

Graph Topology Inference with Derivative-Reproducing Property in RKHS: Algorithm and Convergence Analysis

Mircea Moscu, Ricardo A Borsoi, Cédric Richard, José-Carlos M Bermudez

▶ **To cite this version:**

Mircea Moscu, Ricardo A Borsoi, Cédric Richard, José-Carlos M Bermudez. Graph Topology Inference with Derivative-Reproducing Property in RKHS: Algorithm and Convergence Analysis. IEEE Transactions on Signal and Information Processing over Networks, 2022, 8, pp.78-91. 10.1109/TSIPN.2022.3146050 . hal-03347316

HAL Id: hal-03347316
https://hal.science/hal-03347316

Submitted on 17 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Graph topology inference with derivative-reproducing property in RKHS: algorithm and convergence analysis

Mircea Moscu, Ricardo A. Borsoi, Cédric Richard, José-Carlos M. Bermudez

arXiv:2104.13687v1 [eess.SP] 28 Apr 2021

***Abstract*—In many areas such as computational biology, finance or social sciences, knowledge of an underlying graph explaining the interactions between agents is of paramount importance but still challenging. Considering that these interactions may be based on nonlinear relationships adds further complexity to the topology inference problem. Among the latest methods that respond to this need is a topology inference one proposed by the authors, which estimates a possibly directed adjacency matrix in an online manner. Contrasting with previous approaches based on linear models, the considered model is able to explain nonlinear interactions between the agents in a network. The novelty in the considered method is the use of a derivative-reproducing property to enforce network sparsity, while reproducing kernels are used to model the nonlinear interactions. The aim of this paper is to present a thorough convergence analysis of this method. The analysis is proven to be sane both in the mean and mean square sense. In addition, stability conditions are devised to ensure the convergence of the analyzed method.**

***Index Terms*—convergence analysis, kernel least mean squares, nonlinear topology inference, partial derivative sparsity**

I. Introduction

Knowledge of the network topology in applications such as gene regulation systems [1], socio-economic interactions [2], or brain activity [3] is of utmost importance. The main reason is that processing these data via, e.g., filtering [4], [5], [6], [7], [8] or spectral analysis [9], [10], requires the network structure [11]. Since most graph signal processing algorithms assume that the graph topology is known beforehand, significant progress has been made recently in the estimation of the graph topology from available data, be them links between regions of the brain, genes in a network, or sectors of a market economy. Moreover, the presence of nonlinear interactions in real-world applications led to the development of more general algorithms. One such proposition is the graph topology inference algorithm recently proposed by the authors [12], which is based on derivative-reproducing property to infer nonlinear interactions and promote sparsity. The current paper builds upon this preliminary work by proposing an in-depth and complex stability and performance analysis of this algorithm. The ability of reproducing kernels to model nonlinear relationships between nodal signals is employed in the graph inference process, alongside with kernel dictionaries to mitigate the increasing number of kernel functions due to the online setting, while distributing the computational burden over the agents in the network. Many real world examples, such as social graphs [13], show considerable edge sparsity, which needs a sparsity-inducing framework based on derivative-reproducing kernels. The present paper briefly recalls the considered algorithm and follows up with a complete convergence analysis, both in the mean and mean-square sense.

This work was funded in part by ANR under grant ANR-19-CE48-0002.

M. Moscu is with the Université Côte d'Azur, Nice, France (e-mail: mircea.moscu@oca.eu), Lagrange Laboratory (CNRS, OCA).

R.A. Borsoi is with the Department of Electrical Engineering, Federal University of Santa Catarina (DEE–UFSC), Florianópolis, SC, Brazil, and with the Lagrange Laboratory, Université Côte d'Azur, Nice, France. e-mail: raborsoi@gmail.com.

C. Richard is with the Université Côte d'Azur, Nice, France (e-mail: cedric.richard@unice.fr), Lagrange Laboratory (CNRS, OCA).

J.-C.M. Bermudez is with the DEE–UFSC, Florianópolis, SC, Brazil. e-mail: j.bermudez@ieee.org.

A. Prior works

Early advances in topology inference are put forward in [14], where log-likelihood-based method of covariance estimation is introduced. Similarly, in [15] the graphical Lasso is employed in order to estimate the precision matrix from the available data. In [16], the authors advocate that connectivity can be recovered from estimated spectral templates, while [17] adapts the method for band-limited signals, i.e., signals whose Graph Fourier Transform is sparse. The authors of [18] introduce a method designed for topology inference for the case when the measured signals are non-stationary. An online adaptive algorithm is developed in [19], where the authors assume a linear model governing the interactions. Linearity of interactions and signal stationarity are assumed in [20] to devise an ADMM algorithm.

The modeling of nonlinear interactions has proved to be a relevant research topic. Some existing methods of higher-order link inference yield satisfactory results when applied in optimizing power grids [21] or in predicting social behavior [22]. In the presence of nonlinear phenomena, works such as [23], [24] focus on polynomial structural equation models, while the authors of [25] use their nonlinear counterparts. They, however, have some limitations, such as assuming prior knowledge of certain connections or the form of the nonlinear basis functions. Reproducing kernels have seen extensive use in graph topology inference problems. Such a solution was developed in [26], where an unsupervised kernel-based method is implemented. One particularity of the algorithm is that it requires, as a parameter, the number of sought edges. It also offers the possibility of statistical significance testing when setting this parameter. Another work is [27] where kernels,

chosen to best fit the available data, model nonlinear relationships between nodes based on measurements at successive time instants. The multi-kernel approach in [28] uses partial correlations to encode graph topology and ℓ_p-norm regression to enhance the performance. In [29], a thorough analysis of the kernel-based topology inference problem is given. This work focuses on capturing nonlinear and dynamic links. A review of state-of-the art methods in topology inference is in [30].

None of these works, however, bases its rationale on kernel derivatives. Moreover, no analysis of this family of methods exists in the literature. For these reasons, we put forth the current paper, which first recalls the algorithm and then proposes an in-depth convergence analysis.

B. Notations and definitions

Normal font letters denote scalars, while boldface lowercase and uppercase letters stand for column vectors and matrices, respectively. Uppercase calligraphic letters denote sets. Their cardinality is denoted by $|\cdot|$. The (u,v)-th entry of a matrix $\boldsymbol{X}$ is denoted by $[\boldsymbol{X}]_{u,v}$ or by x_{uv}. Notation $\boldsymbol{\alpha} = \text{col}\{\{\alpha_p\}_p\}$ denotes the column vector obtained by stacking all entries α_p for all available indexes p. Finally, $\mathbb{E}\{\cdot\}$ is the expectation operator.

A graph $\mathcal{G}$ consists of a set $\mathcal{N}$ of $(N+1)$ nodes, and a set $\mathcal{E}$ of edges such that if nodes m and n are linked, then we have $(m,n) \in \mathcal{E}$. For undirected graphs, these node pairs are unordered. At node level, we collect a real-valued signal and we organize it as a column vector $[y_1(i), \ldots, y_{N+1}(i)]^\top$, where $y_n(i)$ is the signal sample at node n and time instant i. The adjacency matrix $\boldsymbol{A}$ [31], [32], is defined as an $(N+1)\times(N+1)$ matrix whose entries a_{nm} are zero if $(m,n) \notin \mathcal{E}$ and set to one otherwise. Throughout the paper we shall consider that there is no self-loop, i.e., $a_{nn} = 0$, for all n.

C. Paper outline

The paper is organized as follows. Sections II and III recall the problem previously considered in [12] and the online algorithm, respectively. Section IV then proceeds with the main contribution of the current paper, consisting of the analysis of the algorithm, both in mean and mean square sense. In order to validate the pertinence of the analysis, Section V showcases a set of experimental simulations. Finally, a set of concluding remarks is given in Section VI.

II. Local problem formulation

Consider an $(N+1)$-node graph with adjacency matrix $\boldsymbol{A}$ that models a system such as a brain network or a power grid. In this context, the electrical activity of different brain regions [33], [34], or the voltage angle per bus [28], numbered from 1 to $(N+1)$, can be measured sequentially at different time instants $i \in \mathbb{N}$, leading to a graph signal given by $[y_1(i), \ldots, y_{N+1}(i)]^\top \in \mathbb{R}^{N+1}$ that changes over time. Signal $y_n(i)$ at each node $n = 1, \ldots, N+1$ of the graph is nonlinearly coupled to the signals at all nodes in its neighborhood, according to the topology described in $\boldsymbol{A}$. Moreover, we assume that nonlinear relationships between signals at different nodes occur as have been reported in many applications; see, e.g., the case of brain connectivity [35], [36].

Recent methods have considered additive nonlinear models [37], [38], which represent the local measurements at each node n as:

$$y_n(i) = \sum_{m \in \mathcal{N}\setminus\{n\}} f_{nm}(\boldsymbol{y}_m(i)) + \varrho_n(i)\,, \tag{1}$$

where $\boldsymbol{y}_m(i) = \big[y_m(i), \ldots, y_m(i-L_m+1)\big]^\top$, for $L_m \geq 1$. Parameter L_m endows the algorithm with memory-like capacities. This characteristic is desirable in a number of applications such as brain topology inference where there is a 10–20 ms delay in signal propagation between nodes [39]. Functions $f_{nm} : \mathbb{R}^{L_m} \to \mathbb{R}$ represent the interactions between the different nodes in the network, and $\varrho_n(i)$ denotes innovation noise. By relating how each node $m \in \mathcal{N}\setminus\{n\}$ affects node n, functions f_{nm} encode the connectivity in $\boldsymbol{A}$ directly as $a_{nm} = 0$ if, and only if, $f_{nm} \equiv 0$, where a_{nm} is the (n,m)-th entry of $\boldsymbol{A}$. For ease of notation, we shall assume that $n \equiv (N+1)$, i.e., we identify n as the $(N+1)$-th node of the graph, which allows us to denote $\mathcal{N}\setminus\{n\} = \{1, \ldots, N\}$. The topology inference problem then consists of finding the set of functions f_{nm} which best represent the available graph signal measurements $y_n(i)$ based on the previous model. This can be performed using both batch-based (see, e.g., [37]) and online strategies [38], [40], [41].

Models such as (1) do not consider general nonlinear interactions between multiple nodes, as they assume an additive model for $y_n(i)$ [42]. This can be rather limiting since the nonlinearity introduced in the model through the functions f_{nm} is local, i.e., it only acts on the signals at each node individually. To overcome this issue, we propose to consider the following general nonlinear model:

$$y_n(i) = f_n(\boldsymbol{y}(i)) + \varrho_n(i)\,, \tag{2}$$

where $\boldsymbol{y}(i) = [\boldsymbol{y}_1(i)^\top, \ldots, \boldsymbol{y}_N(i)^\top]^\top$ and function f_n describes the interaction between the signals at nodes $\mathcal{N}\setminus\{n\}$ and the signal $y_n(i)$ at node n. Compared to (1), model (2) captures more complex relationships between nodes without relying on generalized linear model, rendering it more general.

The method we propose relates f_n to the underlying graph topology $\boldsymbol{A}$ by quantifying how a node m affects node n through the corresponding partial derivative, i.e.:

$$\begin{array}{c}\text{node } m \text{ does not affect}\\ \text{node } n \text{ (i.e., } a_{mn} = 0)\end{array} \iff \left\| \frac{\partial f_n(\boldsymbol{y})}{\partial \boldsymbol{y}_m} \right\| = 0\,, \tag{3}$$

under the assumption that f_n is continuously differentiable, where the expectation is taken w.r.t. $\boldsymbol{y}$.

Estimating the graph topology can then be performed by learning function f_n that fits the data as per model (2), while additionally considering penalties to promote a sparsely connected graph, as in many real-world applications [43]. Afterwards, the topology can be recovered via relation (3). By constraining f_n to belong to a Reproducing Kernel Hilbert Space (RKHS) $\mathcal{H}$ associated with a positive definite repro-

ducing kernel $\kappa(\cdot,\cdot)$, this problem can be formulated in batch form as [12]:

$$\min_{f_n\in\mathcal{H}} \frac{1}{2i}\sum_{\ell=1}^{i}\left\|y_n(\ell)-f_n(\boldsymbol{y}(\ell))\right\|^2 + \eta\left(\sum_{m=1}^{N}\sqrt{\frac{1}{i}\sum_{p=1}^{i}\sum_{q=1}^{L_m}\left(\frac{\partial f_n(\boldsymbol{y}(p))}{\partial y_{m,q}}\right)^2}+\psi\big(\|f_n\|_{\mathcal{H}}\big)\right), \tag{4}$$

where $y_{m,q}$ represents the q-th entry of $\boldsymbol{y}_m$. The first term in (4) is a data fitting term that averages the reconstruction error over all available samples. The second one is a convex group-sparsity inducing penalty [44], [45], which promotes sparsity in the estimated derivatives averaged over all observations $\boldsymbol{y}(p)$ up to time instant i, and consequently on adjacency matrix $\boldsymbol{A}$. The last term, with $\psi:\mathbb{R}\to[0,\infty)$, is a monotonically increasing function of $\|f_n\|_{\mathcal{H}}$ with small magnitude in order to guarantee the uniqueness of the solution of (4). Parameter η makes a trade off between data the fitting term and the regularizers.

Problem (4) allows us to introduce sparsity in f_n without the inherent limitations of an additive model. However, the sparsity-promoting penalty term prevents the application of previously established Representation Theorems that guarantee the existence of a finite-dimensional representation of the solution. Assuming that kernel $\kappa(\cdot,\cdot)$ is twice differentiable, the following holds [46]:

$$\mathcal{H}\ni\frac{\partial f_n(\boldsymbol{y})}{\partial y_{m,q}}=\langle f_n,\kappa_{\partial_{m,q}}(\cdot,\boldsymbol{y})\rangle_{\mathcal{H}}\,, \tag{5}$$

where:

$$\kappa_{\partial_{m,q}}(\cdot,\boldsymbol{y}(q))=\left.\frac{\partial\kappa(\cdot,\boldsymbol{a})}{\partial a_{m,q}}\right|_{\boldsymbol{a}=\boldsymbol{y}(q)}. \tag{6}$$

Thus, for sufficiently smooth kernels, the derivative of functions in $\mathcal{H}$ belongs to the same $\mathcal{H}$, in turn meaning that they can be evaluated in the form of simple inner products. The solution of (4) can now be written in a finite dimensional form, similarly to the proposition in [45]. By generalizing the Representer Theorem in the aforementioned paper, we obtain a Representer Theorem for our approach. This result is stated in the following theorem:

Theorem 1. *Let $\mathcal{H}$ be a RKHS whose associated reproducing kernel $\kappa(\cdot,\cdot)$ is at least twice differentiable. Then, the solution of the optimization problem* (4) *can be written as:*

$$f_n^*=\sum_{p=1}^{i}\alpha_p\kappa(\cdot,\boldsymbol{y}(p)) + \sum_{m=1}^{N}\sum_{q=1}^{i}\sum_{\ell=1}^{L_m}\beta_{m,q,\ell}\kappa_{\partial_{m,\ell}}(\cdot,\boldsymbol{y}(q))\,. \tag{7}$$

Proof. The proof of this theorem is presented in Appendix A. □

Expression (7) can then be substituted in problem (4) in order to obtain a finite-dimensional optimization problem.

In the sequel, for the sake of simplicity in the notation and ease of comprehension, we shall consider the case $L_m=1$ for all m, i.e., $\boldsymbol{y}_m(i)=y_m(i)$ and $\kappa_{\partial_{m,q}}(\cdot,\boldsymbol{y})=\kappa_{\partial_m}(\cdot,\boldsymbol{y})$.

III. An online algorithm

The number of coefficients α_p and $\beta_{m,q}$ in solution (7) can become prohibitive as i increases with each new measurement and associated kernel function. Kernel dictionaries are a solution to this issue. These dictionaries can be defined either *a priori* [47] or can admit a new candidate kernel function only if it satisfies a certain sparsification rule [48], such as the approximate linear dependence criterion [49], [50], the Nyström method [51], dictionary-based random Fourier features [52], orthogonal projection [53] or the novelty criterion [54].

We consider a dictionary-based framework, where each node n in the network is equipped with a dictionary of kernel functions (and their derivatives) of the following form:

$$\mathcal{D}_n=\left\{\kappa(\cdot,\boldsymbol{y}(\omega_j)):\omega_j\in\mathcal{I}_n^i\right\}, \tag{8}$$

where $\mathcal{I}_n^i\subset\{1,\dots,i-1\}$ represents the set of time indices of elements selected for the dictionary, before instant i. To build $\mathcal{D}_n$ iteratively, we consider the coherence criterion as the chosen method for dictionary sparsification [55]. This entails the fact that, after a sufficient number of samples i, only a small number $|\mathcal{I}_n^i|\ll i$ of coefficients will be needed. A candidate kernel function $\kappa(\cdot,\boldsymbol{y}(i))$ is added to $\mathcal{D}_n$ if the following sparsification condition holds [55]:

$$\max_{\omega_j\in\mathcal{I}_n^i}|\kappa(\boldsymbol{y}(i),\boldsymbol{y}(\omega_j))|\le\xi_n\,, \tag{9}$$

where $\xi_n\in[0,1)$ determines the level of sparsity and the coherence of the dictionary. The number of entries in the dictionary satisfies $|\mathcal{I}_n^i|<\infty$ when $i\to\infty$ [55]. Note that the dictionary $\mathcal{D}_n$ can also be set *a priori* using, e.g., an uniform grid sampling [47]. In this case, with a slight abuse of notation, we represent the dictionary using a set of time indices $\mathcal{I}_n^i\subset\mathbb{Z}\setminus\mathbb{N}$, attributing to the dictionary elements negative time indexes. We rewrite (7) as:

$$f_{n,\mathcal{D}_n}=\sum_{p=1}^{|\mathcal{D}_n|}\alpha_p\kappa(\cdot,\boldsymbol{y}(\omega_p)) + \sum_{m=1}^{N}\sum_{q=1}^{|\mathcal{D}_n|}\beta_{m,q}\kappa_{\partial_m}(\cdot,\boldsymbol{y}(\omega_q))\,. \tag{10}$$

For notation compactness, we shall write $\boldsymbol{\alpha}=\mathrm{col}\{\{\alpha_p\}_p\}$, and $\boldsymbol{\beta}=\mathrm{col}\{\{\boldsymbol{\beta}_m\}_m\}$ with $\boldsymbol{\beta}_m=\mathrm{col}\{\{\beta_{m,q}\}_q\}$.

The dictionary-based representation $f_{n,\mathcal{D}_n}$ in (10) can now be plugged directly into the batch problem (4) to obtain a solution depending on a small amount of coefficients, for all i. However, this would still entail a growth in complexity with time i, since the kernel functions have to be evaluated at all available data points. To address this issue and devise a computationally scalable algorithm, we consider an alternative cost function to (4). Specifically, we substitute the averaging operations in the data fitting term and in the derivative penalty term in (4), which use all data points up to time instant i, by

the corresponding expected values. Setting $\psi \equiv 0$ without loss of generality, this leads to:

$$\min_{f_{n,\mathcal{D}_n}} \frac{1}{2}\,\mathbb{E}\left\{\left\|y_n(i) - f_{n,\mathcal{D}_n}(\boldsymbol{y}(i))\right\|^2\right\} + \eta \sum_{m=1}^{N} \sqrt{\mathbb{E}\left\{\left(\frac{\partial f_{n,\mathcal{D}_n}(\boldsymbol{y}(i))}{\partial y_m(i)}\right)^2\right\}}. \tag{11}$$

where the expectations are w.r.t. $\boldsymbol{y}(i)$ and conditioned on the dictionary (8).

Combining (11) with (10) as shown in Appendix B, we get:

$$\min_{\boldsymbol{\gamma}} J(\boldsymbol{\gamma}) = \frac{1}{2}\,\mathbb{E}\Big\{\left\|y_n(i) - \boldsymbol{\gamma}^\top \boldsymbol{s}(i)\right\|^2\Big\} + \eta \sum_{m=1}^{N} \sqrt{\boldsymbol{\gamma}^\top \boldsymbol{R}_{tt,m}(i)\,\boldsymbol{\gamma}}\,, \tag{12}$$

with $\boldsymbol{s}(i) = \begin{bmatrix}\boldsymbol{z}(i)\\ \boldsymbol{k}(i)\end{bmatrix}$, $\boldsymbol{\gamma} = \begin{bmatrix}\boldsymbol{\beta}\\ \boldsymbol{\alpha}\end{bmatrix}$, and $\boldsymbol{t}_m(p) = \begin{bmatrix}\boldsymbol{\ell}_m(p)\\ \boldsymbol{\zeta}_m(p)\end{bmatrix}$, where:

$$\boldsymbol{R}_{tt,m}(i) \triangleq \mathbb{E}\{\boldsymbol{t}_m(i)\boldsymbol{t}_m^\top(i)\} \tag{13}$$

$$\boldsymbol{k}(i) = \operatorname{col}\Big\{\kappa(\boldsymbol{y}(i), \boldsymbol{y}(\omega_q))\Big\}_{q=1}^{|\mathcal{D}_n|} \tag{14}$$

$$\boldsymbol{z}(i) = \left[\boldsymbol{z}_1^\top(i), \ldots, \boldsymbol{z}_N^\top(i)\right]^\top \tag{15}$$

$$[\boldsymbol{z}_m(i)]_q = \left.\frac{\partial \kappa(\boldsymbol{y}(i), \boldsymbol{y}(\omega_q))}{\partial y_m(\omega_q)}\right|_{q=1,\ldots,|\mathcal{D}_n|} \tag{16}$$

$$\boldsymbol{\zeta}(i) = \left[\boldsymbol{\zeta}_1^\top(i), \ldots, \boldsymbol{\zeta}_N^\top(i)\right]^\top \tag{17}$$

$$[\boldsymbol{\zeta}_m(i)]_q = \left.\frac{\partial \kappa(\boldsymbol{y}(i), \boldsymbol{y}(\omega_q))}{\partial y_m(i)}\right|_{q=1,\ldots,|\mathcal{D}_n|} \tag{18}$$

$$\boldsymbol{\ell}_m(i) = \left[\boldsymbol{\ell}_{1,m}^\top(i), \ldots, \boldsymbol{\ell}_{N,m}^\top(i)\right]^\top \tag{19}$$

$$[\boldsymbol{\ell}_{m_1,m_2}(i)]_q = \left.\frac{\partial^2 \kappa(\boldsymbol{y}(i), \boldsymbol{y}(\omega_q))}{\partial y_{m_1}(\omega_q)\partial y_{m_2}(i)}\right|_{q=1,\ldots,|\mathcal{D}_n|} \tag{20}$$

Note that $\boldsymbol{\gamma}$ is now the optimization variable in problem (12). Quantities (14)–(20) can be computed in closed form when an explicit expression for kernel $\kappa(\cdot,\cdot)$ is provided. Expressions in the case of the Gaussian kernel are given in Appendix C. A closed form expression of $\boldsymbol{R}_{tt,m}$ for Gaussian random variables is given in the supplementary material.

Calculating the gradient of (12) leads to:

$$\nabla J(\boldsymbol{\gamma}) = -\mathbb{E}\Big\{\boldsymbol{s}(i)\big[y_n(i) - \boldsymbol{s}^\top(i)\boldsymbol{\gamma}\big]\Big\} + \eta \sum_{m=1}^{N} \frac{\boldsymbol{R}_{tt,m}(i)\boldsymbol{\gamma}}{\sqrt{\boldsymbol{\gamma}^\top \boldsymbol{R}_{tt,m}(i)\,\boldsymbol{\gamma}}}\,. \tag{21}$$

Approximating the first expectation in (21) by the instantaneous estimate $\boldsymbol{s}(i)\big[y_n(i) - \boldsymbol{s}^\top(i)\boldsymbol{\gamma}(i)\big]$, and $\boldsymbol{R}_{tt,m}(i)$ by a sample covariance matrix $\hat{\boldsymbol{R}}_{tt,m}(i)$ available at time i, allows us to devise the following subgradient descent algorithm to iteratively minimize $J(\boldsymbol{\gamma})$:

$$\hat{\boldsymbol{\gamma}}(i+1) = \hat{\boldsymbol{\gamma}}(i) + \mu \boldsymbol{s}(i)\big[y_n(i) - \boldsymbol{s}^\top(i)\hat{\boldsymbol{\gamma}}(i)\big] - \mu\eta \sum_{m=1}^{N} \frac{\hat{\boldsymbol{R}}_{tt,m}(i)\hat{\boldsymbol{\gamma}}(i)}{\sqrt{\hat{\boldsymbol{\gamma}}^\top(i)\,\hat{\boldsymbol{R}}_{tt,m}(i)\,\hat{\boldsymbol{\gamma}}(i)}}\,, \tag{22}$$

with μ a small positive step size. We set $x/0 \triangleq 0$ to write the subgradient when the denominator in the last term of (22) equals 0. For the sake of brevity, we introduce the notation:

$$\Delta_m(i) = \sqrt{\hat{\boldsymbol{\gamma}}^\top(i)\,\hat{\boldsymbol{R}}_{tt,m}(i)\,\hat{\boldsymbol{\gamma}}(i)}\,. \tag{23}$$

The covariance matrix $\hat{\boldsymbol{R}}_{tt,m}(i)$ needs to be updated sequentially in environments where the data arrives sequentially. An extensive body of literature concerning sequential estimators exists. The following sample covariance matrix provides such an estimate:

$$\hat{\boldsymbol{R}}_{tt,m}(i) = \alpha\,\hat{\boldsymbol{R}}_{tt,m}(i-1) + (1-\alpha)\,\boldsymbol{t}_m(i)\boldsymbol{t}_m^\top(i)\,, \tag{24}$$

with $\alpha \in [0,1)$ a forgetting factor. In the absence of hypothesis about the covariance matrix, a successful approach so far has been shrinkage estimation [56]. We shall not discuss these details here, which are out of the scope of this paper, but refer the reader to the literature. We shall however assume that the estimator is unbiased, that is:

$$\mathbb{E}\{\hat{\boldsymbol{R}}_{tt,m}(i)\} = \boldsymbol{R}_{tt,m}(i)\,. \tag{25}$$

The time index i will be omitted for stationary variables.

In (22)–(23), observe that each $\Delta_m(i)$ represents an estimate of the squared norm of the partial derivative of f_n with respect to $\boldsymbol{y}_m$. It thus allows us to infer the graph topology by comparing every quantity $\Delta_m(i)$ to a given threshold τ_n such that, following (3), $\hat{a}_{nm}(i) = 1$ if $\Delta_m(i) \geq \tau_n$ and 0 otherwise. These thresholds can be chosen to either obtain an estimated topology which realistically explains the studied process [34], or to obtain a connected graph, i.e., a graph in which there exists a path between all pairs of nodes.

IV. Theoretical analysis

Let $\boldsymbol{v}(i) = \hat{\boldsymbol{\gamma}}(i) - \boldsymbol{\gamma}^*$ be the weight-error vector, where $\hat{\boldsymbol{\gamma}}(i)$ is the current estimate in (22) and $\boldsymbol{\gamma}^*$ is the optimal coefficient vector, obtained as the solution to problem (12).

A. Relating the error in derivatives to the error in coefficients

One difficulty related to the iterative algorithm (22) is that it considers the evolution of the coefficients $\boldsymbol{\gamma}$, instead of the partial derivatives $\partial \hat{f}_{n,\mathcal{D}_n}(\boldsymbol{y})/\partial y_m$. Nevertheless, we can study the convergence of the derivatives by means of filter coefficients as will be shown in the following.

Using (5) and the Cauchy-Schwarz inequality:

$$\left|\frac{\partial \hat{f}_{n,\mathcal{D}_n}(\boldsymbol{y})}{\partial y_m} - \frac{\partial f^*_{n,\mathcal{D}_n}(\boldsymbol{y})}{\partial y_m}\right| = \tag{26}$$

$$= \left|\langle \hat{f}_{n,\mathcal{D}_n}, \kappa_{\partial_m}(\cdot,\boldsymbol{y})\rangle_{\mathcal{H}} - \langle f^*_{n,\mathcal{D}_n}, \kappa_{\partial_m}(\cdot,\boldsymbol{y})\rangle_{\mathcal{H}}\right| = \left|\langle \hat{f}_{n,\mathcal{D}_n} - f^*_{n,\mathcal{D}_n}, \kappa_{\partial_m}(\cdot,\boldsymbol{y})\rangle_{\mathcal{H}}\right| \leq \|\kappa_{\partial_m}(\cdot,\boldsymbol{y})\|_{\mathcal{H}} \|\hat{f}_{n,\mathcal{D}_n} - f^*_{n,\mathcal{D}_n}\|_{\mathcal{H}}\,. \tag{27}$$

Moreover, we also have:

$$\|\hat{f}_{n,\mathcal{D}_n} - f^*_{n,\mathcal{D}_n}\|_{\mathcal{H}} = \|\boldsymbol{s}^\top\hat{\boldsymbol{\gamma}} - \boldsymbol{s}^\top\boldsymbol{\gamma}^*\| = \|\boldsymbol{s}^\top(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*)\| \leq \|\boldsymbol{s}\|\,\|\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*\|\,. \tag{28}$$

By taking the expectation of both sides conditioned on $\hat{\boldsymbol{\gamma}}$, we have:

$$\mathbb{E}\left\{\left|\frac{\partial \hat{f}_{n,\mathcal{D}_n}(\boldsymbol{y})}{\partial y_m}-\frac{\partial f^*_{n,\mathcal{D}_n}(\boldsymbol{y})}{\partial y_m}\right|\,\middle|\,\hat{\boldsymbol{\gamma}}\right\} \leq \mathbb{E}\left\{\|\kappa_{\partial_m}(\cdot,\boldsymbol{y})\|_{\mathcal{H}}\|\boldsymbol{s}\|\,\middle|\,\hat{\boldsymbol{\gamma}}\right\}\|\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^*\|\,. \tag{29}$$

where the first expectation on the r.h.s. is not related to $\hat{f}_{n,\mathcal{D}_n}$ and $f^*_{n,\mathcal{D}_n}$ but depends on the statistics of the input data and on the reproducing kernel and its derivatives. Thus the error in the estimated derivatives is bounded by the error in the coefficients $\hat{\boldsymbol{\gamma}}$. This allows us to study the convergence behavior of the derivatives indirectly by means of the coefficients $\hat{\boldsymbol{\gamma}}$.

B. Simplifying assumptions

To make the analysis of algorithm (22) tractable, we need to introduce some simplifying assumptions.

A1) We assume that vector $\boldsymbol{y}(i)$ is stationary and Gaussian-distributed, with zero-mean and covariance matrix $\boldsymbol{R}_y$, namely:

$$p\big(\boldsymbol{y}(i)\big)=(2\pi)^{-N/2}\det\{\boldsymbol{R}_y\}^{-1/2}\exp\left(-\frac{1}{2}\boldsymbol{y}^\top(i)\boldsymbol{R}_y^{-1}\boldsymbol{y}(i)\right).$$

A2) We assume that $L_m = 1$ for simplicity, but the same analysis can be extended for $L_m > 1$.

A3) We consider the Gaussian kernel:

$$\kappa(\boldsymbol{a},\boldsymbol{b})=\exp\left(-\|\boldsymbol{a}-\boldsymbol{b}\|^2/2\sigma^2\right), \tag{30}$$

due to its capacities as an universal approximator [57].

A4) We make the Modified Independence Assumption, that is, $\boldsymbol{s}(i)\boldsymbol{s}^\top(i)$ is statistically independent of $\boldsymbol{v}(i)$. This assumption has been successfully used in the analysis of various adaptive filtering algorithms [58]. It has also been shown in [59] to be less restrictive when compared to the classical independence assumption.

A5) The elements $\{\boldsymbol{y}(\omega_p)\}$ in the dictionary are set *a priori*. They are thus independent of $\boldsymbol{y}(i)$.

A6) The entries of $\boldsymbol{v}(i)$ are jointly Gaussian distributed. This hypothesis has been used and validated in the analysis of both linear [60], [61] and nonlinear (kernel-based) [62] adaptive algorithms.

We also use the following approximation

$$\hat{\boldsymbol{R}}_{tt,m}(i)\approx\mathbb{E}\big\{\hat{\boldsymbol{R}}_{tt,m}(i)\big\}\overset{(25)}{=}\boldsymbol{R}_{tt,m} \tag{31}$$

to make the analysis tractable. As an example, note that the decay of the estimation error of $\boldsymbol{R}_{tt,m}$ using (24) with i.i.d. samples is in the order of $(1/i)$.

C. Mean weight error analysis

The estimation error $e(i)$ is given by:

$$\begin{aligned}e(i)&\triangleq y_n(i)-f_n(\boldsymbol{y}(i))=y_n(i)-\boldsymbol{s}^\top(i)\hat{\boldsymbol{\gamma}}(i)\\&=\underbrace{y_n(i)-\boldsymbol{s}^\top(i)\boldsymbol{\gamma}^*}_{e_0(i)}-\boldsymbol{s}^\top(i)\boldsymbol{v}(i)\,.\end{aligned} \tag{32}$$

Considering (22) and (32), we can write $\boldsymbol{v}(i+1)$ as:

$$\begin{aligned}\boldsymbol{v}(i+1)=\boldsymbol{v}(i)&-\mu\boldsymbol{s}(i)\boldsymbol{s}^\top(i)\boldsymbol{v}(i)+\mu\boldsymbol{s}(i)e_0(i)\\&-\mu\eta\sum_{m=1}^{N}\frac{\hat{\boldsymbol{R}}_{tt,m}(i)\left[\boldsymbol{v}(i)+\boldsymbol{\gamma}^*\right]}{\sqrt{\left[\boldsymbol{v}(i)+\boldsymbol{\gamma}^*\right]^\top\hat{\boldsymbol{R}}_{tt,m}(i)\left[\boldsymbol{v}(i)+\boldsymbol{\gamma}^*\right]}}\,,\end{aligned} \tag{33}$$

where $e_0(i)=y_n(i)-\boldsymbol{s}^\top(i)\boldsymbol{\gamma}^*$ is the optimal estimation error. Taking the expected value of both sides of (33), and using A4, we obtain the mean weight error model:

$$\begin{aligned}\mathbb{E}\{\boldsymbol{v}(i+1)\}=(\boldsymbol{I}-\mu\boldsymbol{R}_{ss})\mathbb{E}\{\boldsymbol{v}(i)\}+\mu\big(\boldsymbol{r}_{sy}-\boldsymbol{R}_{ss}\boldsymbol{\gamma}^*\big)\\-\mu\eta\sum_{m=1}^{N}\mathbb{E}\left\{\frac{\hat{\boldsymbol{R}}_{tt,m}(i)\left[\boldsymbol{v}(i)+\boldsymbol{\gamma}^*\right]}{\sqrt{\left[\boldsymbol{v}(i)+\boldsymbol{\gamma}^*\right]^\top\hat{\boldsymbol{R}}_{tt,m}(i)\left[\boldsymbol{v}(i)+\boldsymbol{\gamma}^*\right]}}\right\},\end{aligned} \tag{34}$$

with $\boldsymbol{R}_{ss}\triangleq\mathbb{E}\{\boldsymbol{s}(i)\boldsymbol{s}^\top(i)\}$ and $\boldsymbol{r}_{sy}\triangleq\mathbb{E}\{\boldsymbol{s}(i)y_n(i)\}$.

The last term in (34) is a non-trivial expectation involving $\boldsymbol{v}(i)$ and $\hat{\boldsymbol{R}}_{tt,m}(i)$. To proceed, we need to use the simplifying assumption $\hat{\boldsymbol{R}}_{tt,m}(i)\approx\boldsymbol{R}_{tt,m}$ discussed above. This yields:

$$\mathbb{E}\left\{\frac{\hat{\boldsymbol{R}}_{tt,m}(i)\hat{\boldsymbol{\gamma}}(i)}{\sqrt{\hat{\boldsymbol{\gamma}}^\top(i)\hat{\boldsymbol{R}}_{tt,m}(i)\hat{\boldsymbol{\gamma}}(i)}}\right\}\approx\mathbb{E}\left\{\frac{\boldsymbol{R}_{tt,m}\hat{\boldsymbol{\gamma}}(i)}{\sqrt{\hat{\boldsymbol{\gamma}}^\top(i)\boldsymbol{R}_{tt,m}\hat{\boldsymbol{\gamma}}(i)}}\right\}. \tag{35}$$

By successively approximating the expectation of the ratio by the ratio of the expectations, and the expectation of the square root by the square root of the expectation [63, p. 70], we have:

$$\begin{aligned}&\mathbb{E}\left\{\frac{\boldsymbol{R}_{tt,m}\hat{\boldsymbol{\gamma}}(i)}{\sqrt{\hat{\boldsymbol{\gamma}}^\top(i)\boldsymbol{R}_{tt,m}\hat{\boldsymbol{\gamma}}(i)}}\right\}\\&\quad\approx\frac{\boldsymbol{R}_{tt,m}\boldsymbol{\mu}(i)}{\sqrt{\mathrm{Tr}\{\boldsymbol{R}_{tt,m}\boldsymbol{\Sigma}(i)\}+\boldsymbol{\mu}^\top(i)\boldsymbol{R}_{tt,m}\boldsymbol{\mu}(i)}}\,,\end{aligned} \tag{36}$$

which represents the sought-after approximation of the regularization term, with $\boldsymbol{\mu}(i)$ and $\boldsymbol{\Sigma}(i)$ given by:

$$\boldsymbol{\mu}(i)=\mathbb{E}\big\{\hat{\boldsymbol{\gamma}}(i)\big\}=\mathbb{E}\{\boldsymbol{v}(i)\}+\boldsymbol{\gamma}^*\,, \tag{37}$$

and:

$$\begin{aligned}\boldsymbol{\Sigma}(i)&=\mathbb{E}\big\{\hat{\boldsymbol{\gamma}}(i)\hat{\boldsymbol{\gamma}}^\top(i)\big\}-\boldsymbol{\mu}(i)\,\boldsymbol{\mu}^\top(i)\\&=\mathbb{E}\{\boldsymbol{v}(i)\boldsymbol{v}^\top(i)\}-\mathbb{E}\{\boldsymbol{v}(i)\}\mathbb{E}\{\boldsymbol{v}(i)\}^\top\,,\end{aligned} \tag{38}$$

respectively. Using these results in (34), we obtain the following expression for the mean error recursion:

$$\begin{aligned}\mathbb{E}\{\boldsymbol{v}(i+1)\}\approx(\boldsymbol{I}-\mu\boldsymbol{R}_{ss})\mathbb{E}\{\boldsymbol{v}(i)\}+\mu\big(\boldsymbol{r}_{sy}-\boldsymbol{R}_{ss}\boldsymbol{\gamma}^*\big)\\-\mu\eta\sum_{m=1}^{N}\frac{\boldsymbol{R}_{tt,m}\,\boldsymbol{\mu}(i)}{\sqrt{\mathrm{Tr}\{\boldsymbol{R}_{tt,m}\boldsymbol{\Sigma}(i)\}+\boldsymbol{\mu}^\top(i)\,\boldsymbol{R}_{tt,m}\,\boldsymbol{\mu}(i)}}\,.\end{aligned} \tag{39}$$

where $\boldsymbol{\mu}(i)$ and $\boldsymbol{\Sigma}(i)$ have to be calculated at each time i in order to make recursion (39) functional. Expressions (37) and (38) are used, respectively. Specifically, a recursive model is provided in subsection IV-E for calculating $\mathbb{E}\{\boldsymbol{v}(i)\boldsymbol{v}^\top(i)\}$.

D. Mean stability analysis

Iterating (34) starting from $i=0$, we arrive to the following expression:

$$\begin{aligned}
&\mathbb{E}\{\boldsymbol{v}(i+1)\} \\
&= (\boldsymbol{I}-\mu\boldsymbol{R}_{ss})^{i+1}\mathbb{E}\{\boldsymbol{v}(0)\} - \mu\eta\sum_{\ell=0}^{i}(\boldsymbol{I}-\mu\boldsymbol{R}_{ss})^{i-\ell} \\
&\quad\times\sum_{m=1}^{N}\mathbb{E}\left\{\frac{\hat{\boldsymbol{R}}_{tt,m}(\ell)\left[\boldsymbol{v}(\ell)+\boldsymbol{\gamma}^*\right]}{\sqrt{\left[\boldsymbol{v}(\ell)+\boldsymbol{\gamma}^*\right]^{\top}\hat{\boldsymbol{R}}_{tt,m}(\ell)\left[\boldsymbol{v}(\ell)+\boldsymbol{\gamma}^*\right]}}\right\} \\
&\quad+\mu\sum_{\ell=0}^{i}(\boldsymbol{I}-\mu\boldsymbol{R}_{ss})^{i-\ell}(\boldsymbol{r}_{sy}-\boldsymbol{R}_{ss}\boldsymbol{\gamma}^*)\,, \qquad (40)
\end{aligned}$$

where $\mathbb{E}\{\boldsymbol{v}(0)\}$ is the initial condition. $\mathbb{E}\{\boldsymbol{v}(i+1)\}$ converges when $i\to\infty$ if, and only if, all terms on the r.h.s. of (40) converge to finite values. The first term converges to zero as $i\to\infty$ if the matrix $(\boldsymbol{I}-\mu\boldsymbol{R}_{ss})$ is stable. A sufficient condition to ensure the stability of $(\boldsymbol{I}-\mu\boldsymbol{R}_{ss})$ is to choose the step-size μ according to:

$$0<\mu<\frac{2}{\lambda_{\max}(\boldsymbol{R}_{ss})}\,. \qquad (41)$$

where $\lambda_{\max}(\cdot)$ denotes the maximum eigenvalue of its matrix argument. We shall now prove the convergence of the second and third series on the r.h.s. of (40). For compactness, we write the second series as $\sum_{\ell=0}^{i}(\boldsymbol{I}-\mu\boldsymbol{R}_{ss})^{i-\ell}\sum_{m=1}^{N}\boldsymbol{q}_m(\ell)$. To prove the convergence of this series, it is sufficient to prove that the series $\sum_{\ell=0}^{i}\sum_{m=1}^{N}[(\boldsymbol{I}-\mu\boldsymbol{R}_{ss})^{i-\ell}\boldsymbol{q}_m(\ell)]_k$ converges for all k, where $[\cdot]_k$ is the k-th entry of its vector argument. A series is absolutely convergent if each term of the series can be bounded by a term of an absolutely convergent series. Using Jensen's inequality, we have:

$$\begin{aligned}
&\left|[(\boldsymbol{I}-\mu\boldsymbol{R}_{ss})^{i-\ell}\boldsymbol{q}_m(\ell)]_k\right| \\
&\le\left\|(\boldsymbol{I}-\mu\boldsymbol{R}_{ss})^{i-\ell}\boldsymbol{q}_m(\ell)\right\| \\
&\approx\left\|(\boldsymbol{I}-\mu\boldsymbol{R}_{ss})^{i-\ell}\,\mathbb{E}\left\{\frac{\boldsymbol{R}_{tt,m}\left[\boldsymbol{v}(\ell)+\boldsymbol{\gamma}^*\right]}{\sqrt{\left[\boldsymbol{v}(\ell)+\boldsymbol{\gamma}^*\right]^{\top}\boldsymbol{R}_{tt,m}\left[\boldsymbol{v}(\ell)+\boldsymbol{\gamma}^*\right]}}\right\}\right\| \\
&\le\|\boldsymbol{I}-\mu\boldsymbol{R}_{ss}\|^{i-\ell}\;\mathbb{E}\left\{\frac{\|\boldsymbol{R}_{tt,m}\|\,\|\boldsymbol{v}(\ell)+\boldsymbol{\gamma}^*\|}{\|\boldsymbol{v}(\ell)+\boldsymbol{\gamma}^*\|_{\boldsymbol{R}_{tt,m}}}\right\} \\
&\le\|\boldsymbol{I}-\mu\boldsymbol{R}_{ss}\|^{i-\ell}\;\tau_m\,\|\boldsymbol{R}_{tt,m}\|\,, \qquad (42)
\end{aligned}$$

where $\|\boldsymbol{v}\|_{\boldsymbol{R}_{tt,m}}$ denotes $\boldsymbol{v}^{\top}\boldsymbol{R}_{tt,m}\boldsymbol{v}$. The last inequality follows because all norms are equivalent in finite dimensional spaces, that is, there exists a $\tau_m>0$ such that:

$$\|\boldsymbol{v}\|\le\tau_m\,\|\boldsymbol{v}\|_{\boldsymbol{R}_{tt,m}}\,, \qquad (43)$$

for all $\boldsymbol{v}$. The series $\sum_{\ell=0}^{i}\|\boldsymbol{I}-\mu\boldsymbol{R}_{ss}\|^{i-\ell}\sum_{m=1}^{N}\tau_m\,\|\boldsymbol{R}_{tt,m}\|$ is absolutely convergent if the step-size μ is chosen according to (41). The series $\sum_{\ell=0}^{i}\sum_{m=1}^{N}[(\boldsymbol{I}-\mu\boldsymbol{R}_{ss})^{i-\ell}\boldsymbol{q}_m(\ell)]_k$ is therefore absolutely convergent.

To show that the third series in (40) converges, it is sufficient to observe that:

$$\begin{aligned}
&\left|[(\boldsymbol{I}-\mu\boldsymbol{R}_{ss})^{i-\ell}\big(\boldsymbol{r}_{sy}-\boldsymbol{R}_{ss}\boldsymbol{\gamma}^*\big)]_k\right| \\
&\qquad\le\|\boldsymbol{I}-\mu\boldsymbol{R}_{ss}\|^{i-\ell}\,\|\boldsymbol{r}_{sy}-\boldsymbol{R}_{ss}\boldsymbol{\gamma}^*\|\,. \qquad (44)
\end{aligned}$$

The series $\sum_{\ell=0}^{i}\|\boldsymbol{I}-\mu\boldsymbol{R}_{ss}\|^{i-\ell}\,\|\boldsymbol{r}_{sy}-\boldsymbol{R}_{ss}\boldsymbol{\gamma}^*\|$ is absolutely convergent if the step-size μ is chosen according to (41).

To conclude, all series in (40) converge if (41) is satisfied. Therefore, for any initial condition, the algorithm (22) converges in the mean if the step-size μ is chosen according to condition (41).

E. Mean square error analysis

We examine now the mean-square-error behavior of the algorithm (22) by studying $\mathbb{E}\{\boldsymbol{v}(i)\boldsymbol{v}^{\top}(i)\}$. The mean-square-error recursion is given by:

$$\begin{aligned}
&\boldsymbol{V}(i+1)=\mathbb{E}\{\boldsymbol{v}(i+1)\boldsymbol{v}^{\top}(i+1)\} \\
&=\mathbb{E}\Bigg\{\operatorname{sym}\left\{\frac{1}{2}\underbrace{\boldsymbol{v}(i)\boldsymbol{v}^{\top}(i)}_{\boldsymbol{Q}_1}-\mu\underbrace{\boldsymbol{v}(i)\boldsymbol{v}^{\top}(i)\boldsymbol{s}(i)\boldsymbol{s}^{\top}(i)}_{\boldsymbol{Q}_2}\right. \\
&+\mu\underbrace{\boldsymbol{v}(i)\boldsymbol{s}^{\top}(i)e_0(i)}_{\boldsymbol{Q}_3}-\mu^2\underbrace{\boldsymbol{s}(i)\boldsymbol{s}^{\top}(i)\boldsymbol{v}(i)\boldsymbol{s}^{\top}(i)e_0(i)}_{\boldsymbol{Q}_4} \\
&+\frac{\mu^2}{2}\underbrace{\boldsymbol{s}(i)\boldsymbol{s}^{\top}(i)(e_0(i))^2}_{\boldsymbol{Q}_5}+\frac{\mu^2}{2}\underbrace{\boldsymbol{s}(i)\boldsymbol{s}^{\top}(i)\boldsymbol{v}(i)\boldsymbol{v}^{\top}(i)\boldsymbol{s}(i)\boldsymbol{s}^{\top}(i)}_{\boldsymbol{Q}_6} \\
&-\mu\eta\underbrace{\boldsymbol{v}(i)\left[\sum_{m=1}^{N}\frac{\hat{\boldsymbol{R}}_{tt,m}(i)\hat{\boldsymbol{\gamma}}(i)}{\Delta_m(i)}\right]^{\top}}_{\boldsymbol{Q}_7} \\
&+\mu^2\eta\underbrace{\boldsymbol{s}(i)\boldsymbol{s}^{\top}(i)\boldsymbol{v}(i)\left[\sum_{m=1}^{N}\frac{\hat{\boldsymbol{R}}_{tt,m}(i)\hat{\boldsymbol{\gamma}}(i)}{\Delta_m(i)}\right]^{\top}}_{\boldsymbol{Q}_8} \\
&-\mu^2\eta\underbrace{\boldsymbol{s}(i)e_0(i)\left[\sum_{m=1}^{N}\frac{\hat{\boldsymbol{R}}_{tt,m}(i)\hat{\boldsymbol{\gamma}}(i)}{\Delta_m(i)}\right]^{\top}}_{\boldsymbol{Q}_9} \\
&\left.\left.+\frac{\mu^2\eta^2}{2}\underbrace{\left[\sum_{m=1}^{N}\frac{\hat{\boldsymbol{R}}_{tt,m}(i)\hat{\boldsymbol{\gamma}}(i)}{\Delta_m(i)}\right]\left[\sum_{p=1}^{N}\frac{\hat{\boldsymbol{R}}_{tt,p}(i)\hat{\boldsymbol{\gamma}}(i)}{\Delta_p(i)}\right]^{\top}}_{\boldsymbol{Q}_{10}}\right\}\right\}, \qquad (45)
\end{aligned}$$

where $\operatorname{sym}(\boldsymbol{X})=\boldsymbol{X}+\boldsymbol{X}^{\top}$. We shall now calculate the expectations separately. Helpful details are provided in [58]:

$\boldsymbol{Q}_1$ $\mathbb{E}\{\boldsymbol{v}(i)\boldsymbol{v}^{\top}(i)\}=\boldsymbol{V}(i)$ appears on the r.h.s. of (45).

$\boldsymbol{Q}_2$ Using some algebraic manipulations and assumption A4:

$$\begin{aligned}
\mathbb{E}\{\boldsymbol{v}(i)\boldsymbol{v}^{\top}(i)\boldsymbol{s}(i)\boldsymbol{s}^{\top}(i)\}&=\mathbb{E}\{\boldsymbol{v}(i)\boldsymbol{v}^{\top}(i)\}\mathbb{E}\{\boldsymbol{s}(i)\boldsymbol{s}^{\top}(i)\} \\
&=\boldsymbol{V}(i)\boldsymbol{R}_{ss}\,, \qquad (46)
\end{aligned}$$

where $\boldsymbol{R}_{ss}$ is calculated in closed-form in the supplementary material for this paper.

$\boldsymbol{Q}_3$ Using assumption A4, we obtain:

$$\begin{aligned}
\mathbb{E}\{\boldsymbol{v}(i)\boldsymbol{s}^{\top}(i)e_0(i)\}&=\mathbb{E}\{\boldsymbol{v}(i)\}\mathbb{E}\{\boldsymbol{s}^{\top}(i)e_0(i)\} \\
&=\mathbb{E}\{\boldsymbol{v}(i)\}\left(\boldsymbol{r}_{sy}-\boldsymbol{R}_{ss}\boldsymbol{\gamma}^*\right)^{\top}\,, \qquad (47)
\end{aligned}$$

where all these quantities have already been calculated.

$\boldsymbol{Q}_4$ Considering hypothesis A4 and assuming the input signal to be i.i.d., we have:

$$\begin{aligned}
&\big[\mathbb{E}\{\boldsymbol{s}(i)\boldsymbol{s}^\top(i)\boldsymbol{v}(i)\boldsymbol{s}^\top(i)e_0(i)\}\big]_{u,v} \\
&= \sum_{a=1}^{N|\mathcal{D}_n|} \mathbb{E}\{s_u(i)s_a(i)v_a(i)s_v(i)e_0(i)\} \\
&= \sum_{a=1}^{N|\mathcal{D}_n|} \mathbb{E}\{v_a(i)\}\mathbb{E}\{s_u(i)s_a(i)s_v(i)[y_n(i)-\boldsymbol{s}^\top(i)\boldsymbol{\gamma}^*]\} \\
&= -\sum_{b=1}^{N|\mathcal{D}_n|}\sum_{a=1}^{N|\mathcal{D}_n|} \mathbb{E}\{v_a(i)\}\mathbb{E}\{s_u(i)s_a(i)s_v(i)s_b(i)\}\gamma_b^* \\
&\quad + \sum_{a=1}^{N|\mathcal{D}_n|} \mathbb{E}\{v_a(i)\}\mathbb{E}\{s_u(i)s_a(i)s_v(i)y_n(i)\}\,, \qquad (48)
\end{aligned}$$

where both expectations $\mathbb{E}\{s_u(i)s_a(i)s_v(i)y_n(i)\}$ and $\mathbb{E}\{s_u(i)s_a(i)s_v(i)s_b(i)\}$ are provided in closed-form in the supplementary material for this paper.

$\boldsymbol{Q}_5$ This term can be written as:

$$\begin{aligned}
&\big[\mathbb{E}\{\boldsymbol{s}(i)\boldsymbol{s}^\top(i)e_0^2(i)\}\big]_{u,v} \\
&= \mathbb{E}\{s_u(i)s_v(i)[y_n(i)-\boldsymbol{s}^\top(i)\boldsymbol{\gamma}^*]^2\} \\
&= \mathbb{E}\{s_u(i)s_v(i)y_n^2(i)\} \\
&\quad - 2\sum_p \gamma_p^*\mathbb{E}\{s_u(i)s_v(i)s_p(i)y_n(i)\} \\
&\quad + \sum_\ell\sum_m \gamma_\ell^*\gamma_m^*\mathbb{E}\{s_u(i)s_v(i)s_\ell(i)s_m(i)\}\,. \qquad (49)
\end{aligned}$$

$\boldsymbol{Q}_6$ We use the strategy in [58] and assumption A4 in order to obtain the following approximation:

$$\begin{aligned}
&\big[\mathbb{E}\{\boldsymbol{s}(i)\boldsymbol{s}^\top(i)\boldsymbol{v}(i)\boldsymbol{v}^\top(i)\boldsymbol{s}(i)\boldsymbol{s}^\top(i)\}\big]_{u,v} \\
&= \sum_\ell\sum_m \mathbb{E}\{s_u(i)s_\ell(i)v_\ell(i)v_m(i)s_m(i)s_v(i)\} \\
&\approx \sum_\ell\sum_m \mathbb{E}\{s_u(i)s_\ell(i)s_m(i)s_v(i)\}\mathbb{E}\{v_\ell(i)v_m(i)\}\,, \\
&\qquad (50)
\end{aligned}$$

where $\mathbb{E}\{v_\ell(i)v_m(i)\} = \big[\boldsymbol{V}(i)\big]_{\ell,m}$. A closed-form expression of the expectation $\mathbb{E}\{s_u(i)s_\ell(i)s_m(i)s_v(i)\}$ is provided in the supplementary material for this paper.

$\boldsymbol{Q}_7$ Recalling that $\boldsymbol{v}(i) = \hat{\boldsymbol{\gamma}}(i) - \boldsymbol{\gamma}^*$, this term becomes:

$$\begin{aligned}
&\mathbb{E}\left\{\boldsymbol{v}(i)\left[\sum_{m=1}^N \frac{\hat{\boldsymbol{R}}_{tt,m}(i)\hat{\boldsymbol{\gamma}}(i)}{\Delta_m(i)}\right]^\top\right\} \\
&\quad = \sum_{m=1}^N \mathbb{E}\left\{\frac{\hat{\boldsymbol{\gamma}}(i)\hat{\boldsymbol{\gamma}}^\top(i)\hat{\boldsymbol{R}}_{tt,m}(i)}{\Delta_m(i)}\right\} \\
&\quad - \sum_{m=1}^N \boldsymbol{\gamma}^*\mathbb{E}\left\{\frac{\hat{\boldsymbol{\gamma}}^\top(i)\hat{\boldsymbol{R}}_{tt,m}(i)}{\Delta_m(i)}\right\}. \qquad (51)
\end{aligned}$$

The second term on the r.h.s. of (51) can be calculated with (36). To calculate the first term on the r.h.s. of (51), we use the following approximation:

$$\begin{aligned}
&\mathbb{E}\left\{\frac{\hat{\boldsymbol{\gamma}}(i)\hat{\boldsymbol{\gamma}}^\top(i)\hat{\boldsymbol{R}}_{tt,m}(i)}{\Delta_m(i)}\right\} \approx \mathbb{E}\left\{\frac{\hat{\boldsymbol{\gamma}}(i)\hat{\boldsymbol{\gamma}}^\top(i)\boldsymbol{R}_{tt,m}}{\Delta_m(i)}\right\} \\
&\quad \approx \frac{\big[\boldsymbol{\Sigma}(i)+\boldsymbol{\mu}(i)\,\boldsymbol{\mu}^\top(i)\big]\,\boldsymbol{R}_{tt,m}}{\sqrt{\mathrm{Tr}\{\boldsymbol{R}_{tt,m}\boldsymbol{\Sigma}(i)\}+\boldsymbol{\mu}^\top(i)\boldsymbol{R}_{tt,m}\boldsymbol{\mu}(i)}}\,, \qquad (52)
\end{aligned}$$

$\boldsymbol{Q}_8$ By using some algebraic manipulations and A4, we get:

$$\begin{aligned}
&\mathbb{E}\left\{\boldsymbol{s}(i)\boldsymbol{s}^\top(i)\boldsymbol{v}(i)\left[\sum_{m=1}^N \frac{\hat{\boldsymbol{R}}_{tt,m}(i)\hat{\boldsymbol{\gamma}}(i)}{\Delta_m(i)}\right]^\top\right\} \\
&\quad = \mathbb{E}\big\{\boldsymbol{s}(i)\boldsymbol{s}^\top(i)\big\}\mathbb{E}\left\{\boldsymbol{v}(i)\left[\sum_{m=1}^N \frac{\hat{\boldsymbol{R}}_{tt,m}(i)\hat{\boldsymbol{\gamma}}(i)}{\Delta_m(i)}\right]^\top\right\} \\
&\quad = \boldsymbol{R}_{ss}\boldsymbol{Q}_7\,. \qquad (53)
\end{aligned}$$

$\boldsymbol{Q}_9$ By approximating $\hat{\boldsymbol{R}}_{tt,m}(i) \approx \boldsymbol{R}_{tt,m}$ and using assumption A4, we obtain:

$$\begin{aligned}
&\mathbb{E}\left\{\boldsymbol{s}(i)e_0(i)\left[\sum_{m=1}^N \frac{\hat{\boldsymbol{R}}_{tt,m}(i)\hat{\boldsymbol{\gamma}}(i)}{\Delta_m(i)}\right]^\top\right\} \\
&\quad = \mathbb{E}\big\{\boldsymbol{s}(i)e_0(i)\big\}\,\mathbb{E}\left\{\sum_{m=1}^N \frac{\hat{\boldsymbol{R}}_{tt,m}(i)\hat{\boldsymbol{\gamma}}(i)}{\sqrt{\hat{\boldsymbol{\gamma}}^\top(i)\hat{\boldsymbol{R}}_{tt,m}(i)\hat{\boldsymbol{\gamma}}(i)}}\right\}^\top \\
&\quad = (\boldsymbol{r}_{sy} - \boldsymbol{R}_{ss}\boldsymbol{\gamma}^*) \qquad (54) \\
&\qquad \times \sum_{m=1}^N \frac{\boldsymbol{R}_{tt,m}\boldsymbol{\mu}(i)}{\sqrt{\mathrm{Tr}\{\boldsymbol{R}_{tt,m}\boldsymbol{\Sigma}(i)\}+\boldsymbol{\mu}^\top(i)\boldsymbol{R}_{tt,m}\boldsymbol{\mu}(i)}}.
\end{aligned}$$

$\boldsymbol{Q}_{10}$ We successively have:

$$\begin{aligned}
&\mathbb{E}\left\{\left[\sum_{m=1}^N \frac{\hat{\boldsymbol{R}}_{tt,m}(i)\hat{\boldsymbol{\gamma}}(i)}{\Delta_m(i)}\right]\left[\sum_{p=1}^N \frac{\hat{\boldsymbol{R}}_{tt,p}(i)\hat{\boldsymbol{\gamma}}(i)}{\Delta_p(i)}\right]^\top\right\} \\
&\approx \sum_{m=1}^N\sum_{p=1}^N \mathbb{E}\left\{\frac{\boldsymbol{R}_{tt,m}\hat{\boldsymbol{\gamma}}(i)\hat{\boldsymbol{\gamma}}^\top(i)\boldsymbol{R}_{tt,p}}{\sqrt{\hat{\boldsymbol{\gamma}}^\top(i)\boldsymbol{R}_{tt,m}\hat{\boldsymbol{\gamma}}(i)\hat{\boldsymbol{\gamma}}^\top(i)\boldsymbol{R}_{tt,p}\hat{\boldsymbol{\gamma}}(i)}}\right\} \\
&\approx \sum_{m=1}^N\sum_{p=1}^N \frac{\boldsymbol{R}_{tt,m}\left[\boldsymbol{\Sigma}(i)+\boldsymbol{\mu}(i)\,\boldsymbol{\mu}^\top(i)\right]\}\boldsymbol{R}_{tt,p}}{\sqrt{\mathbb{E}\{\hat{\boldsymbol{\gamma}}^\top(i)\boldsymbol{R}_{tt,m}\hat{\boldsymbol{\gamma}}(i)\hat{\boldsymbol{\gamma}}^\top(i)\boldsymbol{R}_{tt,p}\hat{\boldsymbol{\gamma}}(i)\}}}\,. \qquad (55)
\end{aligned}$$

Considering hypothesis A6, the denominator of this expression consists of the expectation of a linear combination of fourth order moments of Gaussian random variables. Using [64], we have:

$$\begin{aligned}
&\mathbb{E}\{\hat{\boldsymbol{\gamma}}^\top(i)\boldsymbol{R}_{tt,m}\hat{\boldsymbol{\gamma}}(i)\hat{\boldsymbol{\gamma}}^\top(i)\boldsymbol{R}_{tt,p}\hat{\boldsymbol{\gamma}}(i)\} \\
&= 4\boldsymbol{\mu}^\top\boldsymbol{R}_{tt,m}\boldsymbol{\Sigma}\boldsymbol{R}_{tt,p}\boldsymbol{\mu} + 2\mathrm{Tr}\big\{\boldsymbol{R}_{tt,m}\boldsymbol{\Sigma}\boldsymbol{R}_{tt,p}\boldsymbol{\Sigma}\big\} \\
&\quad + \Big(\boldsymbol{\mu}^\top\boldsymbol{R}_{tt,m}\boldsymbol{\mu} + \mathrm{Tr}\big\{\boldsymbol{R}_{tt,m}\boldsymbol{\Sigma}\big\}\Big) \\
&\quad \times \Big(\boldsymbol{\mu}^\top\boldsymbol{R}_{tt,p}\boldsymbol{\mu} + \mathrm{Tr}\big\{\boldsymbol{R}_{tt,p}\boldsymbol{\Sigma}\big\}\Big)\,, \qquad (56)
\end{aligned}$$

To conclude, using the results from relations (46)–(55), we arrive at the following recursive equation for $\boldsymbol{V}(i)$:

$$\begin{aligned}
\boldsymbol{V}(i+1) &= \boldsymbol{V}(i) - \mu\,\mathrm{sym}\{\boldsymbol{V}(i)\boldsymbol{R}_{ss}\} + \mu\,\mathrm{sym}\{\boldsymbol{Q}_3\} \\
&\quad - \mu^2\,\mathrm{sym}\{\boldsymbol{Q}_4\} + \mu^2\boldsymbol{Q}_6 - \mu\eta\,\mathrm{sym}\{\boldsymbol{Q}_7\} + \mu^2\eta^2\boldsymbol{Q}_{10} \\
&\quad + \mu^2\boldsymbol{Q}_5 + \mu^2\eta\,\mathrm{sym}\{\boldsymbol{R}_{ss}\boldsymbol{Q}_7\} - \mu^2\eta\,\mathrm{sym}\{\boldsymbol{Q}_9\}\,. \qquad (57)
\end{aligned}$$

The unknown moments are calculated in closed-form in the supplementary material for this paper.

Before going further, we define the Mean Square Deviation:

$$\begin{aligned}\text{MSD}(i) &\triangleq \mathbb{E}\left\{\|\hat{\boldsymbol{\gamma}}(i)-\boldsymbol{\gamma}^*\|^2\right\}\\ &= \mathbb{E}\left\{\|\boldsymbol{v}(i)\|^2\right\} = \text{Tr}\left\{\boldsymbol{V}(i)\right\}\,. \end{aligned} \tag{58}$$

This quantity is useful in determining the performance of the algorithm.

F. Steady-state performance when $\eta = 0$

Assuming a small enough step-size μ that verifies (41), with relation (40) we have:

$$\begin{aligned}\boldsymbol{v}(\infty) &= \lim_{i\to\infty}\mathbb{E}\left\{\boldsymbol{v}(i)\right\}\\ &= \boldsymbol{0}\,.\end{aligned} \tag{59}$$

We shall now evaluate $\boldsymbol{V}(\infty) = \lim_{i\to\infty}\boldsymbol{V}(i)$ using (57). With $\boldsymbol{v}(\infty)=\boldsymbol{0}$, we know that $\boldsymbol{Q}_3(\infty)=\boldsymbol{Q}_4(\infty)=\boldsymbol{0}$. Using vectorization of (57) together with the Kronecker product gives:

$$\text{vec}\{\boldsymbol{V}(\infty)\} = \mu^2\left(\boldsymbol{I}_2-\boldsymbol{F}_0\right)^{-1}\text{vec}\{\boldsymbol{Q}_5\}\,, \tag{60}$$

with:

$$\boldsymbol{F}_0 = \boldsymbol{I}_2 - \mu(\boldsymbol{I}\otimes\boldsymbol{R}_{ss}+\boldsymbol{R}_{ss}\otimes\boldsymbol{I})+\mu^2\boldsymbol{F}_1\,, \tag{61}$$

where operator $\otimes$ is the Kronecker product. In these expressions, $\boldsymbol{I}_2$ is the identity matrix of size $k_s^2\times k_s^2$, and $\boldsymbol{I}$ is the identity matrix of size $k_s\times k_s$, where $k_s=(N+1)|\mathcal{D}_n|$ is the number of entries in block vector $\boldsymbol{s}$. The entries of $\boldsymbol{F}_1$ are:

$$[\boldsymbol{F}_1]_{u+(\ell-1)k_s,m+(v-1)k_s} = \mathbb{E}\{s_u(i)s_\ell(i)s_m(i)s_v(i)\}\,. \tag{62}$$

They are provided in closed-form in the supplementary material for this paper.

Finally, we conclude that:

$$\text{MSD}(\infty) = \text{Tr}\left\{\text{vec}^{-1}\left\{\mu^2\left(\boldsymbol{I}_2-\boldsymbol{F}_0\right)^{-1}\text{vec}\{\boldsymbol{Q}_5\}\right\}\right\}\,. \tag{63}$$

V. Experimental validation

The performance of algorithm (22) has already been illustrated in the preliminary work [12], both with simulated and real-world data. Consequently, the experiments in this section will be restricted to model validation.

A. Signal model

We considered a simulation scenario with i.i.d. zero-mean Gaussian data $\boldsymbol{y}(i)$ with covariance matrix $\boldsymbol{R}_{yy}$ satisfying the following model:

$$\boldsymbol{y}(i) = \boldsymbol{A}\boldsymbol{y}(i)+\boldsymbol{v}(i)\,, \tag{64}$$

where $\boldsymbol{A}$ is the adjacency matrix defined as:

$$\boldsymbol{A} = \begin{pmatrix}0&1&0&1&1\\1&0&1&0&1\\1&0&0&1&0\\0&1&1&0&1\\1&0&1&1&0\end{pmatrix}\,. \tag{65}$$

and $\boldsymbol{v}(i)$ is a zero-mean white Gaussian noise with covariance matrix $\sigma_v^2\boldsymbol{I}_5$ and $\sigma_v=0.05$. Note that $\boldsymbol{R}_{yy}$ satisfies:

$$\boldsymbol{R}_{yy} = [\boldsymbol{I}-\boldsymbol{A}]^{-1}\boldsymbol{R}_{vv}([\boldsymbol{I}-\boldsymbol{A}]^{-1})^\top\,. \tag{66}$$

This model, although not nonlinear, offers exact knowledge of the statistical properties necessary to validate the models derived in this paper. Validation on a purely nonlinear system follows in subsection V-E. The reader is also referred to [12] to assess the methods' behavior in real-world nonlinear settings.

We used our algorithm with the Gaussian kernel (30), with kernel bandwidth $\sigma=1$. Each node stored a dictionary $\mathcal{D}_n$ with 6 elements chosen uniformly in $[-1,1]^4$. Simulations were averaged over 100 Monte-Carlo runs. Results presented hereafter focus on node $n=1$.

B. Computing the optimal coefficients

In order to compute the optimal coefficients, we note that problem (12) can be written equivalently as:

$$\min_{\boldsymbol{\gamma}}\ \frac{1}{2}\boldsymbol{\gamma}^\top\boldsymbol{R}_{ss}\boldsymbol{\gamma}-\boldsymbol{\gamma}^\top\boldsymbol{r}_{sy}+\eta\sum_{m=1}^{N}\|\boldsymbol{C}_m^\top\boldsymbol{\gamma}\|_2\,, \tag{67}$$

where $\boldsymbol{C}_m\boldsymbol{C}_m^\top=\boldsymbol{R}_{tt,m}$. We solved this optimization problem using the CVX package to get $\boldsymbol{\gamma}^*$ [65], [66].

The MSD was evaluated experimentally using:

$$\text{MSD}(i) = \mathbb{E}\left\{\|\hat{\boldsymbol{\gamma}}(i)-\boldsymbol{\gamma}^*\|^2\right\}\,. \tag{68}$$

C. Parameter $\eta = 0$

Figure 1(a) shows the theoretical and experimental learning curves of the entries of the parameter vector $\boldsymbol{\gamma}$. Figure 1(b) shows both the theoretical and experimental MSD curves, as well as the steady-state MSD. The theoretical curves closely follow the theoretical ones, both in the analyses in the mean and mean square sense. They also validate the approximation (36) of the regularization term which is in the update (34).

D. Parameter $\eta > 0$

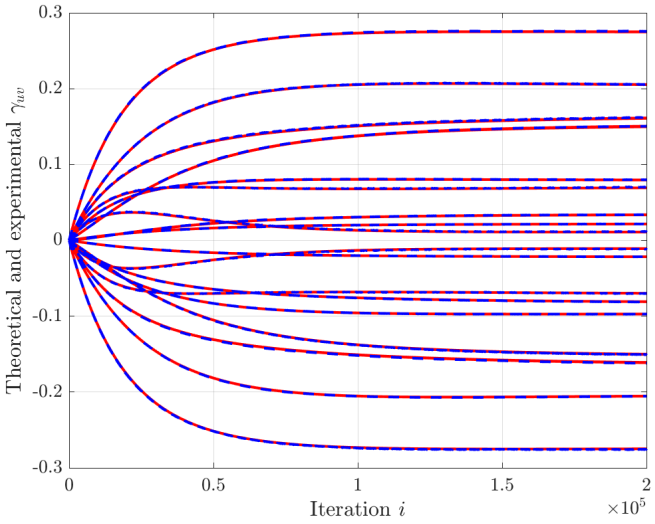
The conditions of this experiment were the same as for the previous case, with $\eta = 1\cdot 10^{-4}$.

The learning curves are presented in Fig. 2. As for the previous case, we observe that the theoretical curves closely fit the experimental ones. This validates the assumptions and approximations employed in the analysis to make it tractable.

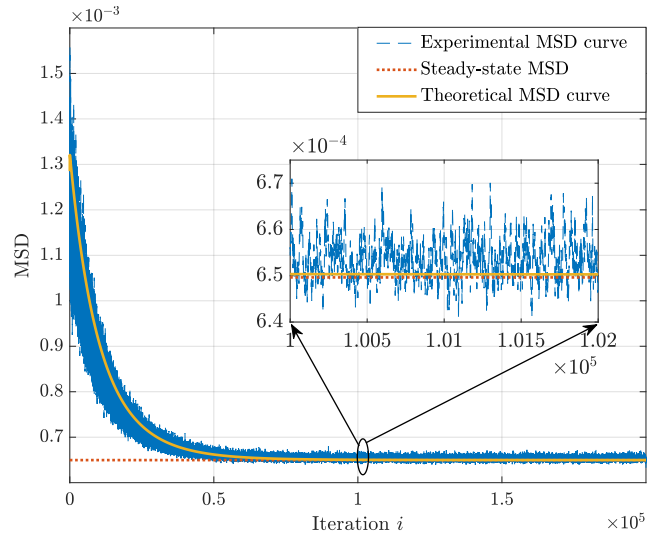
E. Validation with a nonlinear model

In order to evaluate how well the model can predict the the behavior of the algorithm in nonlinear settings, we consider an additional example with a 3-node graph with adjacency matrix given by:

$$\boldsymbol{A} = \begin{pmatrix}0&0&1\\1&0&1\\1&0&0\end{pmatrix}\,, \tag{69}$$

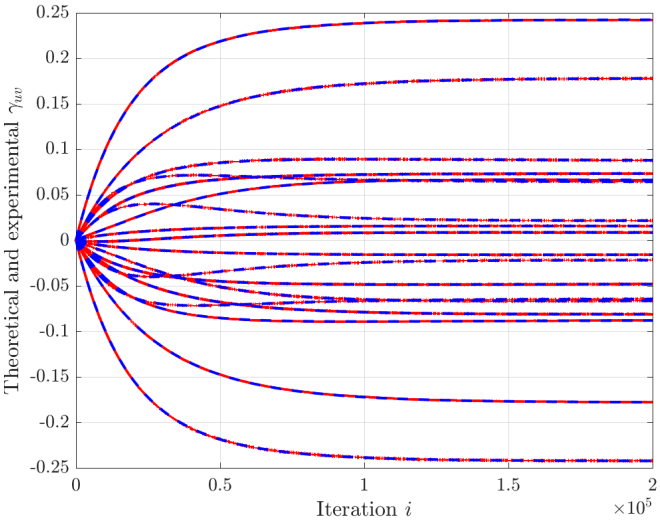


(a) Theoretical and experimental entries of $\boldsymbol{\gamma}$. Blue dashed lines are the experimental curves, while red lines are the theoretical ones

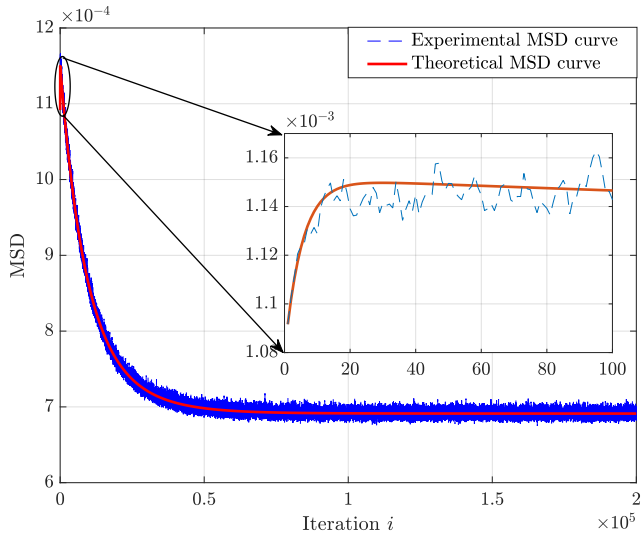


(b) Experimental, steady-state, and theoretical MSD

Fig. 1. Analysis validation in mean and mean square sense, for $\eta = 0$



(a) Theoretical and experimental entries of $\boldsymbol{\gamma}$. Blue dashed lines are the experimental curves, while red lines are the theoretical ones



(b) Experimental and theoretical MSD

Fig. 2. Analysis validation in mean and mean square sense, for $\eta = 1 \cdot 10^{-4}$

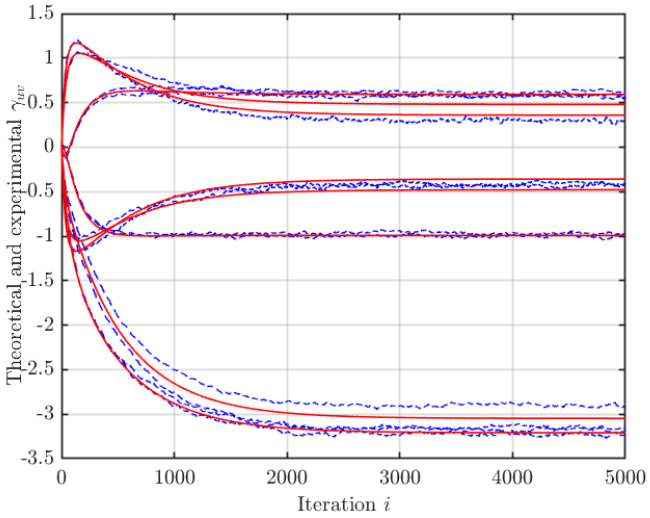
and i.i.d. streaming nodal data generating to agree with the following model:

$$\boldsymbol{y}(i) = \underbrace{\begin{bmatrix} y_1 - K_1(y_3+y_1)^3(K_2 y_1)^{-1} \\ y_2 + \frac{y_2 - K_1(y_3+y_1)^3(K_2y_1)^{-1}}{[0.5+\exp(K_1(y_3+y_1)^3(K_2y_1)^{-1})]^5+1} \\ y_3 + y_1 + K_1(y_3+y_1)^3(K_2y_1)^{-1} \end{bmatrix}}_{\boldsymbol{f}(\boldsymbol{y}(i))} + \boldsymbol{\varrho}(i)$$

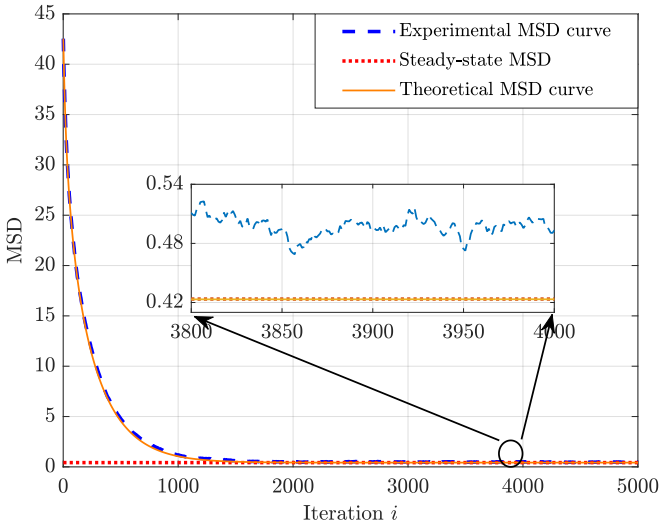
with $K_1 = 8000$ and $K_2 = 27$, $\boldsymbol{\varrho}(i)$ being an i.i.d. zero-mean Gaussian random vector with covariance matrix $\boldsymbol{I}_3$, and $\boldsymbol{f}(\boldsymbol{y}(i)) = \left[f_1(\boldsymbol{y}(i)), f_2(\boldsymbol{y}(i)), f_3(\boldsymbol{y}(i))\right]^\top$ the vector-valued function corresponding to the concatenation of f_n in (2). Notice that the sample index i was omitted inside $\boldsymbol{f}(\boldsymbol{y}(i))$ to simplify the notation. Graph signal $\boldsymbol{y}(i)$ can be sampled from this model by first sampling from $\boldsymbol{\varrho}(i)$ and using the relation $\boldsymbol{y}(i) = (\mathrm{id} - \boldsymbol{f})^{-1}(\boldsymbol{\varrho}(i))$, with $\mathrm{id}(\boldsymbol{x})$ the identity function.

We used our algorithm with the Gaussian kernel (30), with kernel bandwidth $\sigma = 1$. Each node stored a dictionary $\mathcal{D}_n$ with 4 elements chosen uniformly in $[-1, 1] \times [-1, 1]$. Simulations were averaged over 100 Monte-Carlo runs. Results presented hereafter focus on node $n = 1$.

Figure 3 depicts the behavior of the algorithm in mean and mean-square sense without regularization (i.e., for $\eta = 0$). Figure 4 shows the mean and mean-square performance in the case when $\eta = 0.3$. These two figures together show that the conducted analysis is indeed able to predict the behavior of the algorithm without the need of running multiple Monte-Carlo runs. Moreover, even with the highly nonlinear nature of $\boldsymbol{f}$ and $\boldsymbol{y}(i)$ not being Gaussian distributed, the theoretical curves still follow the experimental ones relatively closely, although a small amount of bias can be observed in steady state.

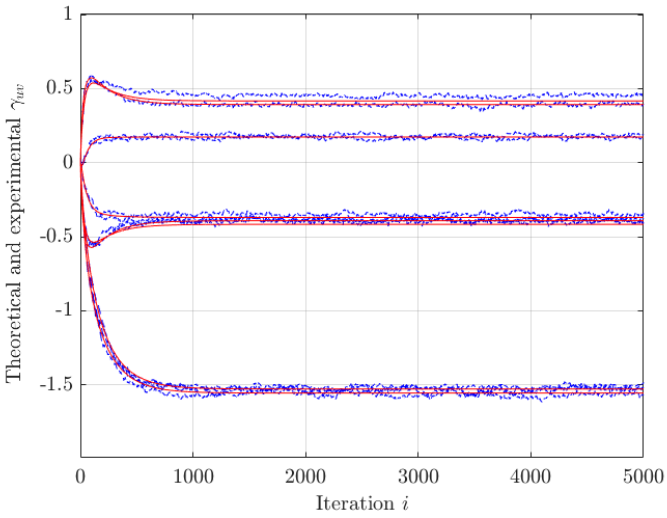


(a) Theoretical and experimental entries of $\boldsymbol{\gamma}$. Blue dashed lines are the experimental curves, while red lines are the theoretical ones

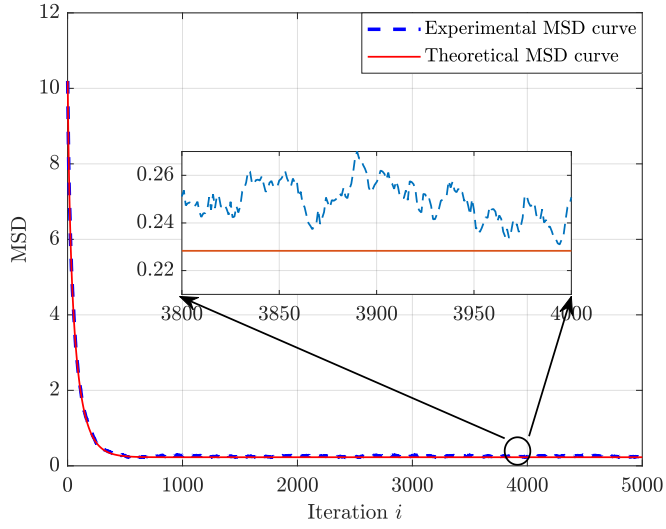


(b) Experimental, steady-state, and theoretical MSD

Fig. 3. Analysis validation in mean and mean-square sense, for $\eta = 0$



(a) Theoretical and experimental entries of $\boldsymbol{\gamma}$. Blue dashed lines are the experimental curves, while red lines are the theoretical ones



(b) Experimental and theoretical MSD

Fig. 4. Analysis validation in mean and mean-square sense, for $\eta = 0.3$

VI. Conclusion

In this paper, an online graph-topology inference method recently proposed by the authors was recalled and analyzed. While previous works mainly focused on models based on additive interactions between the signals at each node, we considered arbitrary nonlinear interactions between the nodes, which render our model much more general. By encoding the links as partial derivatives of some nonlinear functions, our method benefits from the kernel machinery framework to estimate a possibly directed, sparse adjacency matrix. However, it also demonstrates significantly increased complexity that makes its analysis non-trivial. This work proposed a thorough analysis of the algorithm, as well as performance bounds and stability conditions. The derived models characterize the performance of the algorithm as a function of important parameters such as the kernel bandwidth or the step-size. In turn, this allows for precise tuning of such parameters in order to obtain a desired performance in transient or in steady state. Moreover, the behavior of the intuitive partial-derivative-based sparsity regularizer was analyzed in-depth, thus completely characterizing the method. The experimental results, as well as the algorithm analysis, show that the proposed method can lead to more accurate estimates for more general nonlinear systems.

References

[1] The International HapMap Consortium, K. Frazer, D. Ballinger *et al.*, "A second generation human haplotype map of over 3.1 million SNPs," *Nature*, vol. 449, pp. 851 EP –, 10 2007. [Online]. Available: https://doi.org/10.1038/nature06258

[2] R. H. Heiberger, "Predicting economic growth with stock networks," *Physica A: Statistical Mechanics and its Applications*, vol. 489, pp. 102–111, 2018.

[3] M. Kramer, E. D. Kolaczyk, and H. Kirsch, "Emergent network topology at seizure onset in humans," *Epilepsy Research*, vol. 79, pp. 173–86, 2008.

[4] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Graph filters," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 6163–6166.

[5] R. Nassif, C. Richard, J. Chen, and A. H. Sayed, "A graph diffusion LMS strategy for adaptive graph signal processing," in *Proc. Asilomar Conference on Signals, Systems and Computers*, 2017, pp. 1973–1976.

[6] J. Chen, C. Richard, and A. H. Sayed, "Multitask diffusion adaptation over networks with common latent representations," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 3, pp. 563–579, 2017.

[7] M. Coutino, E. Isufi, and G. Leus, "Advances in distributed graph filtering," *arXiv preprint arXiv:1808.03004*, 2018.

[8] A. Koppel, S. Paternain, C. Richard, and A. Ribeiro, "Decentralized online learning with kernels," *IEEE Transactions on Signal Processing*, vol. 66, no. 12, pp. 3240–3255, 2018.

[9] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1644–1656, 2013.

[10] N. Tremblay, G. Puy, P. Borgnat, R. Gribonval, and P. Vandergheynst, "Accelerated spectral clustering using graph filtering of random signals," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 4094–4098.

[11] P. M. Djuric and C. Richard, *Cooperative and Graph Signal Processing: Principles and Applications*. Academic Press, 2018.

[12] M. Moscu, R. Borsoi, and C. Richard, "Online kernel-based graph topology identification with partial-derivative-imposed sparsity," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2021, pp. 2190–2194.

[13] M. Speriosu, N. Sudan, S. Upadhyay, and J. Baldridge, "Twitter polarity classification with label propagation over lexical links and the follower graph," in *Proc. Workshop on Unsupervised Learning in NLP*, 2011, pp. 53–63.

[14] A. P. Dempster, "Covariance selection," *Biometrics*, vol. 28, no. 1, pp. 157–175, 1972.

[15] J. Friedman, T. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical Lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–41, 2008.

[16] S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro, "Network topology inference from spectral templates," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 3, no. 3, pp. 467–483, 2017.

[17] S. Sardellitti, S. Barbarossa, and P. D. Lorenzo, "Graph topology inference based on sparsifying transform learning," *IEEE Transactions on Signal Processing*, vol. 67, no. 7, pp. 1712–1727, 2019.

[18] R. Shafipour, S. Segarra, A. G. Marques, and G. Mateos, "Network topology inference from non-stationary graph signals," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 5870–5874.

[19] M. Moscu, R. Nassif, F. Hua, and C. Richard, "Learning causal networks topology from streaming graph signals," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5.

[20] R. Shafipour, A. Hashemi, G. Mateos, and H. Vikalo, "Online topology inference from streaming stationary graph signals," *Proc. IEEE Data Science Workshop*, pp. 140–144, 2019.

[21] M. Coutino, G. V. Karanikolas, G. Leus, and G. B. Giannakis, "Self-driven graph Volterra models for higher-order link prediction," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 3887–3891.

[22] Q. Yang, M. Coutino, G. Wang, G. B. Giannakis, and G. Leus, "Learning connectivity and higher-order interactions in radial distribution grids," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 5555–5559.

[23] J. Harring, B. Weiss, and J.-C. Hsu, "A comparison of methods for estimating quadratic effects in nonlinear structural equation models," *Psychological Methods*, vol. 17, pp. 193–214, 2012.

[24] W. H. Finch, "Modeling nonlinear structural equation models: A comparison of the two-stage generalized additive models and the finite mixture structural equation model," *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 22, no. 1, pp. 60–75, 2015.

[25] N. Lim, F. D'Alché-Buc, C. Auliac, and G. Michailidis, "Operator-valued kernel-based vector autoregressive models for network inference," *Machine Learning*, vol. 99, no. 3, pp. 489–513, 2015.

[26] C. Lippert, O. Stegle, Z. Ghahramani, and K. Borgwardt, "A kernel method for unsupervised structured network inference," in *Artificial Intelligence and Statistics*, 2009, pp. 368–375.

[27] Y. Shen, B. Baingana, and G. B. Giannakis, "Topology inference of directed graphs using nonlinear structural vector autoregressive models," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 6513–6517.

[28] L. Zhang, G. Wang, and G. B. Giannakis, "Going beyond linear dependencies to unveil connectivity of meshed grids," in *Proc. IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, 2017, pp. 1–5.

[29] G. B. Giannakis, Y. Shen, and G. V. Karanikolas, "Topology Identification and Learning over Graphs: Accounting for Nonlinearities and Dynamics," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 787–807, 2018.

[30] X. Dong, D. Thanou, M. Rabbat, and P. Frossard, "Learning graphs from data: A signal representation perspective," *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 44–63, 2019.

[31] N. Biggs, *Algebraic Graph Theory*. Cambridge University Press, 1993.

[32] A. Sandryhaila and J. M. F. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 80–90, 2014.

[33] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.

[34] Y. Shen, G. B. Giannakis, and B. Baingana, "Nonlinear structural vector autoregressive models with application to directed brain networks," *IEEE Transactions on Signal Processing*, vol. 67, no. 20, pp. 5325–5339, 2019.

[35] W. J. Freeman, "EEG analysis gives model of neuronal template-matching mechanism for sensory search with olfactory bulb," *Biological Cybernetics*, vol. 35, no. 4, pp. 221–234, 1979.

[36] J. A. de Zwart, P. van Gelderen, J. M. Jansma, M. Fukunaga, M. Bianciardi, and J. H. Duyn, "Hemodynamic nonlinearities affect BOLD fMRI response timing and amplitude," *Neuroimage*, vol. 47, no. 4, pp. 1649–1658, 2009.

[37] Y. Shen, B. Baingana, and G. B. Giannakis, "Kernel-based structural equation models for topology identification of directed networks," *IEEE Transactions on Signal Processing*, vol. 65, no. 10, pp. 2503–2516, 2017.

[38] M. Moscu, R. Borsoi, and C. Richard, "Online graph topology inference with kernels for brain connectivity estimation," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2020.

[39] S. Petkoski and V. K. Jirsa, "Transmission time delays organize the brain network synchronization," *A Philosophical Transactions of the Royal Society*, vol. 377, no. 2153, p. 20180132, 2019.

[40] M. Moscu, R. Borsoi, and C. Richard, "Convergence analysis of the graph-topology-inference kernel LMS algorithm," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021.

[41] Y. Shen and G. B. Giannakis, "Online identification of directional graph topologies capturing dynamic and nonlinear dependencies," in *Proc. IEEE Data Science Workshop*, 2018, pp. 195–199.

[42] A. Buja, T. Hastie, and R. Tibshirani, "Linear smoothers and additive models," *The Annals of Statistics*, pp. 453–510, 1989.

[43] M. Danisch, O. Balalau, and M. Sozio, "Listing k-cliques in sparse real-world graphs," in *Proc. World Wide Web Conference*, 2018, pp. 589–598.

[44] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Optimization with sparsity-inducing penalties," *Foundations and Trends in Machine Learning*, vol. 4, no. 1, pp. 1–106, 2011.

[45] L. Rosasco, S. Villa, S. Mosci, M. Santoro, and A. Verri, "Nonparametric sparsity and regularization," *Journal of Machine Learning Research*, vol. 14, pp. 1665–1714, 2013.

[46] D.-X. Zhou, "Derivative reproducing properties for kernel methods in learning theory," *Journal of Computational and Applied Mathematics*, vol. 220, no. 1-2, pp. 456–463, 2008.

[47] J. Chen, W. Gao, C. Richard, and J.-C. M. Bermudez, "Convergence analysis of kernel LMS algorithm with pre-tuned dictionary," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014.

[48] W. Gao, J. Chen, C. Richard, J. Huang, and R. Flamary, "Kernel lms algorithm with forward-backward splitting for dictionary learning," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013.

[49] Y. Engel, S. Mannor, and R. Meir, "The kernel recursive least-squares algorithm," *IEEE Transactions on Signal Processing*, vol. 52, no. 8, pp. 2275–2285, 2004.

[50] A. A. Bueno and M. T. M. Silva, "Gram-Schmidt-based sparsification for kernel dictionary," *IEEE Signal Processing Letters*, vol. 27, pp. 1130–1134, 2020.

[51] C. K. I. Williams and M. Seeger, "Using the Nyström method to speed up kernel machines," in *Proc. Advances in Neural Information Processing Systems 13*, T. K. Leen, T. G. Dietterich, and V. Tresp, Eds. MIT Press, 2001, pp. 682–688. [Online]. Available: http://papers.nips.cc/paper/1866-using-the-nystrom-method-to-speed-up-kernel-machines.pdf

[52] P. Bouboulis, S. Pougkakiotis, and S. Theodoridis, "Efficient KLMS and KRLS algorithms: A random Fourier feature perspective," in *Proc. IEEE Statistical Signal Processing Workshop (SSP)*, 2016, pp. 1–5.

[53] M. Takizawa and M. Yukawa, "Adaptive nonlinear estimation based on parallel projection along affine subspaces in reproducing kernel hilbert space," *IEEE Transactions on Signal Processing*, vol. 63, no. 16, pp. 4257–4269, 2015.

[54] J. Platt, "A resource-allocating network for function interpolation," *Neural Computation*, vol. 3, no. 2, pp. 213–225, 1991.

[55] C. Richard, J.-C. M. Bermudez, and P. Honeine, "Online prediction of time series data with kernels," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1058–1067, 2009.

[56] O. Ledoit and M. Wolf, "Nonlinear shrinkage estimation of large-dimensional covariance matrices," *The Annals of Statistics*, vol. 40, no. 2, pp. 1024 – 1060, 2012.

[57] W. Liu, J. C. Príncipe, and S. Haykin, *Kernel adaptive filtering: a comprehensive introduction*. John Wiley & Sons, Ltd, 2010, ch. 1, pp. 1–26. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470608593.ch1

[58] W. D. Parreira, J.-C. M. Bermudez, C. Richard, and J.-Y. Tourneret, "Stochastic behavior analysis of the Gaussian kernel least-mean-square algorithm," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2208–2222, 2012.

[59] J. Minkoff, "Comment on the "Unnecessary assumption of statistical independence between reference signal and filter weights in feedforward adaptive systems"," *IEEE Transactions on Signal Processing*, vol. 49, no. 5, pp. 1109–, 2001.

[60] Y. Chen, Y. Gu, and A. O. Hero, "Sparse LMS for system identification," in *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2009, pp. 3125–3128.

[61] J. Chen, C. Richard, Y. Song, and D. Brie, "Transient performance analysis of zero-attracting LMS," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1786–1790, 2016.

[62] W. Gao and J. Chen, "Transient performance analysis of zero-attracting Gaussian kernel LMS algorithm with pre-tuned dictionary," *IEEE Access*, vol. 7, pp. 135 770–135 779, 2019.

[63] R. C. Elandt-Johnson and N. L. Johnson, *Survival Models and Data Analysis*. John Wiley & Sons, 1999, vol. 74.

[64] A. Kumar, "Expectation of product of quadratic forms," *Sankhyā: The Indian Journal of Statistics, Series B*, pp. 359–362, 1973.

[65] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[66] ——, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, ser. Lecture Notes in Control and Information Sciences, V. Blondel, S. Boyd, and H. Kimura, Eds. Springer-Verlag Limited, 2008, pp. 95–110, http://stanford.edu/~boyd/graph_dcp.html.

APPENDIX A
PROOF OF THEOREM 1

Proof. Using orthogonal projection, we can decompose any function $f_n \in \mathcal{H}_\kappa$ as the sum of two functions:

$$f_n = f_n^{\parallel} + f_n^{\perp}\,, \tag{70}$$

where $f_n^{\perp}$ is orthonormal to $f_n^{\parallel}$, that is, $\langle f_n^{\parallel}, f_n^{\perp}\rangle_{\mathcal{H}} = 0$), and $f_n^{\parallel}$ can be written as:

$$f_n^{\parallel} = \sum_{\ell=1}^{i} \alpha_\ell \kappa\left(\cdot, \boldsymbol{y}(\ell)\right) + \sum_{m=1}^{N}\sum_{p=1}^{i}\sum_{q=1}^{L_m} \beta_{m,p,q}\kappa_{\partial_{m,q}}\left(\cdot, \boldsymbol{y}(p)\right)\,, \tag{71}$$

where α_ℓ and $\beta_{m,p,q}$ are coefficients. This means that $f_n^{\parallel}$ lies in the span of $\kappa\left(\cdot, \boldsymbol{y}(\ell)\right)$ and $\kappa_{\partial_{m,q}}\left(\cdot, \boldsymbol{y}(p)\right)$, from which the orthogonality condition implies that $\langle f_n^{\perp}, \kappa\left(\cdot, \boldsymbol{y}(\ell)\right)\rangle_{\mathcal{H}} = 0$ and $\langle f_n^{\perp}, \kappa_{\partial_{m,q}}\left(\cdot, \boldsymbol{y}(p)\right)\rangle_{\mathcal{H}} = 0$, for all ℓ, m, p, q.

Assume that the kernel $\kappa(\cdot,\cdot)$ is at least twice differentiable. Then, the following relation holds [46]:

$$\mathcal{H} \ni \frac{\partial f_n(\boldsymbol{y})}{\partial y_{m,q}} = \langle f_n, \kappa_{\partial_{m,q}}(\cdot, \boldsymbol{y})\rangle_{\mathcal{H}}\,, \tag{72}$$

where $y_{m,q}$ is the q-th entry of $\boldsymbol{y}_m$.

Let us plug then decomposition (70) of f_n in problem (4). For the first term, using the reproducing property, we have:

$$\begin{aligned} f_n(\boldsymbol{y}(\ell)) &= \langle f_n, \kappa(\cdot, \boldsymbol{y}(\ell))\rangle_{\mathcal{H}} && (73)\\ &= \langle f_n^{\parallel} + f_n^{\perp}, \kappa(\cdot, \boldsymbol{y}(\ell))\rangle_{\mathcal{H}} && (74)\\ &= \langle f_n^{\parallel}, \kappa(\cdot, \boldsymbol{y}(\ell))\rangle_{\mathcal{H}}\,, && (75)\end{aligned}$$

for all $\ell = 1, \ldots, i$, where $\langle f_n^{\perp}, \kappa(\cdot, \boldsymbol{y}(\ell))\rangle_{\mathcal{H}} = 0$ because $f_n^{\perp}$ is orthogonal to each term in (71). For the second term:

$$\begin{aligned} \frac{\partial f_n(\boldsymbol{y}(p))}{\partial y_{m,q}} &= \left.\frac{\partial f_n(\boldsymbol{y})}{\partial y_{m,q}}\right|_{\boldsymbol{y}=\boldsymbol{y}(p)} && (76)\\ &= \langle f_n^{\parallel} + f_n^{\perp}, \kappa_{\partial_{m,q}}(\cdot, \boldsymbol{y}(p))\rangle_{\mathcal{H}} && (77)\\ &= \langle f_n^{\parallel}, \kappa_{\partial_{m,q}}(\cdot, \boldsymbol{y}(p))\rangle_{\mathcal{H}}\,, && (78)\end{aligned}$$

for all $p = 1, \ldots, i$, where $\langle f_n^{\perp}, \kappa_{\partial_{m,q}}(\cdot, \boldsymbol{y}(p))\rangle_{\mathcal{H}} = 0$ since $f_n^{\perp}$ is perpendicular to each term in (71). For the last term, we have:

$$\begin{aligned} \psi\big(\|f_n\|_{\mathcal{H}}\big) &= \psi\big(\|f_n^{\parallel} + f_n^{\perp}\|_{\mathcal{H}}\big) && (79)\\ &= \psi\big(\|f_n^{\parallel}\|_{\mathcal{H}} + \|f_n^{\perp}\|_{\mathcal{H}}\big)\,. && (80)\end{aligned}$$

Since $\|f_n^{\perp}\|_{\mathcal{H}}$ does not influence the first two terms in problem (4), if ψ is monotonically increasing, then the solution is such that $\|f_n^{\perp}\|_{\mathcal{H}} = 0$. □

APPENDIX B
COMPUTING THE TERMS IN THE OPTIMIZATION PROBLEM THAT DEPEND ON $f_{n,\mathcal{D}_n}$

Let us recall the optimization problem (11):

$$\min_{f_{n,\mathcal{D}_n}} \frac{1}{2}\mathbb{E}\Big\{\Big\|y_n(i) - f_{n,\mathcal{D}_n}(\boldsymbol{y}(i))\Big\|^2\Big\} + \eta\sum_{m=1}^{N}\sqrt{\mathbb{E}\left\{\left(\frac{\partial f_{n,\mathcal{D}_n}(\boldsymbol{y}(i))}{\partial y_m(i)}\right)^2\right\}}\,. \tag{81}$$

We now evaluate the terms involving $f_{n,\mathcal{D}_n}$, using the reproducing property and (5). For the first term, we have:

$$\begin{aligned} f_{n,\mathcal{D}_n}(\boldsymbol{y}(i)) &= \big\langle f_{n,\mathcal{D}_n}, \kappa(\cdot, \boldsymbol{y}(i))\big\rangle_{\mathcal{H}} && (82)\\ &= \sum_{p=1}^{|\mathcal{D}_n|}\alpha_p\big\langle\kappa(\cdot, \boldsymbol{y}(\omega_p)), \kappa(\cdot, \boldsymbol{y}(i))\big\rangle_{\mathcal{H}}\\ &\quad + \sum_{m=1}^{N}\sum_{q=1}^{|\mathcal{D}_n|}\beta_{m,q}\big\langle\kappa_{\partial_m}(\cdot, \boldsymbol{y}(\omega_q)),\, \kappa(\cdot, \boldsymbol{y}(i))\big\rangle_{\mathcal{H}}\,,\end{aligned}$$

from which:

$$\big\langle\kappa(\cdot, \boldsymbol{y}(\omega_p)), \kappa(\cdot, \boldsymbol{y}(i))\big\rangle_{\mathcal{H}} = \kappa(\boldsymbol{y}(\omega_p), \boldsymbol{y}(i))\,, \tag{83a}$$

$$\big\langle\kappa_{\partial_m}(\cdot, \boldsymbol{y}(\omega_q)), \kappa(\cdot, \boldsymbol{y}(i))\big\rangle_{\mathcal{H}} = \frac{\partial\kappa(\boldsymbol{y}(\omega_q), \boldsymbol{y}(i))}{\partial y_m(\omega_q)}\,. \tag{83b}$$

For the second term, we have:

$$\frac{\partial f_{n,\mathcal{D}_n}(\boldsymbol{y}(i))}{\partial y_{m_2}(i)} = \left\langle f_{n,\mathcal{D}_n}, \kappa_{\partial_{m_2}}(\cdot, \boldsymbol{y}(i)) \right\rangle_{\mathcal{H}}$$
$$= \sum_{p=1}^{|\mathcal{D}_n|} \alpha_p \left\langle \kappa(\cdot, \boldsymbol{y}(\omega_p)), \kappa_{\partial_{m_2}}(\cdot, \boldsymbol{y}(i)) \right\rangle_{\mathcal{H}}$$
$$+ \sum_{m_2=1}^{N} \sum_{q=1}^{|\mathcal{D}_n|} \beta_{m_2,q} \left\langle \kappa_{\partial_{m_2}}(\cdot, \boldsymbol{y}(\omega_q)), \kappa_{\partial_{m_2}}(\cdot, \boldsymbol{y}(i)) \right\rangle_{\mathcal{H}}, \quad (84)$$

from which:

$$\left\langle \kappa(\cdot, \boldsymbol{y}(\omega_p)), \kappa_{\partial_{m_2}}(\cdot, \boldsymbol{y}(i)) \right\rangle_{\mathcal{H}} = \frac{\partial \kappa(\boldsymbol{y}(\omega_p), \boldsymbol{y}(i))}{\partial y_{m_2}(i)}, \quad (85a)$$

$$\left\langle \kappa_{\partial_{m_1}}(\cdot, \boldsymbol{y}(\omega_q)), \kappa_{\partial_{m_2}}(\cdot, \boldsymbol{y}(i)) \right\rangle_{\mathcal{H}} = \frac{\partial \kappa(\boldsymbol{y}(\omega_q), \boldsymbol{y}(i))}{\partial y_{m_1}(\omega_q) y_{m_2}(i)}. \quad (85b)$$

APPENDIX C
CALCULATION OF $\boldsymbol{z}_m(i)$, $\boldsymbol{\zeta}_m(i)$, AND $\boldsymbol{\ell}_{m_1,m_2}(i)$

With the Gaussian kernel $\kappa(\boldsymbol{a}, \boldsymbol{b}) = \exp\left(-\|\boldsymbol{a} - \boldsymbol{b}\|^2 / 2\sigma^2\right)$, quantities (16)–(20) are given by:

$$[\boldsymbol{z}_m(i)]_q = \kappa(\boldsymbol{y}(i), \boldsymbol{y}(\omega_q)) \left(\frac{y_m(i) - y_m(\omega_q)}{\sigma^2} \right),$$
$$[\boldsymbol{\zeta}_m(i)]_q = -[\boldsymbol{z}_m(i)]_q$$

$$[\boldsymbol{\ell}_{m_1,m_2}(i)]_q =$$
$$\begin{cases} -\kappa(\boldsymbol{y}(i), \boldsymbol{y}(\omega_q)) \dfrac{y_{m_1}(i) - y_{m_1}(\omega_q)}{\sigma^2} \dfrac{y_{m_2}(i) - y_{m_2}(\omega_q)}{\sigma^2} & , m_1 \neq m_2 \\ -\kappa(\boldsymbol{y}(i), \boldsymbol{y}(\omega_q)) \left(\dfrac{\left(y_{m_1}(i) - y_{m_1}(\omega_q)\right)^2}{\sigma^4} - \dfrac{1}{\sigma^2} \right) & , m_1 = m_2 \end{cases}.$$

Supplementary material for the paper

Graph topology inference with derivative-reproducing property in RKHS: algorithm and convergence analysis

The goal of the supplementary material is to provide detailed computations of quantities $\boldsymbol{R}_{ss}$, $\boldsymbol{R}_{tt,m}$, $\mathbb{E}\{s_u(i)s_v(i)s_a(i)s_b(i)\}$, $\mathbb{E}\{s_u(i)s_v(i)s_a(i)y_n(i)\}$, $\mathbb{E}\{s_u(i)s_v(i)y_n^2(i)\}$ and $\mathbb{E}\{s_u(i)y_n(i)\}$.

We introduce vector $\widetilde{\boldsymbol{y}}(i) = [\boldsymbol{y}^\top(i),\, y_n^\top(i)]^\top$, whose k-th entry is denoted by $\widetilde{y}_k(i)$. The expectations involved in computing $\boldsymbol{R}_{tt,m}$ (with $\boldsymbol{R}_{ss}$ being one of its particular cases), $\mathbb{E}\{s_u(i)s_v(i)s_a(i)s_b(i)\}$, $\mathbb{E}\{s_u(i)s_v(i)s_a(i)y_n(i)\}$, $\mathbb{E}\{s_u(i)s_v(i)y_n^2(i)\}$ and $\mathbb{E}\{s_u(i)y_n(i)\}$ can be expressed generically as[1]:

$$\mathbb{E}\left\{ f(\widetilde{\boldsymbol{y}}) \triangleq (\widetilde{y}_{h_1} - x_{h_2})^{\iota_1}(\widetilde{y}_{h_3} - x_{h_4})^{\iota_2}(\widetilde{y}_{h_5} - x_{h_6})^{\iota_3}(\widetilde{y}_{h_7} - x_{h_8})^{\iota_4} \exp\left(-\frac{1}{2}c_4\widetilde{\boldsymbol{y}}^\top\boldsymbol{B}_0\widetilde{\boldsymbol{y}} - \boldsymbol{x}_4^\top\boldsymbol{B}_4^\top\widetilde{\boldsymbol{y}} + \frac{1}{2}\boldsymbol{x}_4^\top\boldsymbol{Q}_4\boldsymbol{x}_4\right)\right\}, \quad (1)$$

where $\boldsymbol{x}_4 = [\boldsymbol{y}^\top(\omega_p), \boldsymbol{y}^\top(\omega_q), \boldsymbol{y}^\top(\omega_r), \boldsymbol{y}^\top(\omega_s)]^\top$ represents the fixed dictionary elements. Also,

$$\boldsymbol{B}_0 = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & 0 \end{bmatrix} \quad (2)$$

$$c_4 = \frac{1 + \sum_{i=1}^3 c_i}{\sigma^2} \quad (3)$$

$$\boldsymbol{B}_4 = -\frac{1}{\sigma^2}\begin{bmatrix} \boldsymbol{I} & c_3\boldsymbol{I} & c_3c_2\boldsymbol{I} & c_3c_2c_1\boldsymbol{I} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \quad (4)$$

$$\boldsymbol{Q}_4 = -\frac{1}{\sigma^2}\begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & c_3\boldsymbol{I} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & c_3c_2\boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} & c_3c_2c_1\boldsymbol{I} \end{pmatrix} \quad (5)$$

In the previous quantities, c_1, c_2, c_3 represent binary selection variables, accounting for the following possible cases:

$$\begin{cases} c_1 = c_2 = c_3 = 1, \text{ for term } \mathbb{E}\{s_u(i)s_v(i)s_a(i)s_b(i)\} \\ c_1 = 0, c_2 = c_3 = 1, \text{ for term } \mathbb{E}\{s_u(i)s_v(i)s_a(i)y_n(i)\} \\ c_1 = c_2 = 0, c_3 = 1, \text{ for terms } \boldsymbol{R}_{tt,m} \text{ and } \mathbb{E}\{s_u(i)s_v(i)y_n^2(i)\} \\ c_1 = c_2 = c_3 = 0, \text{ for term } \mathbb{E}\{s_b(i)y_n(i)\} \end{cases}. \quad (6)$$

Note that $h_1, h_2, \ldots, h_8$ are different indexes of $\widetilde{\boldsymbol{y}}(i)$ and $\boldsymbol{x}_4$ (not necessarily distinct), and the binary variables $\iota_i \in \{0, 1\}$, for $i = 1, \ldots, 4$, allow us to accommodate lower-order cases and be more flexible.

We now have, for $\widetilde{\boldsymbol{y}}(i)$ jointly Gaussian distributed:

$$\begin{aligned} &\mathbb{E}\left\{ f(\widetilde{\boldsymbol{y}}) \triangleq (\widetilde{y}_{h_1} - x_{h_2})^{\iota_1}(\widetilde{y}_{h_3} - x_{h_4})^{\iota_2}(\widetilde{y}_{h_5} - x_{h_6})^{\iota_3}(\widetilde{y}_{h_7} - x_{h_8})^{\iota_4} \exp\left(-\frac{1}{2}c_4\widetilde{\boldsymbol{y}}^\top\boldsymbol{B}_0\widetilde{\boldsymbol{y}} - \boldsymbol{x}_4^\top\boldsymbol{B}_4^\top\widetilde{\boldsymbol{y}} + \frac{1}{2}\boldsymbol{x}_4^\top\boldsymbol{Q}_4\boldsymbol{x}_4\right)\right\} \\ &= (2\pi)^{-\frac{N+1}{2}} \det\{\boldsymbol{R}_{\widetilde{y}}\}^{-\frac{1}{2}} \exp\left(\frac{1}{2}\boldsymbol{x}_4^\top\boldsymbol{Q}_4\boldsymbol{x}_4\right) \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} (\widetilde{y}_{h_1} - x_{h_2})^{\iota_1}(\widetilde{y}_{h_3} - x_{h_4})^{\iota_2}(\widetilde{y}_{h_5} - x_{h_6})^{\iota_3}(\widetilde{y}_{h_7} - x_{h_8})^{\iota_4} \\ &\quad \times \exp\left(-\frac{1}{2}c_4\widetilde{\boldsymbol{y}}^\top\boldsymbol{B}_0\widetilde{\boldsymbol{y}}\right) \exp\left(-\frac{1}{2}\widetilde{\boldsymbol{y}}^\top\boldsymbol{R}_{\widetilde{y}}^{-1}\widetilde{\boldsymbol{y}}\right) \exp\left(-\boldsymbol{x}_4^\top\boldsymbol{B}_4^\top\widetilde{\boldsymbol{y}}\right) d\widetilde{y}_1 \cdots d\widetilde{y}_{N+1} \quad (7) \\ &= (2\pi)^{-\frac{N+1}{2}} \det\{\boldsymbol{R}_{\widetilde{y}}\}^{-\frac{1}{2}} \exp\left(\frac{1}{2}\boldsymbol{x}_4^\top\boldsymbol{Q}_4\boldsymbol{x}_4\right) \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} (\widetilde{y}_{h_1} - x_{h_2})^{\iota_1}(\widetilde{y}_{h_3} - x_{h_4})^{\iota_2}(\widetilde{y}_{h_5} - x_{h_6})^{\iota_3}(\widetilde{y}_{h_7} - x_{h_8})^{\iota_4} \\ &\quad \times \exp\left(-\frac{1}{2}\widetilde{\boldsymbol{y}}^\top\left[c_4\boldsymbol{B}_0 + \boldsymbol{R}_{\widetilde{y}}^{-1}\right]\widetilde{\boldsymbol{y}} - \boldsymbol{x}_4^\top\boldsymbol{B}_4^\top\widetilde{\boldsymbol{y}}\right) d\widetilde{y}_1 \cdots d\widetilde{y}_{N+1}\,. \quad (8) \end{aligned}$$

[1]Note that the i index was omitted in (1) to simplify the notation.

Note that we can write:

$$
\begin{aligned}
&-\frac{1}{2}\widetilde{\boldsymbol{y}}^\top\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]\widetilde{\boldsymbol{y}}-\boldsymbol{x}_4^\top\boldsymbol{B}_4^\top\widetilde{\boldsymbol{y}}\\
&=-\frac{1}{2}\left(\widetilde{\boldsymbol{y}}^\top\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]\widetilde{\boldsymbol{y}}+\boldsymbol{x}_4^\top\boldsymbol{B}_4^\top\widetilde{\boldsymbol{y}}+\widetilde{\boldsymbol{y}}^\top\boldsymbol{B}_4\boldsymbol{x}_4\right)\\
&=-\frac{1}{2}\left(\widetilde{\boldsymbol{y}}+\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]^{-1}\boldsymbol{B}_4\boldsymbol{x}_4\right)^\top\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]\left(\widetilde{\boldsymbol{y}}+\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]^{-1}\boldsymbol{B}_4\boldsymbol{x}_4\right)\\
&\quad+\frac{1}{2}\boldsymbol{x}_4^\top\boldsymbol{B}_4^\top\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]^{-1}\boldsymbol{B}_4\boldsymbol{x}_4\,. \qquad (9)
\end{aligned}
$$

Thus, relation (8) becomes:

$$
\begin{aligned}
&=(2\pi)^{-\frac{N+1}{2}}\det\{\boldsymbol{R}_{\widetilde{y}}\}^{-\frac{1}{2}}\exp\left(\frac{1}{2}\boldsymbol{x}_4^\top\boldsymbol{Q}_4\boldsymbol{x}_4\right)\int_{\mathbb{R}}\cdots\int_{\mathbb{R}}(\widetilde{y}_{h_1}-x_{h_2})^{\iota_1}(\widetilde{y}_{h_3}-x_{h_4})^{\iota_2}(\widetilde{y}_{h_5}-x_{h_6})^{\iota_3}(\widetilde{y}_{h_7}-x_{h_8})^{\iota_4}\\
&\qquad\times\exp\left(-\frac{1}{2}\left(\widetilde{\boldsymbol{y}}+\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]^{-1}\boldsymbol{B}_4\boldsymbol{x}_4\right)^\top\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]\left(\widetilde{\boldsymbol{y}}+\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]^{-1}\boldsymbol{B}_4\boldsymbol{x}_4\right)\right)\\
&\qquad\times\exp\left(\frac{1}{2}\boldsymbol{x}_4^\top\boldsymbol{B}_4^\top\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]^{-1}\boldsymbol{B}_4\boldsymbol{x}_4\right)d\widetilde{y}_1\cdots d\widetilde{y}_{N+1}\\
&=(2\pi)^{-\frac{N+1}{2}}\det\{\boldsymbol{R}_{\widetilde{y}}\}^{-\frac{1}{2}}\exp\left(\frac{1}{2}\boldsymbol{x}_4^\top\boldsymbol{Q}_4\boldsymbol{x}_4\right)\det\left\{\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]^{-1}\right\}^{\frac{1}{2}}\det\left\{\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]^{-1}\right\}^{-\frac{1}{2}}\\
&\qquad\times\exp\left(\frac{1}{2}\boldsymbol{x}_4^\top\boldsymbol{B}_4^\top\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]^{-1}\boldsymbol{B}_4\boldsymbol{x}_4\right)\int_{\mathbb{R}}\cdots\int_{\mathbb{R}}(\widetilde{y}_{h_1}-x_{h_2})^{\iota_1}(\widetilde{y}_{h_3}-x_{h_4})^{\iota_2}(\widetilde{y}_{h_5}-x_{h_6})^{\iota_3}(\widetilde{y}_{h_7}-x_{h_8})^{\iota_4}\\
&\qquad\times\exp\left(-\frac{1}{2}\left(\widetilde{\boldsymbol{y}}+\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]^{-1}\boldsymbol{B}_4\boldsymbol{x}_4\right)^\top\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]\left(\widetilde{\boldsymbol{y}}+\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]^{-1}\boldsymbol{B}_4\boldsymbol{x}_4\right)\right)d\widetilde{y}_1\cdots d\widetilde{y}_{N+1}\\
&=\det\{\boldsymbol{R}_{\widetilde{y}}\}^{-\frac{1}{2}}\det\left\{\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]^{-1}\right\}^{\frac{1}{2}}\exp\left(\frac{1}{2}\boldsymbol{x}_4^\top\left(\boldsymbol{Q}_4+\boldsymbol{B}_4^\top\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]^{-1}\boldsymbol{B}_4\right)\boldsymbol{x}_4\right)\\
&\qquad\times\mathbb{E}_{p'(\widetilde{\boldsymbol{y}})}\left\{(\widetilde{y}_{h_1}-x_{h_2})^{\iota_1}(\widetilde{y}_{h_3}-x_{h_4})^{\iota_2}(\widetilde{y}_{h_5}-x_{h_6})^{\iota_3}(\widetilde{y}_{h_7}-x_{h_8})^{\iota_4}\right\} \qquad (10)\\
&=\nu(\{h_i\}_{i=1}^8)\,, \qquad (11)
\end{aligned}
$$

with the normal distribution:

$$
p'(\widetilde{\boldsymbol{y}})=\mathcal{N}\left(\boldsymbol{\mu}=-\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]^{-1}\boldsymbol{B}_4\boldsymbol{x}_4,\,\boldsymbol{\Sigma}=\left[c_4\boldsymbol{B}_0+\boldsymbol{R}_{\widetilde{y}}^{-1}\right]^{-1}\right)\,. \qquad (12)
$$

Let us focus on the expectation in (10). We have, for $\iota_i=1, i=\{1,2,3,4\}$:

$$
\begin{aligned}
&\mathbb{E}_{p'(\boldsymbol{y})}\left\{(\widetilde{y}_{h_1}-x_{h_2})^{\iota_1}(\widetilde{y}_{h_3}-x_{h_4})^{\iota_2}(\widetilde{y}_{h_5}-x_{h_6})^{\iota_3}(\widetilde{y}_{h_7}-x_{h_8})^{\iota_4}\right\}\\
&=\mathbb{E}\{\widetilde{y}_{h_1}\widetilde{y}_{h_3}\widetilde{y}_{h_5}\widetilde{y}_{h_7}\}-x_{h_8}\mathbb{E}\{\widetilde{y}_{h_1}\widetilde{y}_{h_3}\widetilde{y}_{h_5}\}-x_{h_6}\mathbb{E}\{\widetilde{y}_{h_1}\widetilde{y}_{h_3}\widetilde{y}_{h_7}\}+x_{h_6}x_{h_8}\mathbb{E}\{\widetilde{y}_{h_1}\widetilde{y}_{h_3}\}\\
&\quad-x_{h_4}\mathbb{E}\{\widetilde{y}_{h_1}\widetilde{y}_{h_5}\widetilde{y}_{h_7}\}+x_{h_4}x_{h_8}\mathbb{E}\{\widetilde{y}_{h_1}\widetilde{y}_{h_5}\}+x_{h_4}x_{h_6}\mathbb{E}\{\widetilde{y}_{h_1}\widetilde{y}_{h_7}\}-x_{h_4}x_{h_6}x_{h_8}\mathbb{E}\{\widetilde{y}_{h_1}\}\\
&\quad-x_{h_2}\mathbb{E}\{\widetilde{y}_{h_3}\widetilde{y}_{h_5}\widetilde{y}_{h_7}\}+x_{h_2}x_{h_8}\mathbb{E}\{\widetilde{y}_{h_3}\widetilde{y}_{h_5}\}+x_{h_2}x_{h_6}\mathbb{E}\{\widetilde{y}_{h_3}\widetilde{y}_{h_7}\}-x_{h_2}x_{h_6}x_{h_8}\mathbb{E}\{\widetilde{y}_{h_3}\}\\
&\quad+x_{h_2}x_{h_4}\mathbb{E}\{\widetilde{y}_{h_5}\widetilde{y}_{h_7}\}-x_{h_2}x_{h_4}x_{h_8}\mathbb{E}\{\widetilde{y}_{h_5}\}-x_{h_2}x_{h_4}x_{h_6}\mathbb{E}\{\widetilde{y}_{h_7}\}+x_{h_2}x_{h_4}x_{h_6}x_{h_8}\,. \qquad (13)
\end{aligned}
$$

Since $h_i, i\in\{1,3,5,7\}$ represent solely generic placeholders for any actual index, it suffices to compute only a few expectations from the previous relation, which can then be used for any other combination. We have:

$$
\begin{aligned}
\mathbb{E}_{p'(\widetilde{\boldsymbol{y}})}\left\{\widetilde{y}_{h_1}\widetilde{y}_{h_3}\widetilde{y}_{h_5}\widetilde{y}_{h_7}\right\}&=\mu_{h_1}\mu_{h_3}\mu_{h_5}\mu_{h_7}+\Sigma_{h_1,h_3}\Sigma_{h_5,h_7}+\Sigma_{h_1,h_5}\Sigma_{h_3,h_7}+\Sigma_{h_1,h_7}\Sigma_{h_3,h_5}\\
&\quad+\mu_{h_1}\mu_{h_3}\Sigma_{h_5,h_7}+\mu_{h_1}\mu_{h_5}\Sigma_{h_3,h_7}+\mu_{h_1}\mu_{h_7}\Sigma_{h_3,h_5}\\
&\quad+\mu_{h_3}\mu_{h_5}\Sigma_{h_1,h_7}+\mu_{h_3}\mu_{h_7}\Sigma_{h_1,h_5}+\mu_{h_5}\mu_{h_7}\Sigma_{h_1,h_3}\,, \qquad (14)\\
\mathbb{E}_{p'(\boldsymbol{y})}\left\{\widetilde{y}_{h_1}\widetilde{y}_{h_3}\widetilde{y}_{h_5}\right\}&=\mu_{h_1}\mu_{h_3}\mu_{h_5}+\mu_{h_5}\Sigma_{h_1,h_3}+\mu_{h_3}\Sigma_{h_1,h_5}+\mu_{h_1}\Sigma_{h_3,h_5}\,, \qquad (15)\\
\mathbb{E}_{p'(\boldsymbol{y})}\left\{\widetilde{y}_{h_1}\widetilde{y}_{h_3}\right\}&=\mu_{h_1}\mu_{h_3}+\Sigma_{h_1,h_3}\,, \qquad (16)\\
\mathbb{E}_{p'(\boldsymbol{y})}\left\{\widetilde{y}_{h_1}\right\}&=\mu_{h_1}\,. \qquad (17)
\end{aligned}
$$

APPENDIX D
CASES CORRESPONDING TO $\boldsymbol{R}_{tt,m}$ ($c_1 = c_2 = 0, c_3 = 1$)

Since $\boldsymbol{t}_m(p) = \begin{bmatrix}\boldsymbol{\ell}_m(p)\\ \boldsymbol{\zeta}_m(p)\end{bmatrix}$, the matrix $\boldsymbol{R}_{tt,m}$ can be written as:

$$\begin{aligned}\boldsymbol{R}_{tt,m} &= \mathbb{E}\{\boldsymbol{t}_m(i)\boldsymbol{t}_m^\top(i)\}\\ &= \mathbb{E}\left\{\begin{bmatrix}\boldsymbol{\ell}_m(i)\\ \boldsymbol{\zeta}_m(i)\end{bmatrix}\begin{bmatrix}\boldsymbol{\ell}_m^\top(i) & \boldsymbol{\zeta}_m^\top(i)\end{bmatrix}\right\}\\ &= \begin{bmatrix}\boldsymbol{R}_{\ell\ell,m} & \boldsymbol{R}_{\zeta\ell,m}^\top\\ \boldsymbol{R}_{\zeta\ell,m} & \boldsymbol{R}_{\zeta\zeta,m}\end{bmatrix},\end{aligned} \tag{18}$$

where in the last step we used the fact that the signals are identically distributed (for different i).

For ease of comprehension, we recall the definitions of $\boldsymbol{z}(i)$, $\boldsymbol{\zeta}(i)$ and $\boldsymbol{\ell}_m(i)$:

$$\begin{aligned}\boldsymbol{z}(i) &= \left[\boldsymbol{z}_1^\top(i), \ldots, \boldsymbol{z}_N^\top(i)\right]^\top, \qquad [\boldsymbol{z}_m(i)]_q = \left.\frac{\partial\kappa(\boldsymbol{y}(i), \boldsymbol{y}(\omega_q))}{\partial y_m(\omega_q)}\right|_{q=1,\ldots,|\mathcal{D}_n|},\\ \boldsymbol{\zeta}(i) &= \left[\boldsymbol{\zeta}_1^\top(i), \ldots, \boldsymbol{\zeta}_N^\top(i)\right]^\top, \qquad [\boldsymbol{\zeta}_m(i)]_q = \left.\frac{\partial\kappa(\boldsymbol{y}(i), \boldsymbol{y}(\omega_q))}{\partial y_m(i)}\right|_{q=1,\ldots,|\mathcal{D}_n|},\\ \boldsymbol{\ell}_m(i) &= \left[\boldsymbol{\ell}_{1,m}^\top(i), \ldots, \boldsymbol{\ell}_{N,m}^\top(i)\right]^\top, \qquad [\boldsymbol{\ell}_{m_1,m_2}(i)]_q = \left.\frac{\partial^2\kappa(\boldsymbol{y}(i), \boldsymbol{y}(\omega_q))}{\partial y_{m_1}(\omega_q)\partial y_{m_2}(\omega_q)}\right|_{q=1,\ldots,|\mathcal{D}_n|}.\end{aligned}$$

Their explicit forms for the particular case of the Gaussian kernel are:

$$\begin{aligned}[\boldsymbol{z}_m(i)]_q &= -[\boldsymbol{\zeta}_m(i)]_q\\ &= \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y}(i) - \boldsymbol{y}(\omega_q)\|^2\right)\left(\frac{[\boldsymbol{y}(i)]_m - [\boldsymbol{y}(\omega_q)]_m}{\sigma^2}\right),\\ [\boldsymbol{\ell}_{m_1,m_2}]_q &= \begin{cases}-\exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y}(i) - \boldsymbol{y}(\omega_q)\|^2\right)\frac{[\boldsymbol{y}(i)]_{m_1} - [\boldsymbol{y}(\omega_q)]_{m_1}}{\sigma^2}\frac{[\boldsymbol{y}(i)]_{m_2} - [\boldsymbol{y}(\omega_q)]_{m_2}}{\sigma^2} & m_1 \neq m_2\\ -\exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y}(i) - \boldsymbol{y}(\omega_q)\|^2\right)\left(\frac{\left([\boldsymbol{y}(i)]_{m_1} - [\boldsymbol{y}(\omega_q)]_{m_1}\right)^2}{\sigma^4} - \frac{1}{\sigma^2}\right) & m_1 = m_2\end{cases}.\end{aligned}$$

A. Computing $\boldsymbol{R}_{\ell\ell,m}$

$$
\begin{aligned}
&\big[\boldsymbol{R}_{\ell\ell,m}\big]_{(a-1)|\mathcal{D}_n|+p,\,(b-1)|\mathcal{D}_n|+q} \\
&\quad= \big[\mathbb{E}\{\boldsymbol{\ell}_m(i)\boldsymbol{\ell}_m^{\top}(i)\}\big]_{(a-1)|\mathcal{D}_n|+p,\,(b-1)|\mathcal{D}_n|+q} \\
&\quad= \mathbb{E}\big\{[\boldsymbol{\ell}_{a,m}]_p[\boldsymbol{\ell}_{b,m}]_q\big\} \\
&\quad= \begin{cases}
\mathbb{E}\Big\{ \dfrac{(y_a(i)-y_a(\omega_p))}{\sigma^2}\dfrac{(y_m(i)-y_m(\omega_p))}{\sigma^2}\dfrac{(y_b(i)-y_b(\omega_q))}{\sigma^2}\dfrac{(y_m(i)-y_m(\omega_q))}{\sigma^2} \\
\qquad \times \exp\left(-\dfrac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2 - \dfrac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2\right)\Big\}, & a\neq m,\, b\neq m \\
\mathbb{E}\Big\{ \dfrac{(y_a(i)-y_a(\omega_p))}{\sigma^2}\dfrac{(y_m(i)-y_m(\omega_p))}{\sigma^2}\left(\dfrac{\big(y_b(i)-y_b(\omega_q)\big)^2}{\sigma^4}-\dfrac{1}{\sigma^2}\right) \\
\qquad \times \exp\left(-\dfrac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2 - \dfrac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2\right)\Big\}, & a\neq m,\, b=m \\
\mathbb{E}\Big\{ \dfrac{(y_b(i)-y_b(\omega_q))}{\sigma^2}\dfrac{(y_m(i)-y_m(\omega_q))}{\sigma^2}\left(\dfrac{\big(y_a(i)-y_a(\omega_p)\big)^2}{\sigma^4}-\dfrac{1}{\sigma^2}\right) \\
\qquad \times \exp\left(-\dfrac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2 - \dfrac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2\right)\Big\}, & a=m,\, b\neq m \\
\mathbb{E}\Big\{ \left(\dfrac{\big(y_a(i)-y_a(\omega_p)\big)^2}{\sigma^4}-\dfrac{1}{\sigma^2}\right)\left(\dfrac{\big(y_b(i)-y_b(\omega_q)\big)^2}{\sigma^4}-\dfrac{1}{\sigma^2}\right) \\
\quad \times \exp\left(-\dfrac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2 - \dfrac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2\right)\Big\}, & a=m,\, b=m
\end{cases}
\end{aligned}
\tag{19}
$$

which is equivalent to:

$$
\begin{aligned}
&\left[\boldsymbol{R}_{\ell\ell,m}\right]_{(a-1)|\mathcal{D}_n|+p,\,(b-1)|\mathcal{D}_n|+q} \\
&= \begin{cases}
\dfrac{1}{\sigma^8}\nu(\{h_i\}_{i=1}^8), \text{ with } \begin{cases} h_1 = a \\ h_2 = a \\ h_3 = m \\ h_4 = m \\ h_5 = b \\ h_6 = N+b \\ h_7 = m \\ h_8 = N+m \end{cases}, & \text{when } a \neq m,\, b \neq m \\
\dfrac{1}{\sigma^8}\nu(\{h_i\}_{i=1}^8) \\ -\dfrac{1}{\sigma^2}\left[\boldsymbol{R}_{zz}\right]_{(a-1)|\mathcal{D}_n|+p,(m-1)|\mathcal{D}_n|+p}, \text{ with } \begin{cases} h_1 = a \\ h_2 = a \\ h_3 = b \\ h_4 = b \\ h_5 = b \\ h_6 = N+b \\ h_7 = b \\ h_8 = N+b \end{cases}, & \text{when } a \neq m,\, b = m \\
\dfrac{1}{\sigma^8}\nu(\{h_i\}_{i=1}^8) \\ -\dfrac{1}{\sigma^2}\left[\boldsymbol{R}_{zz}\right]_{(b-1)|\mathcal{D}_n|+q,(m-1)|\mathcal{D}_n|+q}, \text{ with } \begin{cases} h_1 = b \\ h_2 = b \\ h_3 = a \\ h_4 = a \\ h_5 = a \\ h_6 = N+a \\ h_7 = a \\ h_8 = N+a \end{cases}, & \text{when } a = m,\, b \neq m \\
\dfrac{1}{\sigma^8}\nu(\{h_i\}_{i=1}^8) \\ -\dfrac{1}{\sigma^2}\left[\boldsymbol{R}_{zz}\right]_{(a-1)|\mathcal{D}_n|+p,(a-1)|\mathcal{D}_n|+p} \\ -\dfrac{1}{\sigma^2}\left[\boldsymbol{R}_{zz}\right]_{(b-1)|\mathcal{D}_n|+q,(b-1)|\mathcal{D}_n|+q} \\ +\dfrac{1}{\sigma^4}\left[\boldsymbol{R}_{kk}\right]_{p,q}, \text{ with } \begin{cases} h_1 = a \\ h_2 = a \\ h_3 = a \\ h_4 = a \\ h_5 = b \\ h_6 = N+b \\ h_7 = b \\ h_8 = N+b \end{cases}, & \text{when } a = m,\, b = m
\end{cases}.
\end{aligned}
\tag{20}
$$

B. Computing $\boldsymbol{R}_{\zeta\ell,m}$

$$
\begin{aligned}
&\left[\boldsymbol{R}_{\zeta\ell,m}\right]_{p,\,(b-1)|\mathcal{D}_n|+q} \\
&\quad= \left[\mathbb{E}\{\boldsymbol{\zeta}_m(i)\boldsymbol{\ell}_m^\top(i)\}\right]_{p,\,(b-1)|\mathcal{D}_n|+q} \\
&\quad= \mathbb{E}\left\{[\boldsymbol{\zeta}_m]_p[\boldsymbol{\ell}_{b,m}]_q\right\} \\
&\quad= \begin{cases}
-\mathbb{E}\Bigg\{\dfrac{(y_m(i)-y_m(\omega_p))}{\sigma^2}\dfrac{(y_b(i)-y_b(\omega_q))}{\sigma^2}\dfrac{(y_m(i)-y_m(\omega_q))}{\sigma^2} & \\
\times\exp\left(-\dfrac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2-\dfrac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2\right)\Bigg\} & , \quad b\neq m \\
-\mathbb{E}\Bigg\{\dfrac{(y_b(i)-y_b(\omega_p))}{\sigma^2}\left(\dfrac{\left(y_b(i)-y_b(\omega_q)\right)^2}{\sigma^4}-\dfrac{1}{\sigma^2}\right) & \\
\times\exp\left(-\dfrac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2-\dfrac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2\right)\Bigg\} & , \quad b=m
\end{cases}.
\end{aligned} \tag{21}
$$

which becomes:

$$
\left[\boldsymbol{R}_{\zeta\ell,m}\right]_{p,\,(b-1)|\mathcal{D}_n|+q} = \begin{cases}
-\dfrac{1}{\sigma^6}\nu(\{h_i\}_{i=1}^8),\ \text{with}\ \begin{cases}\left.\begin{matrix}h_1=\bullet\\h_2=\bullet\end{matrix}\right\}\iota_1=0\\ h_3=m\\ h_4=m\\ h_5=b\\ h_6=N+b\\ h_7=m\\ h_8=N+m\end{cases}, & \text{when } b\neq m \\
-\dfrac{1}{\sigma^6}\nu(\{h_i\}_{i=1}^8) & \\
+\dfrac{1}{\sigma^2}\left[\boldsymbol{R}_{kz}\right]_{q,(b-1)|\mathcal{D}_n|+p},\ \text{with}\ \begin{cases}\left.\begin{matrix}h_1=\bullet\\h_2=\bullet\end{matrix}\right\}\iota_1=0\\ h_3=b\\ h_4=b\\ h_5=b\\ h_6=N+b\\ h_7=b\\ h_8=N+b\end{cases}, & \text{when } b=m
\end{cases}, \tag{22}
$$

where the bullet $\bullet$ means that the index is irrelevant, due to its corresponding term not being present in the generic form (1).

C. Computing $\boldsymbol{R}_{ss}$ (including $\boldsymbol{R}_{zz,m}$, $\boldsymbol{R}_{zz,m_1,m_2}$, $\boldsymbol{R}_{kz,m}$ and $\boldsymbol{R}_{kk,m}$)

Since $\boldsymbol{s}(i)=\begin{bmatrix}\boldsymbol{z}(i)\\ \boldsymbol{k}(i)\end{bmatrix}$, all the matrices detailed in this section can be used in computing $\boldsymbol{R}_{ss}$.

$$
\begin{aligned}
\left[\boldsymbol{R}_{\zeta\zeta,m}\right]_{p,\,q} &= \left[\mathbb{E}\{\boldsymbol{\zeta}_m(i)\boldsymbol{\zeta}_m^\top(i)\}\right]_{p,\,q} \\
&= \mathbb{E}\left\{[\boldsymbol{\zeta}_m]_p[\boldsymbol{\zeta}_m]_q\right\} \\
&= \mathbb{E}\left\{\frac{(y_m(i)-y_m(\omega_p))}{\sigma^2}\frac{(y_m(i)-y_m(\omega_q))}{\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2-\frac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2\right)\right\}. \qquad (23)
\end{aligned}
$$

$$
= \frac{1}{\sigma^4}\nu(\{h_i\}_{i=1}^8),\ \text{with}\ \begin{cases}\left.\begin{matrix}h_1=\bullet\\h_2=\bullet\end{matrix}\right\}\iota_1=0\\ h_3=m\\ h_4=m\\ \left.\begin{matrix}h_5=\bullet\\h_6=\bullet\end{matrix}\right\}\iota_3=0\\ h_7=m\\ h_8=N+m\end{cases}. \tag{24}
$$

Generalizing to the case $m_1 \neq m_2$, we straightforwardly have:

$$\begin{aligned}\left[\boldsymbol{R}_{\zeta\zeta,m_1,m_2}\right]_{p,\,q} &= \left[\mathbb{E}\{\boldsymbol{\zeta}_{m_1}(i)\boldsymbol{\zeta}_{m_2}^{\top}(i)\}\right]_{p,\,q} \\ &= \mathbb{E}\big\{[\boldsymbol{\zeta}_{m_1}]_p[\boldsymbol{\zeta}_{m_2}]_q\big\} \\ &= \mathbb{E}\left\{\frac{(y_{m_1}(i)-y_{m_1}(\omega_p))}{\sigma^2}\frac{(y_{m_2}(i)-y_{m_2}(\omega_q))}{\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2-\frac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2\right)\right\}. \qquad (25)\end{aligned}$$

$$= \frac{1}{\sigma^4}\nu(\{h_i\}_{i=1}^8), \text{ with } \begin{cases} \left.\begin{matrix}h_1=\bullet\\ h_2=\bullet\end{matrix}\right\}\iota_1=0 \\ h_3=m_1 \\ h_4=m_1 \\ \left.\begin{matrix}h_5=\bullet\\ h_6=\bullet\end{matrix}\right\}\iota_3=0 \\ h_7=m_2 \\ h_8=N+m_2 \end{cases}. \qquad (26)$$

Particularizing to $\boldsymbol{R}_{kz,m}$ and using the fact that for the Gaussian kernel $[\boldsymbol{z}_m]_q = -[\boldsymbol{\zeta}_m]_q$, we have:

$$\begin{aligned}\left[\boldsymbol{R}_{kz,m}\right]_{p,\,q} &= -\left[\mathbb{E}\{\boldsymbol{k}(i)\boldsymbol{\zeta}_m^{\top}(i)\}\right]_{p,\,q} \\ &= \mathbb{E}\big\{[\boldsymbol{k}]_p[\boldsymbol{z}_m]_q\big\} \\ &= \mathbb{E}\left\{\frac{(y_{m_2}(i)-y_{m_2}(\omega_q))}{\sigma^2}\exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2-\frac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2\right)\right\}. \qquad (27)\end{aligned}$$

$$= \frac{1}{\sigma^2}\nu(\{h_i\}_{i=1}^8), \text{ with } \begin{cases} \left.\begin{matrix}h_1=\bullet\\ h_2=\bullet\end{matrix}\right\}\iota_1=0 \\ \left.\begin{matrix}h_3=\bullet\\ h_4=\bullet\end{matrix}\right\}\iota_2=0 \\ \left.\begin{matrix}h_5=\bullet\\ h_6=\bullet\end{matrix}\right\}\iota_3=0 \\ h_7=m \\ h_8=N+m \end{cases}. \qquad (28)$$

Particularizing to $\boldsymbol{R}_{kk}$, we have:

$$\begin{aligned}\left[\boldsymbol{R}_{kk}\right]_{p,\,q} &= \left[\mathbb{E}\{\boldsymbol{k}(i)\boldsymbol{k}^{\top}(i)\}\right]_{p,\,q} \\ &= \mathbb{E}\big\{[\boldsymbol{k}]_p[\boldsymbol{k}]_q\big\} \\ &= \mathbb{E}\left\{\exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2-\frac{1}{2\sigma^2}\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2\right)\right\}. \qquad (29)\end{aligned}$$

$$= \nu(\{h_i\}_{i=1}^8), \text{ with } \begin{cases} \left.\begin{matrix}h_1=\bullet\\ h_2=\bullet\end{matrix}\right\}\iota_1=0 \\ \left.\begin{matrix}h_3=\bullet\\ h_4=\bullet\end{matrix}\right\}\iota_2=0 \\ \left.\begin{matrix}h_5=\bullet\\ h_6=\bullet\end{matrix}\right\}\iota_3=0 \\ \left.\begin{matrix}h_7=\bullet\\ h_8=\bullet\end{matrix}\right\}\iota_4=0 \end{cases}. \qquad (30)$$

APPENDIX E
CASES CORRESPONDING TO $\mathbb{E}\{s_u(i)s_v(i)s_a(i)s_b(i)\}$ ($c_1 = c_2 = c_3 = 1$)

Recall that $\boldsymbol{s}(i) = \begin{bmatrix}\boldsymbol{z}(i)\\ \boldsymbol{k}(i)\end{bmatrix}$. We have the following terms:

a) Term $\mathbb{E}\{z_u(i)z_a(i)z_b(i)z_v(i)\}$*:* $u,v=1,\ldots,N|\mathcal{D}_n|$, $a,b=1,\ldots,N|\mathcal{D}_n|$

$$\begin{aligned}
&\mathbb{E}\{z_u(i)z_a(i)z_a(i)z_v(i)\}\\
&\quad=\mathbb{E}\Bigg\{\frac{(y_{m_1}(i)-y_{m_1}(\omega_p))}{\sigma^2}\frac{(y_{m_2}(i)-y_{m_2}(\omega_q))}{\sigma^2}\frac{(y_{m_3}(i)-y_{m_3}(\omega_r))}{\sigma^2}\frac{(y_{m_4}(i)-y_{m_4}(\omega_s))}{\sigma^2}\\
&\qquad\times\exp\left(-\frac{1}{2\sigma^2}\left(\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_r)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_s)\|^2\right)\right)\Bigg\} \qquad (31)\\
&\quad=\frac{1}{\sigma^8}\nu(\{h_i\}_{i=1}^8),\text{ with }\begin{cases}h_1=h_2=m_1=\left\lceil\frac{u}{|\mathcal{D}_n|}\right\rceil\\ h_3=h_4-N=m_2=\left\lceil\frac{a}{|\mathcal{D}_n|}\right\rceil\\ h_5=h_6-2N=m_3=\left\lceil\frac{b}{|\mathcal{D}_n|}\right\rceil\\ h_7=h_8-3N=m_4=\left\lceil\frac{v}{|\mathcal{D}_n|}\right\rceil\end{cases},\begin{cases}\omega_p=\mathrm{mod}(u-1,|\mathcal{D}_n|)+1\\ \omega_q=\mathrm{mod}(a-1,|\mathcal{D}_n|)+1\\ \omega_r=\mathrm{mod}(b-1,|\mathcal{D}_n|)+1\\ \omega_s=\mathrm{mod}(v-1,|\mathcal{D}_n|)+1\end{cases}. \qquad (32)
\end{aligned}$$

b) Term $\mathbb{E}\{k_u(i)z_a(i)z_b(i)z_v(i)\}$*:* $u=1,\ldots,|\mathcal{D}_n|$, $v=1,\ldots,N|\mathcal{D}_n|$, $a,b=1,\ldots,N|\mathcal{D}_n|$

$$\begin{aligned}
\mathbb{E}\{k_u(i)z_a(i)z_b(i)z_v(i)\}&=\mathbb{E}\Bigg\{\frac{(y_{m_2}(i)-y_{m_2}(\omega_q))}{\sigma^2}\frac{(y_{m_3}(i)-y_{m_3}(\omega_r))}{\sigma^2}\frac{(y_{m_4}(i)-y_{m_4}(\omega_s))}{\sigma^2}\exp\Bigg(-\frac{1}{2\sigma^2}\Big(\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2\\
&\quad+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_r)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_s)\|^2\Big)\Bigg)\Bigg\} \qquad (33)\\
&=\frac{1}{\sigma^6}\nu(\{h_i\}_{i=1}^8),\text{ with }\begin{cases}h_1=h_2=m_1=\bullet\\ h_3=h_4-N=m_2=\left\lceil\frac{a}{|\mathcal{D}_n|}\right\rceil\\ h_5=h_6-2N=m_3=\left\lceil\frac{b}{|\mathcal{D}_n|}\right\rceil\\ h_7=h_8-3N=m_4=\left\lceil\frac{v}{|\mathcal{D}_n|}\right\rceil\end{cases},\begin{cases}\omega_p=u\\ \omega_q=\mathrm{mod}(a-1,|\mathcal{D}_n|)+1\\ \omega_r=\mathrm{mod}(b-1,|\mathcal{D}_n|)+1\\ \omega_s=\mathrm{mod}(v-1,|\mathcal{D}_n|)+1\end{cases}. \qquad (34)
\end{aligned}$$

c) Term $\mathbb{E}\{k_u(i)k_a(i)z_b(i)z_v(i)\}$*:* $u=1,\ldots,|\mathcal{D}_n|$, $v=1,\ldots,N|\mathcal{D}_n|$, $a=1,\ldots,|\mathcal{D}_n|$, $b=1,\ldots,N|\mathcal{D}_n|$

$$\begin{aligned}
\mathbb{E}\{k_u(i)k_a(i)z_b(i)z_v(i)\}&=\mathbb{E}\Bigg\{\frac{(y_{m_3}(i)-y_{m_3}(\omega_r))}{\sigma^2}\frac{(y_{m_4}(i)-y_{m_4}(\omega_s))}{\sigma^2}\exp\Bigg(-\frac{1}{2\sigma^2}\Big(\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2\\
&\quad+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_r)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_s)\|^2\Big)\Bigg)\Bigg\} \qquad (35)\\
&=\frac{1}{\sigma^4}\nu(\{h_i\}_{i=1}^8),\text{ with }\begin{cases}h_1=h_2=m_1=\bullet\\ h_3=h_4-N=m_2=\bullet\\ h_5=h_6-2N=m_3=\left\lceil\frac{b}{|\mathcal{D}_n|}\right\rceil\\ h_7=h_8-3N=m_4=\left\lceil\frac{v}{|\mathcal{D}_n|}\right\rceil\end{cases},\begin{cases}\omega_p=u\\ \omega_q=a\\ \omega_r=\mathrm{mod}(b-1,|\mathcal{D}_n|)+1\\ \omega_s=\mathrm{mod}(v-1,|\mathcal{D}_n|)+1\end{cases}. \qquad (36)
\end{aligned}$$

d) Term $\mathbb{E}\{k_u(i)k_a(i)k_b(i)z_v(i)\}$*:* $u=1,\ldots,|\mathcal{D}_n|$, $v=1,\ldots,N|\mathcal{D}_n|$, $a,b=1,\ldots,|\mathcal{D}_n|$

$$\begin{aligned}
\mathbb{E}\{k_u(i)k_a(i)k_b(i)z_v(i)\}&=\mathbb{E}\Bigg\{\frac{(y_{m_4}(i)-y_{m_4}(\omega_s))}{\sigma^2}\exp\Bigg(-\frac{1}{2\sigma^2}\Big(\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2\\
&\quad+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_r)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_s)\|^2\Big)\Bigg)\Bigg\} \qquad (37)\\
&=\frac{1}{\sigma^2}\nu(\{h_i\}_{i=1}^8),\text{ with }\begin{cases}h_1=h_2=m_1=\bullet\\ h_3=h_4-N=m_2=\bullet\\ h_5=h_6-2N=m_3=\bullet\\ h_7=h_8-3N=m_4=\left\lceil\frac{v}{|\mathcal{D}_n|}\right\rceil\end{cases},\begin{cases}\omega_p=u\\ \omega_q=a\\ \omega_r=b\\ \omega_s=\mathrm{mod}(v-1,|\mathcal{D}_n|)+1\end{cases}. \qquad (38)
\end{aligned}$$

e) Term $\mathbb{E}\{k_u(i)k_a(i)k_a(i)k_v(i)\}$*:* $u,v=1,\dots,|\mathcal{D}_n|$, $a,b=1,\dots,|\mathcal{D}_n|$

$$\begin{aligned}&\mathbb{E}\{k_u(i)k_a(i)k_b(i)k_v(i)\}\\&\quad=\exp\left(-\frac{1}{2\sigma^2}\left(\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_r)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_s)\|^2\right)\right)\Bigg\} \qquad (39)\\&\quad=\nu(\{h_i\}_{i=1}^8),\text{ with }\begin{cases}h_1=h_2=m_1=\bullet\\h_3=h_4-N=m_2=\bullet\\h_5=h_6-2N=m_3=\bullet\\h_7=h_8-3N=m_4=\bullet\end{cases},\begin{cases}\omega_p=u\\\omega_q=a\\\omega_r=b\\\omega_s=v\end{cases}. \qquad (40)\end{aligned}$$

APPENDIX F
CASES CORRESPONDING TO $\mathbb{E}\{s_u(i)s_a(i)s_b(i)y_n(i)\}$ $(c_1=0,c_2=c_3=1)$

These terms can easily be computed using the results in section E.

a) Term $\mathbb{E}\{z_u(i)z_a(i)z_b(i)y_n(i)\}$*:* $u=1,\dots,N|\mathcal{D}_n|$, , $a,b=1,\dots,N|\mathcal{D}_n|$

We adapt relations (31)–(32), thus obtaining:

$$\begin{aligned}\mathbb{E}\{z_u(i)z_a(i)z_b(i)y_n(i)\}&=\mathbb{E}\Bigg\{\frac{(y_{m_1}(i)-y_{m_1}(\omega_p))}{\sigma^2}\frac{(y_{m_2}(i)-y_{m_2}(\omega_q))}{\sigma^2}\frac{(y_{m_3}(i)-y_{m_3}(\omega_r))}{\sigma^2}y_n(i)\\&\quad\times\exp\left(-\frac{1}{2\sigma^2}\left(\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_r)\|^2\right)\right)\Bigg\} \qquad (41)\\&=\frac{1}{\sigma^6}\nu(\{h_i\}_{i=1}^8),\text{ with }\begin{cases}h_1=h_2=m_1=\left\lceil\frac{u}{|\mathcal{D}_n|}\right\rceil\\h_3=h_4-N=m_2=\left\lceil\frac{a}{|\mathcal{D}_n|}\right\rceil\\h_5=h_6-2N=m_3=\left\lceil\frac{b}{|\mathcal{D}_n|}\right\rceil\\h_7=h_8-3N=m_4=n\end{cases},\begin{cases}\omega_p=\mathrm{mod}(u-1,|\mathcal{D}_n|)+1\\\omega_q=\mathrm{mod}(a-1,|\mathcal{D}_n|)+1\\\omega_r=\mathrm{mod}(b-1,|\mathcal{D}_n|)+1\\\omega_s=\bullet\end{cases}. \qquad (42)\end{aligned}$$

b) Term $\mathbb{E}\{z_u(i)k_a(i)z_b(i)y_n(i)\}$*:* $u=1,\dots,N|\mathcal{D}_n|$, $a=1,\dots,|\mathcal{D}_n|$, $b=1,\dots,N|\mathcal{D}_n|$

We adapt relations (33)–(34), thus obtaining:

$$\begin{aligned}\mathbb{E}\{z_u(i)k_a(i)z_b(i)y_n(i)\}&=\mathbb{E}\Bigg\{\frac{(y_{m_1}(i)-y_{m_1}(\omega_p))}{\sigma^2}\frac{(y_{m_3}(i)-y_{m_3}(\omega_r))}{\sigma^2}y_n(i)\\&\quad\times\exp\left(-\frac{1}{2\sigma^2}\left(\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_r)\|^2\right)\right)\Bigg\} \qquad (43)\\&=\frac{1}{\sigma^4}\nu(\{h_i\}_{i=1}^8),\text{ with }\begin{cases}h_1=h_2=m_1=\left\lceil\frac{u}{|\mathcal{D}_n|}\right\rceil\\h_3=h_4-N=m_2=\bullet\\h_5=h_6-2N=m_3=\left\lceil\frac{b}{|\mathcal{D}_n|}\right\rceil\\h_7=h_8-3N=m_4=n\end{cases},\begin{cases}\omega_p=\mathrm{mod}(u-1,|\mathcal{D}_n|)+1\\\omega_q=a\\\omega_r=\mathrm{mod}(b-1,|\mathcal{D}_n|)+1\\\omega_s=\bullet\end{cases}. \qquad (44)\end{aligned}$$

c) Term $\mathbb{E}\{z_u(i)k_a(i)k_b(i)y_n(i)\}$*:* $u=1,\dots,N|\mathcal{D}_n|$, $a,b=1,\dots,|\mathcal{D}_n|$

We adapt relations (35)–(36), thus obtaining:

$$\begin{aligned}\mathbb{E}\{z_u(i)k_a(i)k_b(i)y_n(i)\}&=\mathbb{E}\Bigg\{\frac{(y_{m_1}(i)-y_{m_1}(\omega_p))}{\sigma^2}y_n(i)\\&\quad\times\exp\left(-\frac{1}{2\sigma^2}\left(\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_p)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_q)\|^2+\|\boldsymbol{y}(i)-\boldsymbol{y}(\omega_r)\|^2\right)\right)\Bigg\} \qquad (45)\\&=\frac{1}{\sigma^2}\nu(\{h_i\}_{i=1}^8),\text{ with }\begin{cases}h_1=h_2=m_1=\left\lceil\frac{u}{|\mathcal{D}_n|}\right\rceil\\h_3=h_4-N=m_2=\bullet\\h_5=h_6-2N=m_3=\bullet\\h_7=h_8-3N=m_4=n\end{cases},\begin{cases}\omega_p=\mathrm{mod}(u-1,|\mathcal{D}_n|)+1\\\omega_q=a\\\omega_r=b\\\omega_s=\bullet\end{cases}. \qquad (46)\end{aligned}$$

d) Term $\mathbb{E}\{k_u(i)k_a(i)k_b(i)y_n(i)\}$*:* $u = 1, \ldots, |\mathcal{D}_n|$, $a, b = 1, \ldots, |\mathcal{D}_n|$

We adapt relations (39)–(40), thus obtaining:

$$\mathbb{E}\{k_u(i)k_a(i)k_b(i)k_v(i)\} = y_n(i) \exp\left(-\frac{1}{2\sigma^2}\left(\|\boldsymbol{y}(i) - \boldsymbol{y}(\omega_p)\|^2 + \|\boldsymbol{y}(i) - \boldsymbol{y}(\omega_q)\|^2 + \|\boldsymbol{y}(i) - \boldsymbol{y}(\omega_r)\|^2\right)\right)\Bigg\} \tag{47}$$

$$= \nu(\{h_i\}_{i=1}^8), \text{ with } \begin{cases} h_1 = h_2 = m_1 = \bullet \\ h_3 = h_4 - N = m_2 = \bullet \\ h_5 = h_6 - 2N = m_3 = \bullet \\ h_7 = h_8 - 3N = m_4 = n \end{cases}, \begin{cases} \omega_p = u \\ \omega_q = a \\ \omega_r = b \\ \omega_s = \bullet \end{cases}. \tag{48}$$

APPENDIX G
CASES CORRESPONDING TO $\mathbb{E}\{s_a(i)s_b(i)y_n^2(i)\}$ $(c_1 = c_2 = 0, c_3 = 1)$

These terms can easily be computed using the results in section F.

a) Term $\mathbb{E}\{y_n(i)z_a(i)z_b(i)y_n(i)\}$*:* , $a, b = 1, \ldots, N|\mathcal{D}_n|$

We adapt relations (41)–(42), thus obtaining:

$$\mathbb{E}\{y_n(i)z_a(i)z_b(i)y_n(i)\} = \mathbb{E}\Bigg\{y_n(i)\frac{(y_{m_2}(i) - y_{m_2}(\omega_q))}{\sigma^2}\frac{(y_{m_3}(i) - y_{m_3}(\omega_r))}{\sigma^2}y_n(i)$$
$$\times \exp\left(-\frac{1}{2\sigma^2}\left(\|\boldsymbol{y}(i) - \boldsymbol{y}(\omega_q)\|^2 + \|\boldsymbol{y}(i) - \boldsymbol{y}(\omega_r)\|^2\right)\right)\Bigg\} \tag{49}$$

$$= \frac{1}{\sigma^4}\nu(\{h_i\}_{i=1}^8), \text{ with } \begin{cases} h_1 = h_2 = m_1 = n \\ h_3 = h_4 - N = m_2 = \left\lceil \frac{a}{|\mathcal{D}_n|} \right\rceil \\ h_5 = h_6 - 2N = m_3 = \left\lceil \frac{b}{|\mathcal{D}_n|} \right\rceil \\ h_7 = h_8 - 3N = m_4 = n \end{cases}, \begin{cases} \omega_p = \bullet \\ \omega_q = \mathrm{mod}(a-1, |\mathcal{D}_n|) + 1 \\ \omega_r = \mathrm{mod}(b-1, |\mathcal{D}_n|) + 1 \\ \omega_s = \bullet \end{cases}. \tag{50}$$

b) Term $\mathbb{E}\{y_n(i)k_a(i)z_b(i)y_n(i)\}$*:* $a = 1, \ldots, |\mathcal{D}_n|$, $b = 1, \ldots, N|\mathcal{D}_n|$

We adapt relations (43)–(44), thus obtaining:

$$\mathbb{E}\{y_n(i)k_a(i)z_b(i)y_n(i)\} = \mathbb{E}\Bigg\{y_n(i)\frac{(y_{m_3}(i) - y_{m_3}(\omega_r))}{\sigma^2}y_n(i)\exp\left(-\frac{1}{2\sigma^2}\left(\|\boldsymbol{y}(i) - \boldsymbol{y}(\omega_q)\|^2 + \|\boldsymbol{y}(i) - \boldsymbol{y}(\omega_r)\|^2\right)\right)\Bigg\} \tag{51}$$

$$= \frac{1}{\sigma^2}\nu(\{h_i\}_{i=1}^8), \text{ with } \begin{cases} h_1 = h_2 = m_1 = n \\ h_3 = h_4 - N = m_2 = \bullet \\ h_5 = h_6 - 2N = m_3 = \left\lceil \frac{b}{|\mathcal{D}_n|} \right\rceil \\ h_7 = h_8 - 3N = m_4 = n \end{cases}, \begin{cases} \omega_p = \bullet \\ \omega_q = a \\ \omega_r = \mathrm{mod}(b-1, |\mathcal{D}_n|) + 1 \\ \omega_s = \bullet \end{cases}. \tag{52}$$

c) Term $\mathbb{E}\{y_n(i)k_a(i)k_b(i)y_n(i)\}$*:* $a, b = 1, \ldots, |\mathcal{D}_n|$

We adapt relations (45)–(46), thus obtaining:

$$\mathbb{E}\{y_n(i)k_a(i)k_a(i)y_n(i)\} = \mathbb{E}\Bigg\{y_n(i)y_n(i)\exp\left(-\frac{1}{2\sigma^2}\left(\|\boldsymbol{y}(i) - \boldsymbol{y}(\omega_q)\|^2 + \|\boldsymbol{y}(i) - \boldsymbol{y}(\omega_r)\|^2\right)\right)\Bigg\} \tag{53}$$

$$= \nu(\{h_i\}_{i=1}^8), \text{ with } \begin{cases} h_1 = h_2 = m_1 = n \\ h_3 = h_4 - N = m_2 = \bullet \\ h_5 = h_6 - 2N = m_3 = \bullet \\ h_7 = h_8 - 3N = m_4 = n \end{cases}, \begin{cases} \omega_p = \bullet \\ \omega_q = a \\ \omega_r = b \\ \omega_s = \bullet \end{cases}. \tag{54}$$

APPENDIX H
CASES CORRESPONDING TO $\mathbb{E}\{s_b(i)y_n(i)\}$ $(c_1 = c_2 = c_3 = 0)$

These terms can easily be computed using the results in section G.

a) Term $\mathbb{E}\{z_b(i)y_n(i)\}$*:* , $b = 1, \ldots, N|\mathcal{D}_n|$
We adapt relations (49)–(50), thus obtaining:

$$\mathbb{E}\{z_b(i)y_n(i)\} = \mathbb{E}\left\{ \frac{(y_{m_3}(i) - y_{m_3}(\omega_r))}{\sigma^2} y_n(i) \exp\left(-\frac{1}{2\sigma^2} \left(\|\boldsymbol{y}(i) - \boldsymbol{y}(\omega_r)\|^2 \right) \right) \right\} \tag{55}$$

$$= \frac{1}{\sigma^2} \nu(\{h_i\}_{i=1}^8), \text{ with } \begin{cases} h_1 = h_2 = m_1 = \bullet \\ h_3 = h_4 - N = m_2 = \bullet \\ h_5 = h_6 - 2N = m_3 = \left\lceil \frac{y}{|\mathcal{D}_n|} \right\rceil \\ h_7 = h_8 - 3N = m_4 = n \end{cases}, \begin{cases} \omega_p = \bullet \\ \omega_q = \bullet \\ \omega_r = \mathrm{mod}(b-1, |\mathcal{D}_n|) + 1 \\ \omega_s = \bullet \end{cases} . \tag{56}$$

b) Term $\mathbb{E}\{k_y(i)y_n(i)\}$*:* , $b = 1, \ldots, |\mathcal{D}_n|$
We adapt relations (53)–(54), thus obtaining:

$$\mathbb{E}\{k_b(i)y_n(i)\} = \mathbb{E}\left\{ y_n(i) \exp\left(-\frac{1}{2\sigma^2} \left(\|\boldsymbol{y}(i) - \boldsymbol{y}(\omega_r)\|^2 \right) \right) \right\} \tag{57}$$

$$= \nu(\{h_i\}_{i=1}^8), \text{ with } \begin{cases} h_1 = h_2 = m_1 = \bullet \\ h_3 = h_4 - N = m_2 = \bullet \\ h_5 = h_6 - 2N = m_3 = \bullet \\ h_7 = h_8 - 3N = m_4 = n \end{cases}, \begin{cases} \omega_p = \bullet \\ \omega_q = \bullet \\ \omega_r = b \\ \omega_s = \bullet \end{cases} . \tag{58}$$