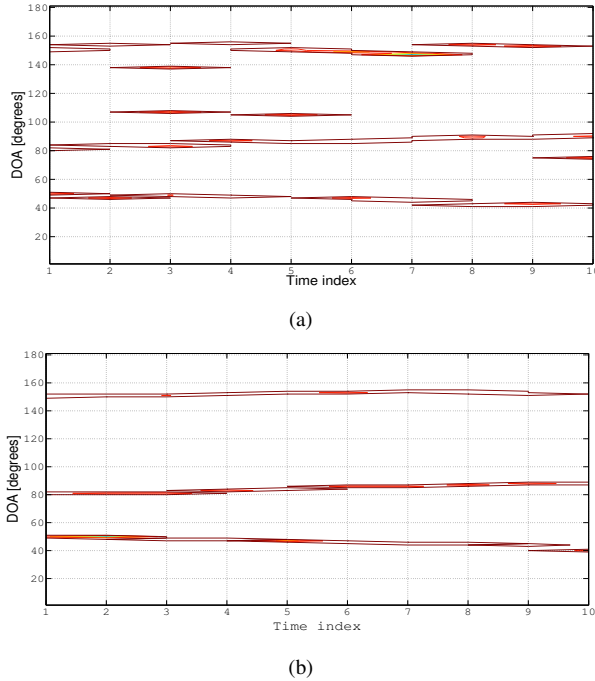


Fig. 2. The contour plot of the LASSO based DOA spatial spectrogram for multiple snapshots of moving objects at SNR=0 dB. The motion starts from DOAs 50, 80, and 150 degrees with velocity $\mathbf{v} = [-1 \ 1 \ 0]$ degrees per snapshot. Figure (a) shows the disjoint process while Figure (b) show the proposed method.

increases polynomially. The results are obviously better than the disjoint process.

Figure 2(b) shows the result of applying the LASSO based DOA tracking method to the multiple snapshot case with the true DOAs moving with velocity $\mathbf{v} = [-1 \ 1 \ 0]$ degrees per snapshot. One may compare the results to the separate process in Figure 2(a). Clearly, the proposed method is more robust to the measurement noise.

VI. CONCLUDING REMARKS

In this paper we discussed the possibility of combining the multiple measurement information to obtain a low-complexity sequential estimation method. We assumed a narrow-band far-field array measurement and a simple random walk evolution model over time. We obtained a sequential Bayesian estimator by updating the weights in the weighted LASSO based DOA

estimator in each snapshot. This was done by reinterpreting the Laplacian prior in the LASSO as an approximation to the more accurate Bernoulli model.

We showed that the method decreases the uncertainty by increasing the difference between weights of active and non-active generalized sources at each time innovation up to a limit defined by the evolution model uncertainty. We observe that the first order Bernoulli approximation works well. At low SNRs, the method occasionally fails due to the fact that there might not be any solution with a correct number of sources. This shows the main drawback of assuming a fixed number of sources and achieving such a solution by a greedy method. Furthermore, there is no possibility of adding or removing new sources, which is a benefit of using the Bernoulli model. The solution with a certain level of sparsity is sometimes hard to follow by the algorithm. In this manner a model order selection technique can be applied in each step. The current method also fails when the sources get too close to each other. It might be further enhanced by improving the approximation in (10) or the model in (17).

REFERENCES

- [1] S.B. Park, C.S. Ryu, and K.K. Lee, "Multiple target angle tracking algorithm using predicted angles," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 30, no. 2, pp. 643–648, apr 1994.
- [2] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B, (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- [3] D.M. Malioutov, M. Cetin, and A.S. Willsky, "Source localization by enforcing sparsity through a laplacian prior: an svd-based approach," *IEEE Workshop on Statistical Signal Processing*, pp. 573–576, sept.-1 oct. 2003.
- [4] T. Park and G. Casella, "The bayesian lasso," *Journal of the American Statistical Association*, vol. 103, pp. 681–686, 2008.
- [5] H Zou, "The adaptive lasso and its oracle properties," *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1418–1429, 2006.
- [6] M.R. Morelande and D. Musicki, "Fast multiple target tracking using particle filters," in *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05. 44th IEEE Conference on*, dec. 2005, pp. 530–535.
- [7] Ashkan Panahi and Mats Viberg, "Maximum a posteriori based regularization parameter selection," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, may 2011, pp. 2452–2455.
- [8] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [9] C. Fevotte, B. Torresani, L. Daudet, and S.J. Godsill, "Sparse linear regression with structured priors and application to denoising of musical audio," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 16, no. 1, pp. 174–185, jan. 2008.
- [10] D. Kanevsky, A. Carmi, L. Horesh, P. Gurfil, B. Ramabhadran, and T.N. Sainath, "Kalman filtering for compressed sensing," in *Information Fusion (FUSION), 2010 13th Conference on*, july 2010, pp. 1–8.
- [11] Stegun Irene A. Abramowitz Milton, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, 1964.
- [12] Bradley Efron, Trevor Hastie, Lain Johnstone, and Robert Tibshirani, "Least angle regression," *Annals of Statistics*, vol. 32, pp. 407–499, 2004.
- [13] P. Stoica and Y. Selen, "Model-order selection: a review of information criterion rules," *Signal Processing Magazine, IEEE*, vol. 21, no. 4, pp. 36–47, July 2004.