Learning sparse linear dynamic networks in a hyper-parameter free setting

Arun Venkitaraman, * **Håkan Hjalmarsson, and Bo Wahlberg**

* *Divison of Decision and Control Systems, KTH Royal Institute of Technology, Stockholm, Sweden. (e-mail: {arunv,hjalmars,bo}@ kth.se).*

Abstract: We address the issue of estimating the topology and dynamics of sparse linear dynamic networks in a hyperparameter-free setting. We propose a method to estimate the network dynamics in a computationally efficient and parameter tuning-free iterative framework known as SPICE (Sparse Iterative Covariance Estimation). Our approach does not assume that the network is undirected and is applicable even with varying noise levels across the modules of the network. We also do not assume any explicit prior knowledge on the network dynamics. Numerical experiments with realistic dynamic networks illustrate the usefulness of our method.

Keywords: Dynamic networks

1. INTRODUCTION

Estimation of dynamic networks have been of increasing interest in the systems and control community. Networks are used to efficiently model the signal flow or the dependencies between the different interconnected modules in a control system (Materassi and Innocenti, 2010). Given the set of signals which jointly evolve over time across the nodes of a network, it is often of interest to estimate how they influence each other (Chiuso and Pillonetto, 2012). In other words, one wishes to estimate the topology and dynamics of the network from the observed signals.

In the case of linear dynamic networks with J nodes or modules, the evolution of the signal w_i at the ith node over time t is modelled as (Shi et al., 2019; Van den Hof et al., 2018)

$$w_i(t) = \sum_{j=1, j \neq i} G_{ij}(q)w_j(t) + H_i(q)e_i(t), \qquad (1)$$

where G_{ij} denotes the transfer function of the module connecting the jth node to the ith node, H_i denotes the transfer function for the noise or innovation $e_i(t)$ at the ith node with variance σ_i^2, and q^{-1} denotes the unit delay operation. By concatenating the signal equations at all the J nodes, we have the complete model given by

$$\mathbf{w}(t) = \mathbf{G}(q)\mathbf{w}(t) + \mathbf{H}(q)\mathbf{e}(t),$$

where $\mathbf{G}(i,j) = G_{ij}$ and $\mathbf{H} = \mathrm{diag}(H_1, \cdots, H_J)$. The system matrix $\mathbf{G}$ reflects the underlying topology (presence or absence of connections) of the system: $G_{ij} \neq 0$ indicates an edge going from node j to node i. Let us further assume that the network is sparse in the number of edges (which is typically the case in practical systems), and that the generating dynamic process is stable.

Then the model in (1) may be equivalently described in the prediction error minimization framework (Van den Hof et al., 2018; Zorzi and Chiuso, 2015) as

$$w_i(t) = w_i(t|t-1) + e_i(t),$$

where

$$w_i(t|t-1) = \sum_{j=1}^{J} \sum_{k=1}^{+\infty} \theta_{ij}^{(k)} q^{-k} w_i(t) \qquad (2)$$

is the best one-step predictor of $w_i(t)$ given the past data until time $t-1$, where $\theta_{ij}^{(k)}$ denotes the kth impulse response coefficient relating the signal at node j to that at node i (Chiuso and Pillonetto, 2012). Collecting the signal values at all the J nodes upto time $t-1$ in the vector $\mathbf{a}_i(t)$, and the corresponding impulse response coefficients in the vector θ_i, the signal model at the ith node becomes

$$w_i(t) = \mathbf{a}_i(t)^\top \theta_i + e_i(t).$$

Since the network is sparse, θ_i is also sparse (block sparse, more specifically): $\theta_{ij}^{(k)} = 0$, $\forall k$ if $G_{ij} = 0$. Thus, learning both the topology and dynamics of the underlying network is equivalent to the estimation of the sparse θ_i for every node. While different approaches have been pursued to address this problem (Chiuso and Pillonetto, 2012; Jahandari and Materassi, 2018; Shi et al., 2019) they typically involve the estimation of hyperparameters or hyperpriors. As a result, they become demanding both in the amount of data and the computational complexity, specially when one deals with networks of moderate sizes.

Our goal in this work is to develop a method that learns a sparse linear dynamic network from the given network data in a computationally efficient and hyperparameter-free manner. We do this by building upon an iterative tuning-free covariance-matching based sparse estimation technique known as SPICE (Sparse Iterative Covariance Estimation) (Stoica et al., 2011; Zachariah and Stoica, 2015) Specifically, the highlights of our method are as follows:

- *No hyperparameter-tuning or cross-validation involved* unlike many approaches where the number of hyperparameters scales linearly with the number of nodes in the network
- *Computational efficiency* as a result of the parallel structure and no hyperparameter tuning involved
- *Minimal model assumptions* since we do not assume prior knowledge or structure on the predictor parameters
- *Accomodates varying noise levels across nodes (Heteroscedastic noise)* since our approach inherently treats every node independently

1.1 Related Work

Various approaches have been considered in the estimation of topology and dynamics in network systems. Frequency-domain methods have been proposed for identification of networks (Materassi and Innocenti, 2010; Shahrampour and Preciado, 2015) Module-based analysis and characterization of dynamic networks and conditions on identifiability under diverse system conditions have been studied (Everitt et al., 2016; Gevers and Bazanella, 2015; Hendrickx et al., 2019; Materassi and Salapaka, 2016; Van den Hof et al., 2013; Van den Hof et al., 2012). Regularized regression approaches which leverage on model sparsity have been explored(Bolstad et al., 2011; Jahandari and Materassi, 2018; Materassi et al., 2013; Yuan and Lin, 2006). By modelling the transfer function parameters as structured random variables, Bayesian approaches have been also developed (Chiuso and Pillonetto, 2012; Everitt et al., 2016; Everitt et al., 2018; Shi et al., 2019; Zorzi and Chiuso, 2015). Among these, the works of (Chiuso and Pillonetto, 2012) and (Shi et al., 2019) are the closest in context to our contribution. However, unlike our approach, these works involve computation of hyperparameters along with the use of explicit prior structure. We note that one of our motivations for using the SPICE framework comes from the recent work of (Venkitaraman and Zachariah, 2019), which employs SPICE in estimating static sparse partial correlation graphs from multivariate data.

2. THE LINEAR DYNAMIC NETWORK SYSTEM MODEL

As noted in the Introduction, an equivalent model for (1) in the prediction error minimization framework is given by

$$w_i(t) = w_i(t|t-1) + e_i(t),$$

where

$$w_i(t|t-1) = \sum_{j=1, j\neq i}^{J} \frac{G_{ij}(q)}{H_i(q)} w_j(t) + (1 - H_i(q)^{-1}) w_i(t),$$

is the best one-step predictor of $w_i(t)$ given the past data until time $t-1$. Then, by defining the impulse responses

$$(G_{ij}(q)/H_i(q)) \triangleq \sum_{k=1}^{+\infty} \theta_{ij}^{(k)} q^{-k}$$

$$(1 - H_i(q)^{-1}) \triangleq \sum_{k=1}^{+\infty} \theta_{ii}^{(k)} q^{-k},$$

we arrive at the predictor form in (2). In practice, however, one considers the finite-impulse response (FIR) approximation:

$$w_i(t|t-1) = \sum_{j=1}^{J} \sum_{k=1}^{K} \theta_{ij}^{(k)} q^{-k} w_i(t).$$

Notice that the summation over k is now over a finite window of K delays or past samples. The larger the K, better would be the model. Let $\mathbf{w}_i^N$ denote the vector containing N consecutive samples of w_i from t to $t + N$. Further, let us define the matrices $A_j \in \mathbb{R}^{N \times K}$ for $j = 1, \cdots, J$ as

$$A_j = \begin{bmatrix} w_j(t-1) & w_j(t-2) & \cdots w_j(t-K+1) \\ w_j(t) & w_j(t-2) & \cdots w_j(t-K) \\ \vdots & \vdots & \vdots \\ w_j(t+N-1) & w_j(t+N-2) & \cdots w_j(t+N-K) \end{bmatrix}.$$

Then, we have that where

$$\mathbf{w}_i^N = \sum_{j=1}^{J} A_j \theta_{ij} + \mathbf{e}_i^N$$

where $\theta_{ij} = [\theta_{ij}^{(1)}, \cdots, \theta_{ij}^{(K)}]^\top \in \mathbb{R}^{JK}$, and $\mathbf{e}_i^N$ denotes the noise or innovation over the N time samples. Equivalently, by defining $\mathbf{A}_i = [A_i \cdots A_J] \in \mathbb{R}^{N \times JK}$ and $\theta_i = [\theta_{i1} \cdots \theta_{iJ}]^\top$, we obtain the predictor model:

$$\mathbf{w}_i^N = \mathbf{A}_i \theta_i + \mathbf{e}_i^N,$$

where θ_i is sparse, since $\theta_{ij} = 0$ when there is no edge from node j to node i.

3. LEARNING SPARSE LINEAR DYNAMIC NETWORKS

Given the network system model, there are broadly two ways of approaching the network estimation problem. The first is to consider that the parameter θ_i is deterministic and unknown. Then, one of the most natural methods would be to solve

$$\min \sum_{i=1}^{J} \|\mathbf{w}_i^N - \mathbf{A}_i \theta_i\|_2^2 \text{ subject to } \|\theta_i\|_0 \le R_i \forall i \quad (3)$$

where $\|\|_0$ denotes the ℓ_0 (pseudo) norm which counts the number of non-zeros in the argument. Such an approach requires specification of the upper bounds R_i, and more importantly is an intractable combinatorial problem (Donoho et al., 2006). Hence, various convex relaxations of (3) are often used, such as the LASSO (least absolute shrinkage and selection operator) (Tibshirani, 1994; Yuan and Lin, 2006) which solve the following

$$\min \sum_{i=1}^{J} \left(\|\mathbf{w}_i^N - \mathbf{A}_i \theta_i\|_2^2 + \sum_j \lambda_{ij} \|\theta_{ij}\|_1 \right) \quad (4)$$

where λ_{ij} are the regularization hyperparameters corresponding to the jth node for the estimation at the ith node. We see that such approaches require cross validation and the number of hyperparameters scale with the number of nodes, if further simplifications such as assuming σ_i^2 to be the same across nodes are not employed.

The second approach to solving the problem is to assume that θ_i is a stochastic variable with a given prior and to treat the estimation in a completely Bayesian setting with hyperpriors or kernels. In such a case, it is usually assumed that the prior on θ_{ij} takes the form of kernels which have their own hyperparameters. The optimal estimate

is then obtained by maximizing the posterior probability. For certain choice of the priors, the Bayesian estimation becomes identical to solving (4). Nevertheless, one has to still compute for the hyperparameters even in this approach.

An alternative, which we pursue in this work, is to take an intermediate path by assuming that θ_i for every i is a random variable with expected value $\mathbf{0}$ and a diagonal covariance, prior to observing the data at time t. More precisely, we have

$$\mathrm{E}(\theta_i|t-1) = 0, \text{ and } \mathrm{Cov}(\theta_i|t-1) = \mathrm{diag}(\Pi_i)$$

where Π_i is the vector of variances of different components of θ_i. Then, the optimal estimate maximum aposteriori (MAP) estimate of θ_i is given by

$$\theta_i = (\mathbf{A}_i^\top \mathbf{A}_i + \sigma_i^2 \mathrm{diag}(\Pi_i))^{-1}\mathbf{A}_i^\top \mathbf{w}_i^N.$$

Then, the hyperparameters Π_i and σ_i^2 are estimated by matching the posterior covariance of $\mathbf{w}_i^N$ given by

$$\mathrm{Cov}(\mathbf{w}_i^N|t-1) = \mathbf{A}_i\mathrm{diag}(\Pi_i)\mathbf{A}_i^\top + \sigma_i^2\mathrm{diag}(\Pi_i),$$

with the sample covariance $\mathbf{w}_i^N(\mathbf{w}_i^N)^\top$. Such a covariance-matching treatment is well studied in sparse estimation problems. Further, the resulting estimation is known to be equivalent to solving the weighted square root LASSO (Belloni et al., 2011; Stoica et al., 2014):

$$\min_{\theta_i} \|\mathbf{w}_i^N - \mathbf{A}_i\theta_i\|_2^2 + \sum_j \sum_k \lambda_{jk}|\theta_{ij}^{(k)}|, \qquad (5)$$

where λ_{jk} is given by the ℓ_2 norm of the jkth column of $\mathbf{A}_j$. This approach of sparse estimation through covariance matching is known as SPICE (Sparse Iterative Covariance Estimation).

Thus, we advocate the use of the SPICE framework for learning sparse dynamic networks. As a result, our method offers the desirable properties for the problem:

- it is free from hyperparameter tuning − the effective hyperparameters are specific to the nodes and can hence accommodate noise levels that vary from one node to the other (heteroscedastic noise),
- it does not assume any explicit additional knowledge or prior on the parameters,
- it is computationally efficient (scales as $\mathcal{O}(JK^2N^2)$) and even allows for an online formulation (Zachariah and Stoica, 2015), and lastly,
- the formulation makes no assumptions on the network being directed or undirected.

3.1 On identifiability and uniqueness

We wish to highlight here that our approach estimates a sparse dynamical network entirely in a data-driven way from the given input and output data. As a result, a possible short coming of our approach is that it is not guaranteed theoretically to converge to the true network description. Indeed, it is known that the same input-output data may be associated with multiple network topologies and apriori network specific information in the form of identifiability conditions would be key in identifying the true topology and dynamics exactly (Gevers et al., 2017; Hayden et al., 2016; Weerts et al., 2018). We believe that it should be possible to modify our approach to explicitly incorporate apriori knowledge. This shall be the subject of our future study. However, we also wish to point out that our approach shows a promising performance in terms of the true positive rate and false positive rate in the numerical experiments which we have considered.

4. EXPERIMENTS

We next apply our approach to realistic dynamic network data in two experiments. We consider the following data generation strategy for both the experiments. We first randomly generate graphs of size $J = 8$ at sparsity level $\rho = 0.25$, that is, by considering that only 25% of the total number of possible edges are non-zero in the graph. Having generated the topology, we assign a transfer function over each non-zero edge (i, j) randomly as follows:

- Experiment 1: G_{ij} and H_is are FIR (finite impulse response) filters of length $K = 3$, where the impulse response taps are drawn randomly from the uniform distribution over $[0, 1]$ using the functions $drss(\cdot)$ and that generates a random stable discrete state-space system of a given order.
- Experiment 2: G_{ij}s and H_is are randomly generated stable rational functions with orders randomly chosen between 1 to 5.

The variances σ_i^2 are drawn randomly from the uniform distribution over $[0, 1]$. For a fixed instance of $\mathbf{G}$, we generate corresponding data as

$$\mathbf{w}(t) = [I - \mathbf{G}(q)]^{-1}\mathbf{e}(t)$$

where we consider only those realizations of transfer functions that result in a stable $[I - \mathbf{G}(q)]^{-1}$.

We note that while θ_i estimated by our method will have values very close to zero, they are never usually perfect zeros. This is an observation we share with most sparse estimation techniques. As a result, in order to get exact zeros, one must specify a tolerance or threshold $\delta > 0$ such that $\theta_j(hk)$ is set to zero if $|\theta_j(hk)| < \delta$. We note here that the thresholding operation is not an inherent part of our method, rather is typical to all graph/network estimation methods–specially when topology identification is the motive. In our experiments, we set $\delta = 10^{-1}$.

We measure the performance of our approach using three different metrics: True positive rate (TPR) which is the ratio of the edges rightly identified by the total number of true edges, the false positive rate (FPR) which is the ratio of the edges falsely identified to the ratio of zero-edges or no-edges, and lastly the distance (dis) defined as

$$dis = \sqrt{(FPR)^2 + (1 - TPR)^2}$$

which measures the deviation from the ground truth or oracle estimation which produces no false edges and identifies all existing edges. We wish to note here again that in general it cannot be claimed that our estimate will converge to the true underlying network dynamics. Nevertheless, our goal here is investigate if our approach produces a network estimate that approximates the true network reasonably well.

In Figure 1(a), we show the performance of our approach in terms of the three metrics as a function of the ratio of the available datasamples N to JK (the dimension of θ_i), averaged over all the J nodes and over 50 Monte-Carlo simulations, for Experiment 1. We observe that as
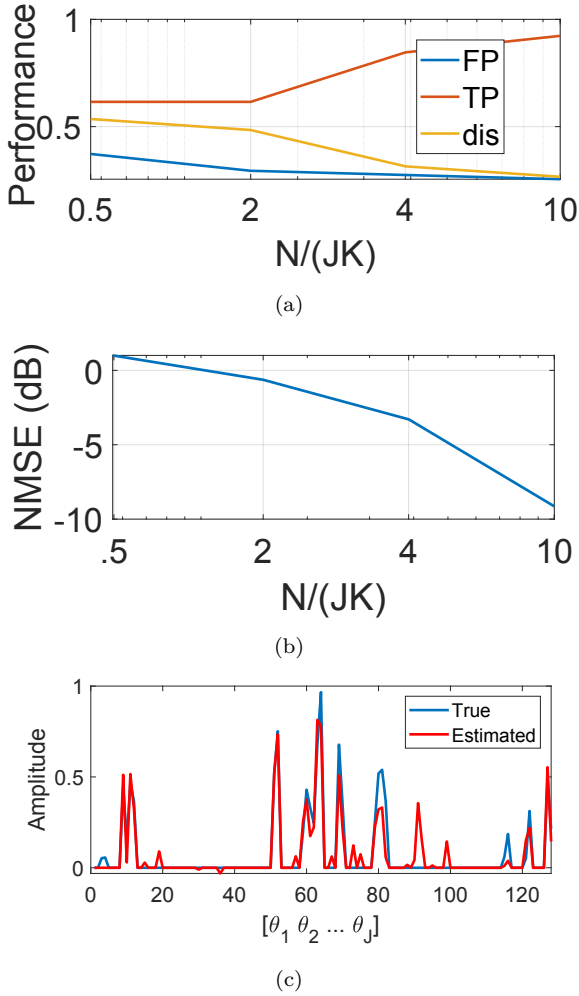
(a)

(b)

(c)

Fig. 1. Results for Experiment 1 (a) Performance, (b) NMSE in the values of the filter taps (c) An instance of the FIR filter taps estimated.
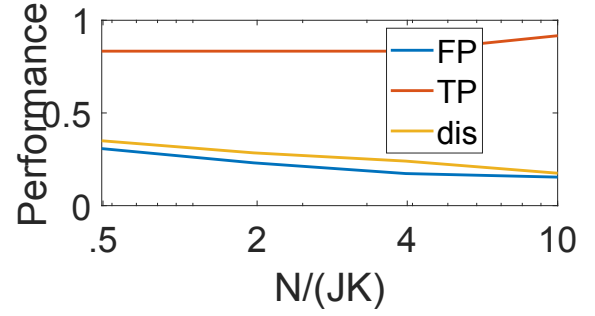

Fig. 2. Performance of our method for Experiment 2.

sparse linear dynamic system. We achieved this by building upon the hyperparameter-free covariance matching framework. Experiments with realistic linear dynamic systems revealed that our approach results in superior topology identification even with moderate number of available data. A possible drawback of our approach is that it does not currently guarantee the uniqueness of the estimated network. Our future work will focus on how our approach may be modified to incorporate identifiability conditions explicitly.

more data becomes available for estimation relative to the number of unknowns, all three metrics improve: TPR gets closer to unity, FPR closer to zero, and *dis* also tends to zero as N is increased. In order to evaluate the closeness of the estimated FIR taps with that of the ground truth, we evaluate the NMSE in estimation of θ_j averaged over all nodes, and over the Monte Carlo simulations. The NMSE is shown in Figure 1(b) and shows that NMSE improves progressively as N is increased from very small to large values. Figure 1(c), we show an instance of the true FIR impulse responses used to generate the data and those estimated by our method for $N = 1 \times JK$. We observe that our estimates for the FIR dynamics almost coincide with the ground-truth.

In Figure 2, we show the performance of our method for Experiment 2 in terms of the three metrics as N is increased. Once again, we observe the same trend that the estimation improves in terms of all the three metrics as N becomes large.

5. CONCLUSIONS

We proposed a computationally efficient and tuning-free method for learning the topology and dynamics of a

REFERENCES

Belloni, A., Chernozhukov, V., and Wang, L. (2011). Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika*, 791–806.

Bolstad, A., Van Veen, B.D., and Nowak, R. (2011). Causal network inference via group sparse regularization. *Trans. Sig. Proc.*, 59(6), 2628–2641.

Chiuso, A. and Pillonetto, G. (2012). A bayesian approach to sparse dynamic network identification. *Automatica*, 48(8), 1553–1565.

Donoho, D.L., Elad, M., and Temlyakov, V.N. (2006). Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inf. Theor.*, 52(1), 6–18.

Everitt, N., Bottegal, G., Rojas, C.R., and Hjalmarsson, H. (2016). Identification of modules in dynamic networks: An empirical bayes approach. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, 4612–4617.

Everitt, N., Bottegal, G., and Hjalmarsson, H. (2018). An empirical bayes approach to identification of modules in dynamic networks. *Automatica*, 91, 144 – 151.

Gevers, M. and Bazanella, A.S. (2015). Identification in dynamic networks: Identifiability and experiment design issues. In *2015 54th IEEE Conference on Decision and Control (CDC)*, 4005–4010.

Gevers, M., Bazanella, A.S., and Parraga, A. (2017). On the identifiability of dynamical networks. *IFAC-PapersOnLine*, 50(1), 10580 – 10585. 20th IFAC World Congress.

Hayden, D., Chang, Y.H., Goncalves, J., and Tomlin, C.J. (2016). Sparse network identifiability via compressed sensing. *Automatica*, 68, 9 – 17. doi: https://doi.org/10.1016/j.automatica.2016.01.008.

Hendrickx, J.M., Gevers, M., and Bazanella, A.S. (2019). Identifiability of dynamical networks with partial node measurements. *IEEE Transactions on Automatic Control*, 64(6), 2240–2253.

Jahandari, S. and Materassi, D. (2018). Topology identification of dynamical networks via compressive sensing. *IFAC-PapersOnLine*, 51(15), 575 – 580.

Materassi, D. and Innocenti, G. (2010). Topological identification in networks of dynamical systems. *IEEE Transactions on Automatic Control*, 55(8), 1860–1871.

Materassi, D., Innocenti, G., Giarré, L., and Salapaka, M. (2013). Model identification of a network as compressing sensing. *Systems & Control Letters*, 62(8), 664 – 672.

Materassi, D. and Salapaka, M.V. (2016). Graphoid-based methodologies in modeling, analysis, identification and control of networks of dynamic systems. In *2016 American Control Conference (ACC)*, 4661–4675.

Shahrampour, S. and Preciado, V.M. (2015). Topology identification of directed dynamical networks via power spectral analysis. *IEEE Transactions on Automatic Control*, 60(8), 2260–2265.

Shi, S., Bottegal, G., and Van den Hof, P.M.J. (2019). Bayesian topology identification of linear dynamic networks. *CoRR*, abs/1903.06205. URL http://arxiv.org/abs/1903.06205.

Stoica, P., Babu, P., and Li, J. (2011). SPICE: A sparse covariance-based estimation method for array processing. *IEEE Trans. Signal Processing*, 59(2), 629–638.

Stoica, P., Zachariah, D., and Li, J. (2014). Weighted SPICE: A unifying approach for hyperparameter-free sparse estimation. *Digital Signal Processing*, 33, 1–12.

Tibshirani, R. (1994). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58, 267–288.

Van den Hof, Dankers, A., Heuberger, P.S., and Bombois, X. (2013). Identification of dynamic models in complex networks with prediction error methods—basic methods for consistent module estimates. *Automatica*, 49(10), 2994 – 3006.

Van den Hof , P.M., Dankers, A.G., and Weerts, H. (2018). Identification in dynamic networks. *Computers & Chemical Engineering*, 109, 23 – 29.

Van den Hof, P.M.J., Dankers, A.G., Heuberger, P.S.C., and Bombois, X. (2012). Identification in dynamic networks with known interconnection topology. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, 895–900.

Venkitaraman, A. and Zachariah, D. (2019). Learning sparse graphs for prediction of multivariate data processes. *IEEE Signal Process. Lett.*, 26(3), 495–499.

Weerts, H.H., Van den Hof, P.M., and Dankers, A.G. (2018). Identifiability of linear dynamic networks. *Automatica*, 89, 247 – 258. doi: https://doi.org/10.1016/j.automatica.2017.12.013.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1), 49–67.

Zachariah, D. and Stoica, P. (2015). Online hyperparameter-free sparse estimation method. *IEEE Trans. Signal Processing*, 63(13), 3348–3359. doi: 10.1109/TSP.2015.2421472.

Zorzi, M. and Chiuso, A. (2015). A bayesian approach to sparse plus low rank network identification. In *2015 54th IEEE Conference on Decision and Control (CDC)*, 7386–7391.