High Precision Variational Bayesian Inference of Sparse Linear Networks *

Junyang Jin a, Ye Yuan b,*, and Jorge Gonçalves a

a The Luxembourg Centre for Systems Biomedicine, avenue du Swing, 4367 Belvaux, Luxembourg
b School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, 430074, Hubei, China

Abstract
Sparse networks can be found in a wide range of applications, such as biological and communication networks. Inference of such networks from data has been receiving considerable attention lately, mainly driven by the need to understand and control internal working mechanisms. However, while most available methods have been successful at predicting many correct links, they also tend to infer many incorrect links. Precision is the ratio between the number of correctly inferred links and all inferred links, and should ideally be close to 100%. For example, 50% precision means that half of inferred links are incorrect, and there is only a 50% chance of picking a correct one. In contrast, this paper infers links of discrete-time linear networks with very high precision, based on variational Bayesian inference and Gaussian processes. Our method can handle limited datasets, does not require full-state measurements and effectively promotes both system stability and network sparsity. On several of examples, Monte Carlo simulations illustrate that our method consistently has 100% or nearly 100% precision, even in the presence of noise and hidden nodes, outperforming several state-of-the-art methods. The method should be applicable to a wide range of network inference contexts, including biological networks and power systems.

Key words: System Identification; Variational Inference; Dynamical Structure Function; Network Inference; Sparse Networks

1 INTRODUCTION
In systems biology, mathematical modelling has been central to the study of biological networks. Dynamical models are frequently developed to predict the behaviour of systems in response to external or internal stimulus for example, drug treatment or mutation. Yet only input-output dynamics are learned without exploring the topology, whereas in many other applications, comprehensive knowledge of the network topology is required. For example, the information about the structure of control systems is essential for fault diagnosis. Hence, both the inference of system dynamics and the detection of network topology are important.

Sparsity and stability are fundamental properties of most real-world networks. Communication networks, as artificial systems, are designed to be stable for robust operation and sparse to reduce energy consumption. Thus, sparsity and stability are primary constraints in network inference, especially in the case of limited data source or high amount of noise. When dealing with practical networks, often, not all the nodes in the network can be measured, because of either high experimental cost or technical limitations. The difficulty here is that many inference methods nonetheless assume full-state measurements. It is important to reconsider this issue carefully.

In recent years, kernel-based methods have become popular in the system identification community [13]. For linear systems, the methods effectively impose system stability and greatly simplify the estimation of model complexity. Kernel-based methods have successfully identified SISO continuous linear time invariant (LTI) models [13] and been further extended to discrete LTI systems [4, 5, 13]. In particular, kernel-based methods have been used to infer sparse networks described by Granger causality [6]. They have been further developed to infer the so-called sparse plus low rank networks where it is assumed that the majority of nodes can be described by a few other components that are not accessible for observation [19]. Nevertheless, kernel-based methods are not ideal for topology detection. Due to local optimal solutions, it is very difficult to achieve 100% precision.

Variational inference (VI), as empirical Bayes, is a Bayesian deterministic approximation technique that has been applied to a number of cases, including sparse regression models [1] and neural networks [2]. With well-posed models of exponential families, VI is as efficient computationally as empirical Bayes [9]. Whilst rigorous evaluation remains elusive, Monte Carlo simulations show that VI can be more accurate than empirical Bayes [9, 10]. More importantly, VI is able to estimate model evidence for each possible model structure: this

* For correspondence: yye@hust.edu.cn.
enables evaluation of the relative confidence between models.

This paper combines Gaussian processes and VI to infer sparse networks\footnote{A full version of this brief paper can be found in [11].} Dynamical structure functions are used to describe sparse linear networks where the information of hidden nodes is encoded via transfer functions. VI is employed to identify system dynamics and infer network topology. Moreover, by applying backward selection strategies, the proposed method encourages high inference precision. Monte Carlo simulations show that our method produces more reliable networks than kernel-based methods under various experimental conditions, such as different topologies, noise levels, kernel functions and number of data points. Most importantly, the proposed method always achieves 100% or nearly 100% precision so that almost all inferred links are correct.

The paper is organized as follows. Section 2 introduces dynamical structure function and formulates the full Bayesian model. Section 3 discusses network inference using variational inference and analyses algorithm properties. Section 4 compares the method with other approaches via Monte Carlo simulations. Finally, Section 5 concludes and discusses further developments in this field.

## 2 MODEL FORMULATION

### 2.1 The dynamical structure function

The sparse network of \( n \) nodes is described by a linear state space model as follows:

\[
\begin{align*}
x(t+1) &= Ax(t) + Bu(t) + Be(t) \\
y(t) &= Cx(t),
\end{align*}
\]

where \( x \in \mathbb{R}^n \) are states of the system, each of which represents a node. \( u \in \mathbb{R}^m \) denote inputs. \( y \in \mathbb{R}^p \) present the measurements of the states. \( e \in \mathbb{R}^q \) are i.i.d white Gaussian noise with zero mean and covariance matrix \( P_e \). Without loss of generality, \( P_e \) is assumed to be diagonal. \( A \in \mathbb{R}^{n \times n}, B_u \in \mathbb{R}^{n \times m}, B_e \in \mathbb{R}^{n \times q} \) and \( C \in \mathbb{R}^{n \times p} \) are system matrices.

It is normal in practice that some of the nodes are unobservable (hidden states). Assume the first \( p < n \) states are observable (i.e. \( C = [I, 0] \)). To avoid inferring hidden states, they are removed from the model. Dynamical structure functions (DSF) encode the information of hidden states via transfer functions [17]:

\[
Y = QY + PU + HE,
\]

where \( Q, P \) and \( H \) are transfer matrices, each element of which is a strictly proper transfer function, indicating that the network is a causal system [16]. Matrix \( Q \) implies the connectivity among observable nodes, whose diagonal elements are zero. \( P \) and \( H \) matrices relate inputs and process noise to nodes, respectively. The topology of the network (i.e. model structure) is reflected by the zero structure of these three matrices. For example, if \([Q]_{ij} = 0\), the \( j \)th node does not control the \( i \)th node. Model structures are denoted by \( \mathcal{M}_0 \) and \( \mathcal{M}_p \) presents the number of links. In particular, \( \mathcal{M}_0 \) represents the fully-connected topology. The internal dynamics of the network are described by the transfer functions. The order of a transfer function is relevant to the number of hidden states involved in that regulation pathway.

The input-output map of the network is associated with the DSF as \( Y = G_u U + G_e E \) where \( G_u = (I - Q)^{-1} P \) and \( G_e = (I - Q)^{-1} H \). Ideally, the input-output map can be perfectly recovered based on measurement data. Nevertheless, the corresponding DSF may not be unique, meaning that the network topology is unidentifiable. To ensure the inference problem is well-posed, additional constraints are imposed [8]. A sufficient condition for network identifiability is that matrix \( H \) is diagonal so that \( p - 1 \) elements in each column of \([Q, P, H]'\) are known to be zero. In what follows, we make following assumptions so that no prior knowledge of matrix \( P \) (structure of input channels) is required.

**Assumption 1** Noise matrix \( H \) is diagonal, monic (\( \lim_{q \to \infty} qH = I \)) and minimal phase.

Stability and sparsity are the basic nature of many practical networks such as biological networks and power systems. Therefore, we assume the target network is stable and sparse.

**Assumption 2** Each elements of transfer matrices, \( Q \) and \( P \) are stable. Matrices \( Q \) and \( P \) are sparse.

### 2.2 The likelihood distribution

After simple manipulations, the DSF in (2) can be rewritten as:

\[
Y = F_y Y + F_u U + \bar{E}.
\]

where

\[
F_u = (qH)^{-1} P, \quad F_y = I - (qH)^{-1} (I - Q), \quad \bar{E} = q^{-1} E.
\]

Model (3) is a valid causal system. According to the assumptions, transfer matrices, \( F_u \) and \( F_y \) are stable. In addition, since \( H \) is diagonal, \( F_u \) and \( F_y \) have the same zero structure as \( P \) and \( Q \). As a result, \( F_u \) and \( F_y \) are sparse matrices and imply the network topology.

Model (3) is expressed in a non-parametric way. By doing so, the selection of model complexity is avoided and, more importantly, system stability can be promoted effectively. The dynamical system for the \( i \)th target node, is formulated as \( y_i(t) = \sum_{j=1}^{p} \sum_{k=1}^{\infty} h_{ij}^u(k) y_j(t - k) + \sum_{j=1}^{m} \sum_{k=1}^{\infty} h_{ij}^e(k) y_j(t - k) + \epsilon_i(t) \) where \( h_{ij}^u \) and \( h_{ij}^e \) are the impulse responses of transfer functions \([F_y]_{ij}\) and
functions are Tuned/Correlated kernels (TC kernel). TC kernel has been widely used to characterize stable impulse responses [3]. Other valid kernels include Diagonal/Correlated kernel (DC kernel)\(^2\) and second order stable spline kernel (SS kernel)\(^3\). As a result, the prior distribution for \(w\) is:

\[
p(w|\lambda, \beta, \sigma, \mathcal{M}_k) = \prod_{q=1}^{L} \prod_{i=1}^{M_k} \mathcal{N}(w_{q,i}|0, \sigma^{-1}I_{K_{q,i}}),
\]

where \(\beta\) are hyperparameters of TC kernels, which control the exponential decaying rate of impulse responses. \(\lambda\) are scale variables of the kernel functions. In the Bayesian model, \(\lambda\) influence the probability of model structure (i.e. network topology). As \(\lambda_i\) approaches infinity, distribution \(p(w_i|\lambda_i, \beta_i, \sigma)\) deploys an impulse at the origin, enforcing zero impulse responses. In this case, the \(i\)th node or input does not control the target node. As the standard setting of variational inference, noise precision \(\sigma^2\) is also used to scale the covariance matrix.

\[
K_{q,i} \in \mathcal{R}_{T_q \times T_q}, \lambda = [\lambda_1, \ldots, \lambda_{M_k}], \beta = [\beta_1, \ldots, \beta_{M_k}]\quad \text{and}\quad K_{q,i}|_{t,s} = k(t, s; \beta_i), k(t, s; \beta_i) = \beta_i^{\max}(t, s) < 0 < \beta_i < 1, \lambda_i > 0.
\]

Since \(\sigma\) is non-negative, the Gamma distribution is assigned as its conjugate prior. Without specific preference on \(\sigma\), parameters \(a_0, b_0\) of the distribution are set to 0.001, resulting in a non-informative prior: \(p(\sigma|a_0, b_0) = \text{Gamma}(\sigma|a_0, b_0)\). In addition, hyperpriors are assigned to hyperparameters to complete the hierarchy. For hyperparameter \(\lambda_i\), the Gamma distribution is applied as the conjugate prior. Similar to \(\sigma\), a non-informative prior is adopted. For hyperparameter \(\beta_i\), the uniform distribution on (0, 1) is employed as the prior, i.e., \(p(\beta_i) = 0 < \beta_i < 1\).

### 2.4 The full Bayesian model

By incorporating the likelihood and prior distributions, the full Bayesian distribution is:

\[
p(w, \sigma, \lambda, \beta|\mathcal{Y}, \mathcal{M}_k) \
\propto \prod_{q=1}^{L} \prod_{i=1}^{M_k} \left\{ (2\pi\sigma^{-1})^{-\frac{N_q-T_q}{2}} \exp\left\{-\frac{\sigma}{2} \| Y_q - \Phi_q w_q \|_2^2 \right\} \right\}
\]

\[
\times |2\pi\sigma^{-1}K_{q,i}|^{-\frac{1}{2}} \exp\left\{-\frac{\sigma}{2} w_q'K_{q,i}^{-1}w_q \right\}
\]

\[
\times \frac{b_0^{a_0}}{\Gamma(a_0)} \sigma^{a_0-1} e^{-b_0\sigma} \prod_{i=1}^{M_k} \frac{b_0^{a_0}}{\Gamma(a_0)} \lambda_i^{a_0-1} e^{-b_0\lambda_i},
\]

where \(K_{q,i} = \text{blkdiag}\{K_{q,1}, \ldots, K_{q,M_k}\}\) and \(\Lambda_q = \text{diag}\{\Lambda\} \otimes I_{T_q}\).

---

\(^2\) DC kernel \((t, s; \beta_1, \beta_2) = \beta_1^{(t+s)} + \beta_2^{(t-s)}, \beta_1, \beta_2 \in (0, 1)\) and \(\beta_2 \in (-1, 1)\)

\(^3\) SS kernel \((t, s; \beta) = \frac{\beta^{(t+s)}}{2} - \frac{\beta^{\max(t, s)}}{6}, \beta \in (0, 1)\)
3 VARIATIONAL INFERENCE OF DYNAMICAL STRUCTURE FUNCTIONS

3.1 Update of random quantities

For each model structure $M_k$, the corresponding full Bayesian model, $p(w, \sigma, \lambda, \beta | Y, M_k)$ is approximated by a candidate distribution, $Q(w, \sigma, \lambda, \beta | M_k)$ using the mean field factorization:

$$Q(w, \sigma, \lambda, \beta | M_k) = q(w, \sigma | M_k)q(\lambda | M_k)q(\beta | M_k),$$

where $Q(\cdot)$ and $q(\cdot)$ are valid probability distributions. Hereafter, the symbol, $M_k$ is suppressed to simplify the notation.

The factors of (10) are formulated in sequence. In what follows, the terms independent on the random variables of the factor under consideration are ignored for convenience. To begin with, factor $q(w, \sigma)$ is solved as the Gaussian-Gamma distribution:

$$\ln q(w_q, \sigma) = \ln N(w_q | \mu_q, \sigma^{-1} \Sigma_q) - \ln \text{Gamma}(\sigma | a^\sigma, b^\sigma),$$

where

$$\Sigma_q^{-1} = \Phi_q) + E(\lambda_q)E(\lambda^{-1}_q), \quad \mu_q = \Sigma_q \Phi_q \lambda_q, a^\sigma = 2 \sum_{q=1}^{L} N_q - T_q + a_0, b^\sigma = b_0 + \frac{1}{2} \sum_{q=1}^{L} (Y_q \sigma_q + \mu_q \sigma_q^{-1} \mu_q), \quad E(\sigma_q, w_q, w'_q) = \frac{a^\sigma}{b^\sigma} \mu_q t_q + \Sigma_q, \quad E(\sigma, w_q) = \frac{a^\sigma}{b^\sigma} \mu_q.$$

Following the same procedure, factor $q(\lambda)$ is solved as independent Gamma distributions:

$$\ln q(\lambda) = \ln \text{Gamma}(\lambda | a^\lambda, b^\lambda),$$

where

$$a^\lambda = \sum_{q=1}^{L} T_q + a_0, \quad b^\lambda = b_0 + \frac{1}{2} \text{trace} \left[ \sum_{q=1}^{L} E(\lambda^{-1}_q) E(w, \sigma | w'_q) \right].$$

Finally, factor $q(\beta)$ is calculated. Unlike the other random variables, factor $q(\beta)$ has no closed-form expression. Nevertheless, the elements of $\beta$ are independently distributed as:

$$q(\beta) = c_i \prod_{q=1}^{L} |K^{-1}_{q,i}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} E(\lambda) \text{trace} \left[ K^{-1}_{q,i} E(w, \sigma | w'_q) T_{q,i} \right] \right\}.$$

where $c_i$ is the unknown normalization constant.

3.2 Lower bound of model evidence

Combining all the factors above, the lower bound, $L(Q(w, \sigma, \lambda, \beta | M_k))$ of model evidence $p(Y | M_k)$ is:

$$L(Q(w, \sigma, \lambda, \beta | M_k)) = \frac{1}{2} \sum_{q=1}^{L} \ln |\Sigma_q| - \left[ \frac{a^\sigma}{b^\sigma} ||Y_q - \Phi_q \mu_q||^2 + \text{trace}(\Phi_q \Sigma_q \Phi'_q) \right]$$

$$- b_0 \frac{a^\sigma}{b^\sigma} - a^\sigma \ln b^\sigma - \sum_{i=1}^{M_k} a^\lambda_i \ln b^\lambda_i + \frac{M_k \sum_{q=1}^{L} T_q}{2} + M_k (a_0 \ln b_0 - \ln \Gamma(a_0))$$

$$- \sum_{i=1}^{M_k} b^\lambda_i - a^\lambda_i - \ln \Gamma(a^\lambda_i) + \ln c_i,$$

where the constant terms independent on $M_k$ are ignored.

3.3 Algorithm for variational inference

The VI algorithm is summarized in Algorithm 1. Since $q(\beta)$ is only known up to a normalization constant, $E(K^{-1})$ cannot be calculated analytically. Hence, Metropolis-Hastings sampling methods are used to estimate the expectation. With $N$ samples $\{B^k | k = 1, \ldots, N\}$, $E(K^{-1})$ is estimated as $E(K^{-1}) = \frac{1}{N} \sum_{k=1}^{N} (K^{-1}_q)$. In addition, in order to calculate the lower bound of model evidence, one has to estimate the normalization constant, $c_i$ of $q(\beta)$ in (15). Since $\beta_i$ is a scalar on $(0, 1)$, numeric integration methods (e.g. adaptive quadrature [15]) are sufficient to give an accurate estimation given $\int q(\beta_i) d\beta_i = 1$. The theoretical analysis of the algorithm can be found in [11].

Model selection is based on the posterior distribution of model structures (i.e. $p(M_k | y) \propto p(y | M_k) p(M_k)$). Assuming equal probability for each model structure, the distribution is proportional to model evidence, $p(M_k | y) \propto p(y | M_k)$.

The by-product of VI provides a
lower bound of model evidence, which can be used to determine the most probable model structure (network topology). We come up with a heuristic strategy that applies backward selection to select model structures as shown in Algorithm 2.

Algorithm 2 Variational inference of network topology

1: Implement Algorithm 1 with model structure $M_0$.
2: Set the threshold:
3: $R = a s e c \{ \sum_q \| w_{q,1} \|, \ldots, \| w_{q,n+p} \| \} $.
4: for $k = 1:n+p+1$ do
5: $I = \{ i | \sum_q \| w_{q,i} \| \leq R_k \} $.
6: Remove the links in index set $I$, producing $M_k$.
7: Run Algorithm 1 with $M_k$.
8: Store the value of $L(q(w, \sigma, \lambda, \beta | M_k))$.
9: end for.
10: $M_{opt} = \arg\max_{M_k} L(q(w, \sigma, \lambda, \beta | M_k))$.

4 SIMULATION

We conducted Monte Carlo simulations and compared our method with another kernel-based system identification approach. The state-of-the-art method [6] applies empirical Bayes to estimate hyperparameters and to detect network topology, where DC (Kernel_DC), SS (Kernel_SS) and TC (Kernel_TC) kernels were all used for inference. The DSF networks for test were generated randomly with diverse topologies (including an extremely sparse type, a ring structure), simulated under various noise levels and inferred using time-series data of different lengths.

Two criteria are applied to evaluate the performance of algorithms, namely True Positive Rate (TPR) and Precision (PREC). TPR shows the percentage of the true links in the ground truths that were successfully inferred. TPR implies the information richness of the inference result. PREC equals to the rate of the correct links over the all inferred. PREC indicates the reliability or accuracy of the inferred network. Hence, ensuring a high PREC is the first priority in real applications. If, for example, PREC is below 50%, one cannot tell which inferred links are correct. To investigate whether the inferred networks have internal dynamics consistent with the ground truths, the estimated models were simulated to predict the validation dataset that was not used for inference.

4.1 Random DSF networks

100 networks were generated with random topologies and internal dynamics. All networks contained 15 nodes in total. Each node was independently driven by an input that was known and process noise that was assumed to be unknown during inference. DSF models were produced from state space models (1). To be specific, a sparse stable matrix $A \in \mathbb{R}^{15 \times 15}$ was first yielded randomly using the function $\text{sprandn}(n, n, \text{density})$ in Matlab. The brute force strategy was applied to guarantee that matrix $A$ was Hurwitz (that is, no eigenvalue was outside the unit circle of the complex plane) and that no isolated nodes existed in the network.

To simulate the models, inputs and process noise were both i.i.d. white Gaussian signals. The variance of inputs was fixed to 1 whilst that of process noise varied. The Signal-to-Noise ratio is defined as $SNR = 10 \log_{10} \frac{\sigma_u}{\sigma_e}$, where $\sigma_u$ and $\sigma_e$ are signal variance of inputs and noise, respectively. The first 10 states of the models were measured, leaving the rest 5 as hidden nodes. The length of impulse responses after truncation was set to 20. The data for inference were collected with different lengths.

The average TPR and PREC over 100 trials are recorded in Table 1-3. In the best-case scenario (no noise), VI outperforms all the other methods. In particular, almost all the inferred links of VI are correct ($PREC \approx 100\%$), regardless of the number of data points. Meanwhile, VI is able to identify most true links of the ground truths.

With 85 data points, VI recovers the network perfectly. In contrast, Kernel_TC presents the weakest result. PREC of Kernel_TC stays low unless more data points are available for inference. The poor performance of Kernel_TC indicates the effectiveness of VI that uses the same kernel function.

As $SNR$ decreases to 10dB, all methods require more data points to counteract the interference of process noise. Although TPR of VI is slightly lower than Kernel_SS, PREC of VI is much higher than Kernel_SS, which is close to 100%. Kernel_TC also achieves accurate results. Nevertheless, many true links are missed ($TPR < 80\%$). In addition, networks were inferred using prediction error minimization (PEM) without imposing sparsity. It turns out that the inferred networks were always fully connected.

It is remarkable that the inferred networks of VI are highly reliable ($PREC \approx 100\%$) even under the worst-case scenario (that of pure noise). The lack of data points only affects TPR of VI whilst PREC remains high. The gap of TPR between Kernel_SS, Kernel_DC and VI decreases gradually as more data points are available. Similar to the above cases, Kernel_TC is outperformed by VI.

Table 1

| No Noise | 45 | 65 | 85 |
|----------|----|----|----|
| Kernel_DC | 94.3 | 56.9 | 97.7 | 73.1 | 97.9 | 85.2 |
| Kernel_SS | 84.7 | 58.3 | 91.4 | 91.7 | 100 | 100 |
| Kernel_TC | 43.3 | 16.5 | 71.1 | 23.4 | 98.3 | 42.9 |
| VI | 100 | 75.0 | 99.7 | 98.5 | 100 | 100 |

4.2 Ring networks

100 networks with the fixed ring structure were generated and simulated following the same protocol in the last section. Each node was driven by independent process noise. Only one input entered the network through a
single node. Since the network contains a feedback loop and is extremely sparse, it is more challenging to infer.

Table 4 presents the inference result. VI is still able to produce reliable networks with the highest PREC among all the methods (PREC = 100%). More importantly, PREC of the other three cases highly relies on the number of data points whilst that of VI does not. TPR of VI and Kernel_SS is very close. Since the ring network contained only 10 links, VI at most missed 3 true links.

Table 4
| PREC | TPR | PREC | TPR | PREC | TPR |
|------|-----|------|-----|------|-----|
| 100  | 77.6| 72.9 | 89.1| 75.6 | 99.9| 79.1|
| 200  | 64.3| 69.6 | 81.7| 76.1 | 88.9| 77.1|
| 300  | 79.9| 60.3 | 87  | 65.7 | 93.4| 70.9|
| 500  | 100 | 56.5 | 100 | 65  | 100 | 74.9|

5 CONCLUSION AND DISCUSSION

This paper has applied variational inference to identify DSF models given measured time series data. No prior knowledge of the hidden nodes including their number and internal connectivity is required. Both the system dynamics and topology of sparse linear networks can be inferred accurately. The proposed method achieves this by incorporating kernel-based methods to promote system stability and by introducing VI to imposing network sparsity. Monte Carlo simulations imply that our method always produces reliable inference result regardless of challenging experimental conditions (for example, lower number of data points, high levels of noise, and extremely sparse topologies). The inference of links is highly accurate (with nearly 100% confidence): only a few true links are missed. Therefore, the developed method appears highly reliable for real-world applications.

Overall, the value of this approach is that it outperforms kernel-based methods at least in regard to TC kernel. Our method raises the reliability of inference results to the highest level (100% PREC). In particular, out method is applicable to real-world networks where full state measurements are normally unavailable such as gene regulatory networks.

Future developments should take nonlinearity into account. Considering most real-word networks are nonlinear, it is necessary to extend our framework to black-box nonlinear systems.

References

[1] C. Bishop. Pattern recognition and machine learning. Springer New York, 2006.
[2] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. Variational inference: A review for statisticians. Journal of the American Statistical Association, 112(518):859–877, 2017.
[3] G. Bottegal, A. Y. Aravkin, H. Hjalmarsson, and G. Pillonetto. Robust em kernel-based methods for linear system identification. Automatica, 67:114 – 126, 2016.
[4] T. Chen, M.S. Andersen, L. Ljung, A. Chiuso, and G. Pillonetto. System identification via sparse multiple kernel-based regularization using sequential convex optimization techniques. Automatica, 59(11):1–33, 2014.
[5] T. Chen, H. Ohlsson, and L. Ljung. On the estimation of transfer functions, regularizations and gaussian processes revisited. Automatica, 48(8):1525 – 1535, 2012.
[6] A. Chiuso and G. Pillonetto. A Bayesian approach to sparse dynamic network identification. Automatica, pages 1553–1565, 2012.
[7] F. Dinuzzo. Kernels for linear time invariant system identification. SIAM Journal on Control and Optimization, 53(5):3299–3317, 2015.
[8] J. Gonçalves and S. Warnick. Necessary and sufficient conditions for dynamical structure reconstruction of Iti networks. IEEE Trans. Automat. Contr., 53(7):1670–1674, 2008.
[9] M. Hoffman, D.M. Blei, C. Wang, and J. Paisley. Stochastic Variational Inference. ArXiv e-prints, June 2012.
[10] W. R. Jacobs, T. Baldacchino, T. J. Dodd, and S. R. Anderson. Sparse bayesian nonlinear system identification using variational inference. IEEE Transactions on Automatic Control, pages 1–1, 2018.
[11] J. Jin, Y. Yuan, and J. Gonçalves. High precision variational bayesian inference of sparse linear networks. CoRR, abs/1901.00673, 2019.
[12] G. Pillonetto, A. Chiuso, and G.D. Nicolao. Prediction error identification of linear systems: A nonparametric gaussian regression approach. Automatica, 47(2):291 – 305, 2011.
[13] G. Pillonetto, F. Dinuzzo, T. Chen, G. De Nicolao, and L. Ljung. Kernel methods in system identification, machine learning and function estimation: A survey. Automatica, 50(3):657–682, 2014.
[14] C. E. Rasmussen and C. K. I. Williams. Gaussian processes for machine learning. The MIT Press, 2006.
[15] L.F. Shampine. Vectorized adaptive quadrature in matlab. Journal of Computational and Applied Mathematics, 211(2):131 – 140, 2008.
[16] Y. Yuan, A. Rai, E. Yeung, G. Stan, S. Warnick, and J. Gonçalves. A minimal realization technique for the dynamical structure function of a class of lti systems. IEEE Transactions on Control of Network Systems, 4(2):301–311, June 2017.
[17] Y. Yuan, G. Stan, S. Warnick, and J. Gonçalves. Robust dynamical network structure reconstruction. Automatica, 47:1230–1235, 2011.
[18] Z. Zhang and B. Rao. Extension of sbl algorithm for the recovery of block sparse signals with intra-block correlation. IEEE Transactions on Signal Processing, 61:2009–2015, 2013.
[19] M. Zorzi and A. Chiuso. Sparse plus low rank network identification: A nonparametric approach. Automatica, 76:355–366, 2016.