Parameter-Invariant Detection of Unknown Inputs in Networked Systems

James Weimer, Damiano Varagnolo, Miloš S. Stanković, Karl Henrik Johansson

***Abstract*— This work considers the problem of detecting unknown inputs in networked systems whose dynamics are governed by time-varying unknown parameters. We propose a strategy in opposition to the commonly employed approach of first estimating the unknown parameters and then using the estimates as the true parameter values for detection, e.g. maximum-likelihood approaches. The suggested detection scheme employs test statistics that are invariant to the unknown parameters and do not rely on parameter estimation. We specifically consider the case of severe lack of prior knowledge, i.e., the problem of detecting unknown inputs when nothing is known of the system but some primitive structural properties, namely that the system is a linear network, subject to Gaussian noise, and that a certain input signal is either present or not. The aim is thus to analyze the structure and performances of invariant tests in a limiting case, specifically where the amount of prior information is minimal. The developed test is proven to be maximally invariant to the unknown parameters and Uniformly Most Powerful Invariant (UMPI). Simulation results indicate that for arbitrary networked systems the parameter-invariant detector achieves a specified probability of false alarm while ensuring that the probability of detection is maximized.**

***Index Terms*— hypothesis testing, invariant tests, linear systems, time varying systems, networked systems**

I. INTRODUCTION

Many applications, including environmental monitoring [1], building automation [2], wireless communications [3] and power grids [4], have networked dynamics that exploit multitudes of sensors and actuators. As the number of devices increases, so do the possibility of faults. When undetected, these faults can lead to several flavors of detriments: from mild inconveniences in HVAC systems (poor air quality) to disruptive ripple effects in power systems (extended blackouts).

While fault detection algorithms undoubtedly benefit from the knowledge of accurate models, these models are often parametrized by unknown time-varying environmental variables. For instance, in power transmission networks the resistance of the transmission lines varies with temperature and icing conditions, while in building automation the heat transfer between air-masses varies with the humidity, external temperature, and the opening or closing of windows and doors. Under these uncertain conditions, it is common to perform fault detection cascading parameter-estimation algorithms with hypothesis testing ones. These maximum likelihood approaches can yield significantly varying results, based on the accuracy of the estimated parameters, see, e.g., [5, Example 1, page 46].

An alternative approach is to design hypothesis tests that are *invariant to the unknown model parameters*. The benefit of invariant testing approaches comes in that the detector can be designed to have a specified performance independent of the unknown time-varying parameters [6]. It is then natural to ask how much the lack of prior knowledge affects the performance of these invariant tests? This question motivates this work, that queries the limits of invariant testing techniques when the problem is to detect the presence of unknown Gaussian inputs (or faults) in linear time-varying networked systems governed by unknown parameters. More specifically, we propose a maximally-invariant detector when *all* the model parameters are unknown that maximizes the probability of correct detection for a specified probability of false alarm.

Literature review: we start by noticing that classical methods for fault detection in the presence of unknown model parameters exploit Generalized Likelihood Ratio (GLR) strategies [7], summarized in: obtain the Maximum Likelihood (ML) estimates of the parameters under the each hypothesis, then test the likelihood ratio of these estimates. ML solutions are common in almost all problems containing unknown parameters (e.g. indirect adaptive control [8], blind identification and equalization of communication channels [9], machine learning [10], and in fault detection and identification [11]). The primary drawback of ML approaches arises when the parameter estimates converge to a value other than the true parameter value, as a direct consequence of the parameters varying with time and/or the input signal not being suitable for parameter estimation (or system identification) [8]. As a corollary, and as noted in [12], Maximally Invariant (MI) tests also outperform GLR approaches when small data sets are available. However, when the parameter estimation is unbiased, and as the Signal to Noise Ratio (SNR) tends to infinity (e.g., when the number of measurements approaches infinity, see [13]), the performance of the GLR and UMPI strategies are asymptotically equivalent.

The desire to detect faults for arbitrary inputs in systems with unknown parameters motivates the abundant space reserved to invariant strategies in classical textbooks, e.g., [6,

J. Weimer is with the Department of Computer and Information Science, School of Engineering and Applied Sciences, University of Pennsylvania, Philadelphia, PA, USA. Email: `weimerj @seas.upenn.edu`.

D. Varagnolo and K. Johansson are with the ACCESS Linnaeus Centre, School of Electrical Engineering, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. Emails: `{damiano, kallej } @kth.se`. The research leading to these results has received funding from the European Union Seventh Framework Programme [FP7/2007-2013] under grant agreement n°257462 HYCON2 Network of Excellence. The research was also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

M. Stanković is with the Innovation Center, School of Electrical Engineering, University of Belgrade, Serbia and is supported by the EU Marie Curie CIG and Serbian Ministry of Science and Technology. Email: `milsta @kth.se`.

Sec. 4.8], [14, Chap. 6], and the usage in several applications as the detection of structural changes in linear regression models [15], or in spectral properties of disturbances [16].

The parameter-invariant literature focuses mainly in finding invariant methods in known linear models with unknown or partially known covariance matrices [17], [18], [19], [20], [21], and do not consider linear models with unknown time-varying dynamics driven by noise. However, all parameter-invariant efforts generally focus on identifying tests that exploit maximally invariant statistics and establish conditions that ensure Constant False Alarm Rate (CFAR) properties.

Statement of contributions: In this work, and beyond the previous work, we assume the knowledge of *just* the fact that the system dynamics are networked, linear with Gaussian driving noises plus a weak knowledge on the structure of the fault. Under these assumptions, we develop a test that is maximally invariant to the unknown system parameters and show the test to be Uniformly Most Powerful Invariant (UMPI) and have a CFAR property. To the best of the authors' knowledge, the developed test herein is the one requiring the *smallest* amount of prior information among all the detectors proposed in literature for time-varying networked systems. Characterization of the detector thus identifies the best achievable performances when the amount of prior information is minimal.

Structure of the paper: Section II reports the needed basic results and definitions for invariant hypothesis testing. Section III mathematically formulates the hypothesis testing problem. We propose our testing technique along with its statistical characterization in Section IV. Section V numerically compares the performance of the proposed detector with the performance of strategies endowed with more prior information and no prior information for different operating points and systems. Finally Section VI reports some concluding remarks and proposes future extensions. For ease of readability all the proofs are collected in the appendix.

II. Notation and Preliminaries

In this section, and commiserate with [6], we introduce the notation, definitions, and methodology employed in designing UMPI tests.

A. Notation

In this subsection, we illustrate the various variable notations using varying fonts and capitalization of the letter z:

- plain upper case italic fonts $\rightarrow$ constant, Z;
- plain lower case italic fonts $\rightarrow$ scalar (or function with scalar range), z;
- bold lower case italic fonts $\rightarrow$ vector (or function with vectorial range), $\boldsymbol{z}$;
- bold lower case italic fonts with overhead vector $\rightarrow$ vector of concatenated vectors, $\vec{\boldsymbol{z}}$;
- bold upper case italic fonts $\rightarrow$ matrix, $\boldsymbol{Z}$.
- bold upper case italic fonts with overhead vector $\rightarrow$ matrix of concatenated matrices, $\vec{\boldsymbol{Z}}$.

For vectors we write $\boldsymbol{z}_i$ to denote the i-th position of $\boldsymbol{z}$. Similarly, For vectors of vectors we write $\vec{\boldsymbol{z}}_i$ to denote the i-th sub-vector. Lastly, for matrices we write $\boldsymbol{Z}_i$ to be the i-th column of $\boldsymbol{Z}$. We also use $\otimes$ to denote Kronecker products, I_N to be the identity matrix of dimension[1] N, $\boldsymbol{0}$ and $\boldsymbol{1}$ to be a vector of all zeros and all ones, respectively, and $\boldsymbol{e}_j$ to denote the elementary vector consisting of all zeros with a single unit entry in the j-th position. For arbitrary $\boldsymbol{Z}$, we define the following matrices:

$$\begin{aligned}
\boldsymbol{P_Z} &:= \boldsymbol{Z}\left(\boldsymbol{Z}^\top \boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^\top \\
\boldsymbol{P_Z^\perp} &:= \boldsymbol{I} - \boldsymbol{P_Z} \\
\boldsymbol{U_Z} &:= \left\{\boldsymbol{U} \mid \boldsymbol{U}\boldsymbol{U}^\top = \boldsymbol{P_Z},\, \boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}\right\} \\
\boldsymbol{U_Z^\perp} &:= \left\{\boldsymbol{U} \mid \boldsymbol{U}\boldsymbol{U}^\top = \boldsymbol{P_Z^\perp},\, \boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{I}\right\}
\end{aligned} \tag{1}$$

Additionally, we define the following matrix sets:

$$\begin{aligned}
\mathcal{D} &:= \{\boldsymbol{Z} \mid \boldsymbol{Z}_{ij} = 0,\ \forall i \neq j\} \\
\mathcal{D}_+ &:= \{\boldsymbol{Z} \ \mid\ \boldsymbol{Z} \in \mathcal{D},\ \ \boldsymbol{Z}_{ij} > 0,\ \ \forall i \neq j\} \\
\mathcal{L}_{\boldsymbol{z}} &:= \left\{\boldsymbol{I} - \boldsymbol{Z}\boldsymbol{L} \ \mid\ \boldsymbol{Z} \in \mathcal{D},\ \ \boldsymbol{Z}_{jj} = \boldsymbol{z}_j,\ \ \boldsymbol{1}^\top\boldsymbol{L} = \boldsymbol{0}\right\}.
\end{aligned}$$

We note that $\mathcal{D}$ corresponds to the set of all diagonal matrices, while $\mathcal{D}_+$ represents all positive definite diagonal matrices. $\mathcal{L}_{\boldsymbol{z}}$ is the set of all matrices having 1 as a left eigenvalue. Lastly, we employ the notation $\Pr[x|y]$ and $\mathbb{E}[x|y]$ to denote the probability of x given y and the expected value of x given y, respectively, where x and y are random variables.

B. Hypothesis Testing Preliminaries

Let $\boldsymbol{y}$ be a r.v. with probability density $f(\boldsymbol{y}\,; \boldsymbol{d}, \boldsymbol{\delta})$ parametrized in $\boldsymbol{d}$, $\boldsymbol{\delta}$. We define $\boldsymbol{d}$ to be the set of test parameters, and $\boldsymbol{\delta}$ to be the set of nuisance parameters, which induce a *transformation group* $\mathcal{G}$, i.e., a set of endomorphisms $\boldsymbol{g}$ on the space of the realizations $\boldsymbol{y}$ [6, Sec. 4.8]. This group of transformations partitions the measurement space into equivalence classes (or orbits) where points are considered equal if there exist $\boldsymbol{g}, \boldsymbol{g}' \in G$ mapping the first into the second and vice versa.

Definition 1 (***Maximally Invariant Statistic*** **[6]):** A statistic $t[\boldsymbol{y}]$ is said to be maximally invariant w.r.t. a transformation group $\mathcal{G}$ if it is:

invariant: $t[\boldsymbol{g}(\boldsymbol{y})] = t[\boldsymbol{y}], \quad \forall \boldsymbol{g} \in \mathcal{G}$

maximal: $t[\boldsymbol{y}] = t[\acute{\boldsymbol{y}}] \Longrightarrow \acute{\boldsymbol{y}} = \boldsymbol{g}(\boldsymbol{y}), \quad \exists\, \boldsymbol{g} \in \mathcal{G}$

A statistical test ϕ based on an invariant statistic can be said to be an invariant test:

Definition 2 (***Invariant Test*** **[6, Sec. 4.8]):** Let $\mathcal{G}$ be a transformation group, $t[\boldsymbol{y}]$ a statistic, and $\phi(\cdot)$ a hypothesis test. ϕ is said to be invariant w.r.t. $\mathcal{G}$ if

$$\phi\big(t[\boldsymbol{g}(\boldsymbol{y})]\big) = \phi\big(t[\boldsymbol{y}]\big) \tag{2}$$

[1] The subscript is omitted when the dimension is implicit.

for every $\boldsymbol{g} \in \mathcal{G}$.

The statistical performance of an invariant test ϕ is measured in terms of its *size* and *power* (see Definition 3). Invariant tests are desired to be Uniformly Most Powerful Invariant (UMPI):

Definition 3 (*Uniformly Most Powerful Invariant (UMPI) Test* [6, Sec. 4.8]): Let $\mathcal{G}$ be a transformation group corresponding to $\boldsymbol{\delta}$, $t[\boldsymbol{y}]$ a statistic and $\phi(\cdot)$ a test for deciding between $H_0 : \boldsymbol{d} = \boldsymbol{d}_0$ and $H_1 : \boldsymbol{d} = \boldsymbol{d}_1$ that is invariant w.r.t. $\mathcal{G}$. Then $\phi\big(t[\boldsymbol{y}]\big)$ is said to be an *uniformly most powerful invariant* (UMPI) test of size α if for every competing invariant test $\phi'\big(t[\boldsymbol{y}]\big)$ it holds that

$$\text{(size)} \quad \begin{aligned} &\Pr\Big[\phi\big(t[\boldsymbol{y}]\big) = H_1 \;\Big|\; \boldsymbol{d}_0, \boldsymbol{\delta}\Big] = \alpha; \\ &\Pr\Big[\phi'\big(t[\boldsymbol{y}]\big) = H_1 \;\Big|\; \boldsymbol{d}_0, \boldsymbol{\delta}\Big] \leq \alpha; \end{aligned}$$

$$\text{(power)} \quad \begin{aligned} &\Pr\Big[\phi\big(t[\boldsymbol{y}]\big) = H_1 \;\Big|\; \boldsymbol{d}_1, \boldsymbol{\delta}\Big] \geq \\ &\qquad \Pr\Big[\phi'\big(t[\boldsymbol{y}]\big) = H_1 \;\Big|\; \boldsymbol{d}_1, \boldsymbol{\delta}\Big]. \end{aligned}$$

As a remark, thanks to the Karlin-Rubin theorem [6, Sec. 4.7, page 124], a scalar maximally invariant statistic whose likelihood ratio is monotone can be used to construct an UMPI test.

III. Problem formulation

Consider the networked discrete-time linear dynamics

$$\begin{aligned} \boldsymbol{x}(k+1) &= \boldsymbol{A}(k)\boldsymbol{x}(k) + \boldsymbol{B}\boldsymbol{d}(k) + \boldsymbol{w}(k) \\ \boldsymbol{y}(k) &= \boldsymbol{x}(k) + \boldsymbol{v}(k) \end{aligned} \tag{3}$$

where

- $\boldsymbol{x}(k) \in \mathbb{R}^M$ denotes the state of the M-node network;
- $\boldsymbol{A}(k) \in \mathcal{L}_{\boldsymbol{m}}$ is the time-varying network dynamics between the M nodes, assuming a time-invariant network weighting vector, $\boldsymbol{m}$ [2];
- $\boldsymbol{B} \in \mathcal{D}$ represents the time-invariant input matrix;
- $\boldsymbol{y}(k), \boldsymbol{d}(k) \in \mathbb{R}^M$ are the node measurements and node inputs, respectively;
- $\boldsymbol{w}(k), \boldsymbol{v}(k) \in \mathbb{R}^M$ are respectively the Gaussian process noise and Gaussian measurement noise, with moments:

$$\mathbb{E}\left[\boldsymbol{w}(k)\right] = \overline{\boldsymbol{w}}, \qquad \mathbb{E}\left[\boldsymbol{v}(k)\right] = \overline{\boldsymbol{v}}$$

$$\mathbb{E}\left[(\boldsymbol{w}(k) - \overline{\boldsymbol{w}})(\boldsymbol{w}(k') - \overline{\boldsymbol{w}})^\top\right] = \begin{cases} \boldsymbol{\Lambda}(k) & k' = k \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}\left[(\boldsymbol{v}(k) - \overline{\boldsymbol{v}})(\boldsymbol{v}(k') - \overline{\boldsymbol{v}})^\top\right] = \begin{cases} \boldsymbol{\Gamma}(k) & k' = k \\ 0 & \text{otherwise} \end{cases}$$

$$\mathbb{E}\left[(\boldsymbol{v}(k) - \overline{\boldsymbol{v}})(\boldsymbol{w}(k') - \overline{\boldsymbol{w}})^\top\right] = \begin{cases} \boldsymbol{\Omega}(k) & k' = k \\ 0 & \text{otherwise} \end{cases}$$

having the property [3]:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}(k+1) + \boldsymbol{\Lambda}(k) + \boldsymbol{\Gamma}(k) - \boldsymbol{\Omega}(k) - \boldsymbol{\Omega}^\top(k).$$

We assume that the node inputs are composed of a known test signal and an unknown nuisance signal, and wish to test for the existence of the test signal. More formally, let $[0, \ldots, T]$ be the time interval for which we have measurements [4]. We can then write the time-concatenated measurements as $\vec{\boldsymbol{y}} := \left[\boldsymbol{y}^\top(0), \ldots, \boldsymbol{y}^\top(T)\right]^\top$. Additionally, we define the class of input signals considered in this manuscript by writing the time-series concatenation of the j-th node input, $\vec{\boldsymbol{d}}_j := [\boldsymbol{d}_j(0), \ldots, \boldsymbol{d}_j(T)]^\top$, as

$$\vec{\boldsymbol{d}}_j = \vec{\boldsymbol{S}}_j \vec{\boldsymbol{\theta}}_j + \boldsymbol{\mu}_j \vec{\boldsymbol{u}}_j \tag{4}$$

where

- $\vec{\boldsymbol{u}}_j := \left[u_j(0), \ldots, u_j(T)\right]^\top$ is the test input signal;
- $\boldsymbol{\mu}_j \in \{0, 1\}$ is a constant test parameter;
- $\vec{\boldsymbol{S}}_j \in \mathbb{R}^{T \times N_j}$ is a nuisance signal subspace;
- $\vec{\boldsymbol{\theta}}_j \in \mathbb{R}^{N_j}$ is a constant nuisance parameter.

For the system and input described in (3) and (4), we classify the information as either *available* or *unavailable* for testing as follows:

Assumption 1 (*Available Information*):

- the time-series measurements $\boldsymbol{y}(0), \ldots, \boldsymbol{y}(T)$
- the test input signals $\vec{\boldsymbol{u}}_1, \ldots, \vec{\boldsymbol{u}}_M$;
- the nuisance subspaces $\vec{\boldsymbol{S}}_1, \ldots, \vec{\boldsymbol{S}}_M$;
- the network weighting vector, $\boldsymbol{m}$;

Assumption 2 (*Unavailable Information*): For all $k \in \{0, \ldots, T\}$,

- the matrices $\boldsymbol{A}(k)$ and $\boldsymbol{B}$;
- the noise moments $\overline{\boldsymbol{w}}, \overline{\boldsymbol{v}}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}(k), \boldsymbol{\Gamma}(k), \boldsymbol{\Omega}(k)$;
- the parameters $\vec{\boldsymbol{\theta}}_1, \ldots, \vec{\boldsymbol{\theta}}_M$ and $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_M$;
- the initial condition $\boldsymbol{x}(0)$;

Under these assumptions, our binary hypothesis testing problem is formulated as testing whether $\boldsymbol{\mu}_j = 0$ or $\boldsymbol{\mu}_j = 1$:

Assumption 3 (*Hypothesis Test*): $\boldsymbol{\mu}_j$ satisfies either one of the two following hypotheses:

$H_{j,0}$ (null hypothesis):	$\boldsymbol{\mu}_j = 0$
$H_{j,1}$ (alternative hypothesis):	$\boldsymbol{\mu}_j = 1$

In words, both hypotheses assume the actual node inputs, $\vec{\boldsymbol{d}}_j$ to be unknown, since $\boldsymbol{\theta}_j$ is unknown, but with a fixed

[2]This work holds for any time-varying linear system having dynamics with at least one known eigenvector and corresponding eigenvalue. This property is satisfied for systems having linear networked dynamics.

[3]In this work, we allow time-varying correlated process and measurement noise, such that the time-difference of the measurements, i.e. $\boldsymbol{y}(k) - \boldsymbol{y}(k-1)$, has a time invariant covariance. We note that this property is satisfied when the process and measurement noises are i.i.d..

[4]T is assumed to be even for notational simplicity.

and known functional structure. $H_{j,1}$ additionally assumes the presence of a test input $\vec{\boldsymbol{u}}_j$.

Our aim is the following: develop a test that considers a ***specific*** *node input $\ell \in \{1, \ldots, M\}$, and decides among the hypotheses $H_{\ell,0}$ vs. $H_{\ell,1}$ in Assumption 3 using only the information in Assumption 1 and, at the same time, being maximally invariant to the unavailable information in Assumption 2.*

More precisely, we aim to find a test that detects whether node ℓ has a fault independently of whether a fault exists at any other node $j \neq \ell$ (fault isolation) *and* maximizes the probability of detection (power) for any probability of false alarm (size), i.e., we require the detector to be UMPI. Formally, thus, we aim to solve the following:

Problem 1

1) find a statistic $t[\vec{\boldsymbol{y}}]$ that satisfies Definition 1 (maximal invariance) w.r.t. the transformation group induced by nuisances parameters in Assumption 2;
2) find a test $\phi\big(t[\vec{\boldsymbol{y}}]\big)$ that satisfies Definition 3 (UMPI test) w.r.t. to the class of tests based on the previously introduced maximal invariant statistic $t\big[\vec{\boldsymbol{y}}\big]$.

IV. The UMPI Test

Following the methodology introduced in Section II, in this section we solve Problem 1 and provide the primary contribution of this work. To solve Problem 1 requires identifying the group of transformations induced by the unavailable information in Assumption 2. To identify this group, we begin by defining the invertible matrix

$$\boldsymbol{C} := \begin{bmatrix} \boldsymbol{I}_M & & & \\ -\boldsymbol{I}_M & \boldsymbol{I}_M & & \\ & \ddots & \ddots & \\ & & -\boldsymbol{I}_M & \boldsymbol{I}_M \end{bmatrix} \in \{0,1\}^{TM \times TM}$$

and recall that any invertible mapping of the measurements preserves maximal invariance [6] such that the time-series measurements can be written as

$$\boldsymbol{C}\vec{\boldsymbol{y}} = (\boldsymbol{I}_{TM} + \boldsymbol{E}) \begin{bmatrix} \boldsymbol{x}(0) \\ \boldsymbol{d}(0) + \boldsymbol{w}(0) \\ \vdots \\ \boldsymbol{d}(T-1) + \boldsymbol{w}(T-1) \end{bmatrix} + \boldsymbol{C} \begin{bmatrix} \boldsymbol{v}(0) \\ \vdots \\ \boldsymbol{v}(T) \end{bmatrix}, \quad (5)$$

where

$$\boldsymbol{E} := \big[\, \boldsymbol{E}_{ij} \,\big]$$
$$\boldsymbol{E}_{ij} = \begin{cases} \Big(\boldsymbol{A}(i-1) - \boldsymbol{I}_M\Big) \prod_{t=j}^{i-2} \boldsymbol{A}(t) & \text{if } j > i > 1 \\ \boldsymbol{0} & \text{otherwise.} \end{cases}$$

We observe that

$$\boldsymbol{p} := \left[\frac{1}{\boldsymbol{m}_1} \; \ldots \; \frac{1}{\boldsymbol{m}_M}\right]^\top$$

is an eigenvector of $\boldsymbol{E}^\top$ corresponding to the zero eigenvalue, $\boldsymbol{E}^\top \boldsymbol{p} = \boldsymbol{0}$. The transformed measurements (5) can thus be written, for some $\boldsymbol{\epsilon_F} \in \mathbb{R}^{2M+1}$ and $\boldsymbol{\epsilon_p} \in \mathbb{R}^{T(M-1)}$, as

$$\boldsymbol{C}\vec{\boldsymbol{y}} = \left(\boldsymbol{I}_T \otimes \boldsymbol{U_p}\right)\left(\boldsymbol{F}_\ell \boldsymbol{\epsilon_F} + \boldsymbol{\mu}_\ell \boldsymbol{B}_{\ell\ell} \vec{\boldsymbol{u}}_\ell + \boldsymbol{n}\right) + \left(\boldsymbol{I}_T \otimes \boldsymbol{U}_{\boldsymbol{p}}^{\perp}\right)\boldsymbol{\epsilon_p} \quad (6)$$

where

$$\boldsymbol{F}_\ell := \big[\, \vec{\boldsymbol{u}}_1 \, \ldots \, \vec{\boldsymbol{u}}_{\ell-1}, \, \vec{\boldsymbol{u}}_{\ell+1} \, \ldots \, \vec{\boldsymbol{u}}_M, \, \vec{\boldsymbol{S}}_1 \, \ldots \, \vec{\boldsymbol{S}}_M, \, \boldsymbol{1}, \, \boldsymbol{e}_1 \big]$$

and $\boldsymbol{n}$ is a zero-mean Gaussian random variable having covariance

$$\mathbb{E}\left[\boldsymbol{n}\boldsymbol{n}^\top\right] = \begin{bmatrix} \frac{\boldsymbol{p}^\top \boldsymbol{\Gamma}(0)\boldsymbol{p}}{\boldsymbol{p}^\top \boldsymbol{p}} & \pi_0 & & & \\ \pi_0 & \sigma & \pi_1 & & \\ & \ddots & \ddots & \ddots & \\ & & \pi_{T-2} & \sigma & \pi_{T-1} \\ & & & \pi_{T-1} & \sigma \end{bmatrix}$$

with

$$\sigma = \frac{\boldsymbol{p}^\top \left(\boldsymbol{\Lambda}(k) + \boldsymbol{\Gamma}(k) - \boldsymbol{\Omega}(k) - \boldsymbol{\Omega}(\boldsymbol{k})^\top\right)\boldsymbol{p}}{\boldsymbol{p}^\top \boldsymbol{p}} = \frac{\boldsymbol{p}^\top \boldsymbol{\Sigma} \boldsymbol{p}}{\boldsymbol{p}^\top \boldsymbol{p}}$$
$$\pi_k = \frac{\boldsymbol{p}^\top \left(\boldsymbol{\Omega}(k) - \boldsymbol{\Gamma}(k)\right)\boldsymbol{p}}{\boldsymbol{p}^\top \boldsymbol{p}}.$$

The transformed measurements in (6) utilize nuisance parameters $\boldsymbol{\epsilon_F}, \boldsymbol{\epsilon_p}, \sigma, \pi_0, \ldots \pi_k$, each of which is a unique function of the unknown parameters in Assumption 2[5]. We thus define the group of transformations induced by the nuisance parameters in the following lemma.

Lemma 1 (*Nuisance Parameter Transformations*): The group of transformations induced by the nuisance parameters is

$$\mathcal{G}_\ell = \left\{ \boldsymbol{g} \;\middle|\; \begin{aligned} \boldsymbol{g}(\vec{\boldsymbol{y}}) = & \sigma\left(\boldsymbol{I}_T \otimes \boldsymbol{U_p}\right) \boldsymbol{G} \left(\boldsymbol{I}_T \otimes \boldsymbol{U}_{\boldsymbol{p}}^\top\right) \boldsymbol{C}\vec{\boldsymbol{y}} \\ & + \left(\boldsymbol{I}_T \otimes \boldsymbol{U_p}\right)\left(\boldsymbol{F}_\ell \boldsymbol{\epsilon_F} + \boldsymbol{\mu}_\ell \boldsymbol{B}_{\ell\ell} \vec{\boldsymbol{u}}_\ell\right) \\ & + \left(\boldsymbol{I}_T \otimes \boldsymbol{U}_{\boldsymbol{p}}^{\perp}\right)\boldsymbol{\epsilon_p} \end{aligned} \right\} \quad (7)$$

where for some $c, \boldsymbol{c}_0, \ldots, \boldsymbol{c}_{\frac{T}{2}-1} \in \mathbb{R}$

$$\boldsymbol{G} = \boldsymbol{I}_T + c\boldsymbol{e}_1 \boldsymbol{e}_1^\top + \sum_{i=0}^{\frac{T}{2}-1} \boldsymbol{c}_i \boldsymbol{e}_{2i} \boldsymbol{e}_{2i+1}^\top.$$

In words, the group of transformations induced on the time-series measurements, $\vec{\boldsymbol{y}}$, by the nuisance parameters in Assumption 2 is a composition of the following four transformations:

- an unstructured time-series bias, at each time step, in the null-space of $\boldsymbol{p}$, induced by the unknown time-varying networked dynamics, $\boldsymbol{A}(0), \ldots, \boldsymbol{A}(T)$, namely $\left(\boldsymbol{I}_T \otimes \boldsymbol{U}_{\boldsymbol{p}}^{\perp}\right)\boldsymbol{\epsilon_p}$;

[5] The explicit realization of $\boldsymbol{\epsilon_F}$ and $\boldsymbol{\epsilon_p}$ are omitted to lighten the notational complexity. The mere existence of these parameters suffices in this work since they are known to be functions of the underlying unknown nuisance parameters.

- a structured time-series bias, in the direction of $\boldsymbol{p}$, induced by the unknown node inputs, $\boldsymbol{d}_1, \ldots, \boldsymbol{d}_M$, namely $(\boldsymbol{I}_T \otimes \boldsymbol{U}_{\boldsymbol{p}}) \boldsymbol{F}_\ell \boldsymbol{\epsilon}_{\boldsymbol{F}}$;
- a structured time-series bias, in the direction of $\boldsymbol{p}$, induced by the unknown node input gain, $\boldsymbol{B}_{\ell\ell}$, in the direction of $\vec{\boldsymbol{u}}_\ell$, namely $\boldsymbol{\mu}_\ell (\boldsymbol{I}_T \otimes \boldsymbol{U}_{\boldsymbol{p}}) \boldsymbol{B}_{\ell\ell} \vec{\boldsymbol{u}}_\ell$;
- a noise-whitening scaling of the time-series measurements, in the direction of $\boldsymbol{p}$, induced by the unknown moments of the process noise and measurement noise, namely, $\sigma (\boldsymbol{I}_T \otimes \boldsymbol{U}_{\boldsymbol{p}}) \boldsymbol{G} \left(\boldsymbol{I}_T \otimes \boldsymbol{U}_{\boldsymbol{p}}^\top\right) \boldsymbol{C} \vec{\boldsymbol{y}}$.

We now introduce a statistic $t_\ell[\vec{\boldsymbol{y}}]$ that is maximally invariant for the group of transformations $\mathcal{G}_\ell$ introduced in Lemma 1:

Theorem 1 (***Maximally Invariant Statistic***)**:** A statistic that is maximally invariant to $\mathcal{G}_\ell$ and solves Problem 1-1 is

$$t_\ell[\vec{\boldsymbol{y}}] = \frac{\boldsymbol{r}_\ell^\top \boldsymbol{P}_{\boldsymbol{s}_\ell} \boldsymbol{r}_\ell}{\frac{1}{N_\ell - 1} \boldsymbol{r}_\ell^\top \boldsymbol{P}_{\boldsymbol{s}_\ell}^\perp \boldsymbol{r}_\ell} \tag{8}$$

with

$$\boldsymbol{r}_\ell = \boldsymbol{U}_{\boldsymbol{Q}\boldsymbol{F}_\ell}^\perp \boldsymbol{D}_{\boldsymbol{p}} \vec{\boldsymbol{y}}$$
$$\boldsymbol{s}_\ell = \boldsymbol{U}_{\boldsymbol{Q}\boldsymbol{F}_\ell}^\perp \boldsymbol{Q} \vec{\boldsymbol{u}}_\ell$$
$$N_\ell = \frac{T}{2} - \operatorname{rank}\left(\boldsymbol{U}_{\boldsymbol{Q}\boldsymbol{F}_\ell}^\perp\right)$$

assuming

$$\boldsymbol{p} := \begin{bmatrix} \frac{1}{\boldsymbol{m}_1} & \cdots & \frac{1}{\boldsymbol{m}_M} \end{bmatrix}^\top$$
$$\boldsymbol{D}_{\boldsymbol{p}} := \boldsymbol{I}_{\frac{T}{2}} \otimes \begin{bmatrix} -\boldsymbol{U}_{\boldsymbol{p}}^\top & \boldsymbol{U}_{\boldsymbol{p}}^\top \end{bmatrix}$$
$$\boldsymbol{Q} := \boldsymbol{I}_{\frac{T}{2}} \otimes \begin{bmatrix} 0 & 1 \end{bmatrix}$$

Proof: A proof is provided in the appendix. ■

We observe that the maximally invariant statistic in (8) can be equivalently written as a ratio of independent chi-square random variables. This particular ratio is known to follow an F-distribution, which is monotone. Thus, by applying the Karlin-Rubin theorem [6], the following corollary results:

Corollary 1 (***UMPI Test***)**:** A UMPI test w.r.t. $\mathcal{G}_\ell$ of size α that solves Problem 1-2 is

$$\phi(t_\ell[\vec{\boldsymbol{y}}]) = \begin{cases} H_{\ell,0} & \text{if } \; t_\ell[\vec{\boldsymbol{y}}] < \mathcal{F}_{1,N_\ell-1}^{-1}(\alpha) \\ H_{\ell,1} & \text{otherwise.} \end{cases} \tag{9}$$

where $\mathcal{F}_{n,m}^{-1}(\alpha)$ is the inverse central cumulative F-distribution of dimensions n and m.

V. Numerical Examples

We evaluate the test in (9) performing three Monte-Carlo characterization as follows:

1) we fixed a desired probability of false alarms α (0.01, 0.1 and 0.25);
2) we randomly generated 500 stable time-invariant networked systems having $M = 10$ nodes, like (3), as described in Table I (i.e., we discarded the unstable realizations). All unspecified terms are assumed to be zero.;
3) for each of the 500 systems (3) we generated exactly one realization $y_j(1), \ldots, y_j(500)$, $j = 1, \ldots, 10$;
4) for each $T = 1, \ldots, 500$ and each of the 500 systems (3) we executed the following three tests, all with the same desired probability of false alarms α:
   a) full information test: assume the perfect knowledge of the networked dynamics $\boldsymbol{A}$ and $\boldsymbol{B}$; the moments of the process and measurement noises $\overline{\boldsymbol{w}}$, $\overline{\boldsymbol{v}}$, $\boldsymbol{\Gamma}$, $\boldsymbol{\Lambda}$; the parameters $\boldsymbol{\theta}_j$; the initial conditions $\boldsymbol{x}_j(0)$ ($j = 1, \ldots, 10$). Then design the Uniformly Most Powerful (UMP) test for testing $H_{\ell,0}$ vs. $H_{\ell,1}$ given all this information;
   b) UMPI test: our test (9);
   c) no information test: perform a weighted coin flip s.t. the desired probability of false alarms α is met.

$\boldsymbol{A}_{ij}, \boldsymbol{B}_{jj} \sim \mathcal{U}[-0.5, 0.5]$	$\boldsymbol{m}_j \sim \mathcal{U}[1, 2]$
$\overline{\boldsymbol{w}}_j, \overline{\boldsymbol{v}}_j \sim \mathcal{N}(0, 1)$	$\boldsymbol{\Gamma}_{jj}, \boldsymbol{\Lambda}_{jj} \sim \mathcal{U}[0.1, 1]$

TABLE I

Random extraction mechanisms for the generation of the systems (3). $\mathcal{N}$ indicates Gaussian distributions, $\mathcal{U}$ uniform distributions. All the quantities are extracted independently.

The outcomes are then summarized in the following Figures 1, 2 and 3, that plot for each test and each T the average correct detection rate reached over the 500 considered realizations of system 3.

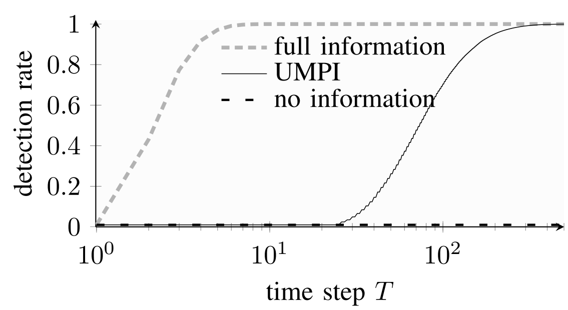

Fig. 1. Monte-Carlo characterization of the detection tests given $\alpha = 0.01$.

From the graphics we draw the following conclusions. All three tests, the full information UMP test, the UMPI test described in this work, and the random coin flip test all yield the same probability of false alarm (by design), but have varying probability of detection. Specifically, we note that the performance of the full-information UMP test is always better than the UMPI test and the coin flip. Before the number of measurements passes the threshold N_ℓ (independent of the chosen α) the UMPI test is equivalent to a coin flipping. This results from the fact that when there are few measurements, all possible measurements can be explained

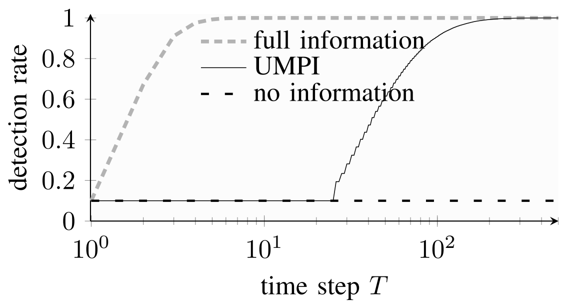

Fig. 2. Monte-Carlo characterization of the detection tests given $\alpha = 0.1$.

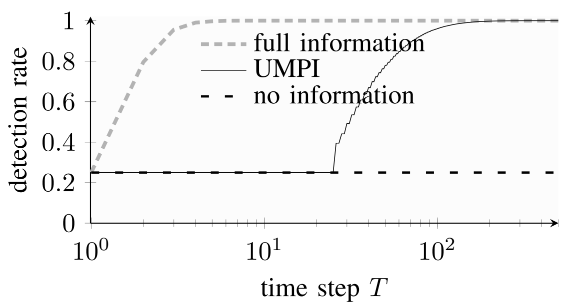

Fig. 3. Monte-Carlo characterization of the detection tests given $\alpha = 0.25$.

by the unknown parameters. Only after N_ℓ measurements is there enough information to begin testing better than random chance. Once the threshold is exceeded, the test starts increasing its correct detection rate, discerning better and better. Eventually it approaches the same performance of the full information-based test, i.e., the best one might desire (with different speeds, depending on the selected probability of false alarms). As expected, the convergence rate of the UMPI test to a high probability of detection decreases as the probability of false alarm decreases. Indicating that to achieve high detection rates with a low probability of false alarm requires, in general, more measurements.

VI. Discussion and Future Works

We considered a hypothesis testing problem defined over networked linear time-varying Gaussian systems, and then derived an Uniformly Most Powerful Invariant detector with Constant False Alarm Rate properties that is tailored for situations where the prior information available is little. Despite the high degree of uncertainty on the system, the offered testing strategy has some power, i.e., it is able to actually detect faults also when the number of measurements is limited.

Clearly the detector's performance, in terms of false positives / negatives rates, is worse than the performance of tests that exploit deeper knowledge of the system (see the numerical results provided in Section V). This accords with the intuition that one should always derive tests that exploit all the information available.

Nonetheless the considered strategy has the valuable property of providing a lower bound on the performance that can be achieved in absence of prior information on a broad class of networked systems. This claim derives from the fact that the derived test has two optimality properties: it is based on a maximally invariant statistic and it is uniformly most powerful. Paraphrasing, every other invariant fault detector defined over the same hypothesis testing problem will have *at best* the same performance of the here proposed strategy (again in terms of false positives / negatives rates). Moreover, to have the same performance, it *must* be essentially based on the same statistic considered here.

The offered strategy raises also an important and interesting research direction. Namely, how should the scheme be compared with strategies that initially estimate the parameters with system identification or Maximum Likelihood approaches and then in cascade perform classical and non-invariant tests. The mathematical problem is in fact to understand if there are conditions for which one of the various strategies is ensured to perform better or worse than the other ones, and why. This is undoubtedly useful in practical scenarios, where one always aim to exploit the best available detector.

References

[1] J. Weimer, B. Sinopoli, and B. Krogh, "An approach to leak detection using wireless sensor networks at carbon sequestration sites," *International Journal of Greenhouse Gas Control*, vol. 9, pp. 243–253, 2012.

[2] F. Oldewurtel, A. Parisio, C. N. Jones, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and M. Morari, "Use of model predictive control and weather forecasts for energy efficient building climate control," *Energy and Buildings*, vol. 45, no. 0, pp. 15 – 27, 2012.

[3] T. Arampatzis, J. Lygeros, and S. Manesis, "A survey of applications of wireless sensors and wireless sensor networks," in *IEEE International Symposium on Intelligent Control, MCCA*, 2005, pp. 719–724.

[4] S. Bolognani, A. Carron, A. Di Vittorio, D. Romeres, L. Schenato, and S. Zampieri, "Distributed multi-hop reactive power compensation in smart micro-grids subject to saturation constraints," in *51st IEEE Conference on Decision and Control*, 2012, pp. 785–790.

[5] V. N. Vapnik, *Statistical learning theory*. New York: Wiley, 1998.

[6] L. L. Scharf, *Statistical Signal Processing, Detection, Estimation, and Time Series Analysis*. Addison-Welsley Publishing Company Inc., Reading, Massachusetts, 1991.

[7] A. S. Willsky and H. L. Jones, "A generalized likelihood ratio approach to state estimation in linear systems subjects to abrupt changes," vol. 13, Nov. 1974, pp. 846–853.

[8] K. J. Astrom and B. Wittenmark, *Adaptive Control*, 2nd ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1994.

[9] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Signals, Systems and Computers, 1991. 1991 Conference Record of the Twenty-Fifth Asilomar Conference on*, Nov, pp. 856–860 vol.2.

[10] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[11] I. Hwang, S. Kim, Y. Kim, and C. Seah, "A survey of fault detection, isolation, and reconfiguration methods," *Control Systems Technology, IEEE Transactions on*, vol. 18, no. 3, pp. 636–653, May.

[12] H. Kim and A. Hero III, "Comparison of GLR and invariant detectors under structured clutter covariance," *Image Processing, IEEE Transactions on*, vol. 10, no. 10, pp. 1509–1520, 2001.

[13] J. Gabriel and S. Kay, "On the relationship between the GLRT and UMPI tests for the detection of signals with unknown parameters," *Signal Processing, IEEE Transactions on*, vol. 53, no. 11, pp. 4194–4203, 2005.

[14] E. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. Springer, 2005.

[15] M. Hušková, "Some invariant test procedures for detection of structural changes," *Kybernetika*, vol. 36, no. 4, pp. 401–414, 2000.

[16] N. Begum and M. L. King, "Most mean powerful invariant test for testing two-dimensional parameter spaces," *Journal of Statistical Planning and Inference*, vol. 134, no. 2, pp. 536 – 548, 2005.

[17] S. Bose and A. Steinhardt, "A maximal invariant framework for adaptive detection with structured and unstructured covariance matrices," *Signal Processing, IEEE Transactions on*, vol. 43, no. 9, pp. 2164–2175, 1995.

[18] K. Noda and H. Ono, "On UMP invariant F-test procedures in a general linear model," *Communications in Statistics-Theory and Methods*, vol. 30, no. 10, pp. 2099–2115, 2001.

[19] E. Conte, A. De Maio, and C. Galdi, "CFAR detection of multidimensional signals: an invariant approach," *Signal Processing, IEEE Transactions on*, vol. 51, no. 1, pp. 142–151, 2003.

[20] A. De Maio, "Rao test for adaptive detection in Gaussian interference with unknown covariance matrix," *Signal Processing, IEEE Transactions on*, vol. 55, no. 7, pp. 3577–3584, 2007.

[21] A. De Maio and E. Conte, "Adaptive detection in Gaussian interference with unknown covariance after reduction by invariance," *Signal Processing, IEEE Transactions on*, vol. 58, no. 6, pp. 2925–2934, 2010.

APPENDIX

This appendix provides a proof for Theorem 1.

Proof:

Invariance: Observing that

$$\boldsymbol{U}_{\boldsymbol{QF_\ell}}^{\perp}\boldsymbol{D_p}\boldsymbol{g}(\vec{\boldsymbol{y}}) = \sigma\boldsymbol{r}_\ell + \boldsymbol{\mu}_\ell\boldsymbol{B}_{\ell\ell}\boldsymbol{s}_\ell \tag{10}$$

then

$$t\big[\boldsymbol{g}(\vec{\boldsymbol{y}})\big] = \frac{\sigma^2\boldsymbol{r}_\ell^\top\boldsymbol{P}_{\boldsymbol{s}_\ell}\boldsymbol{r}_\ell}{\sigma^2\frac{1}{N_\ell-1}\boldsymbol{r}_\ell^\top\boldsymbol{P}_{\boldsymbol{s}_\ell}^{\perp}\boldsymbol{r}_\ell} = \frac{\boldsymbol{r}_\ell^\top\boldsymbol{P}_{\boldsymbol{s}_\ell}\boldsymbol{r}_\ell}{\frac{1}{N_\ell-1}\boldsymbol{r}_\ell^\top\boldsymbol{P}_{\boldsymbol{s}_\ell}^{\perp}\boldsymbol{r}_\ell} = t\big[\vec{\boldsymbol{y}}\big] \tag{11}$$

Maximality: Let

$$\begin{aligned}\boldsymbol{r}_\ell &= \boldsymbol{U}_{\boldsymbol{QF_\ell}}^{\perp}\boldsymbol{D_p}\vec{\boldsymbol{y}}\\ \acute{\boldsymbol{r}}_\ell &= \boldsymbol{U}_{\boldsymbol{QF_\ell}}^{\perp}\boldsymbol{D_p}\acute{\vec{\boldsymbol{y}}}\end{aligned} \tag{12}$$

and

$$\boldsymbol{H} := \left(\boldsymbol{I}_T \otimes \boldsymbol{U}_{\boldsymbol{p}}^\top\right)\boldsymbol{C} \tag{13}$$

then it holds that

$$\begin{aligned}
& t\left[\acute{\vec{\boldsymbol{y}}}\right] = t\left[\vec{\boldsymbol{y}}\right]\\
\longrightarrow & \frac{\boldsymbol{r}_\ell^\top\boldsymbol{P}_{\boldsymbol{s}_\ell}\boldsymbol{r}_\ell}{\frac{1}{N_\ell-1}\boldsymbol{r}_\ell^\top\boldsymbol{P}_{\boldsymbol{s}_\ell}^{\perp}\boldsymbol{r}_\ell} = \frac{\acute{\boldsymbol{r}}_\ell^\top\boldsymbol{P}_{\boldsymbol{s}_\ell}\acute{\boldsymbol{r}}}{\frac{1}{N_\ell-1}\acute{\boldsymbol{r}}_\ell^\top\boldsymbol{P}_{\boldsymbol{s}_\ell}^{\perp}\acute{\boldsymbol{r}}_\ell}\\
\longrightarrow & \acute{\boldsymbol{r}}^\top\left(\boldsymbol{P}_{\boldsymbol{s}_\ell} - \boldsymbol{I}\frac{\boldsymbol{r}_\ell^\top\boldsymbol{P}_{\boldsymbol{s}_\ell}\boldsymbol{r}_\ell}{\boldsymbol{r}_\ell^\top\boldsymbol{r}_\ell}\right)\acute{\boldsymbol{r}}_\ell = 0\\
\longrightarrow & \boldsymbol{s}_\ell^\top\acute{\boldsymbol{r}}_\ell = \sigma\boldsymbol{s}_\ell^\top\boldsymbol{r}_\ell \quad \exists\sigma\in\mathbb{R}\\
\longrightarrow & \acute{\boldsymbol{r}}_\ell = \sigma\boldsymbol{r}_\ell - \boldsymbol{P}_{\boldsymbol{s}_\ell}^{\perp}\left(\sigma\boldsymbol{r}_\ell - \acute{\boldsymbol{r}}_\ell\right)\\
\longrightarrow & \acute{\boldsymbol{r}}_\ell = \sigma\boldsymbol{r}_\ell + \boldsymbol{\mu}_\ell\boldsymbol{B}_{\ell\ell}\boldsymbol{s}_\ell, \quad \exists\boldsymbol{\mu}_\ell, \boldsymbol{B}_{\ell\ell}\in\mathbb{R}\\
\longrightarrow & \boldsymbol{D_p}\acute{\vec{\boldsymbol{y}}} = \sigma\boldsymbol{D_p}\vec{\boldsymbol{y}} - \boldsymbol{P}_{\boldsymbol{QF_\ell}}^{\perp}\left(\sigma\boldsymbol{D_p}\vec{\boldsymbol{y}} + \boldsymbol{\mu}_\ell\boldsymbol{B}_{\ell\ell}\boldsymbol{Q}\vec{\boldsymbol{u}}_\ell - \boldsymbol{D_p}\acute{\vec{\boldsymbol{y}}}\right)\\
\longrightarrow & \boldsymbol{D_p}\acute{\vec{\boldsymbol{y}}} = \sigma\boldsymbol{D_p}\vec{\boldsymbol{y}} + \boldsymbol{\mu}_\ell\boldsymbol{B}_{\ell\ell}\boldsymbol{Q}\vec{\boldsymbol{u}}_\ell + \boldsymbol{QF_\ell}\boldsymbol{\epsilon_F}, \quad \exists\boldsymbol{\epsilon_F}\in\mathbb{R}^{2M+1}\\
\longrightarrow & \boldsymbol{H}\acute{\vec{\boldsymbol{y}}} = \sigma\boldsymbol{H}\vec{\boldsymbol{y}} + \boldsymbol{\mu}_\ell\boldsymbol{B}_{\ell\ell}\boldsymbol{Q}\vec{\boldsymbol{u}}_\ell + \boldsymbol{QF_\ell}\boldsymbol{\epsilon_F}\\
& \quad - \boldsymbol{P}_{\boldsymbol{Q}}^{\perp}\left(\sigma\boldsymbol{H}\vec{\boldsymbol{y}} + \boldsymbol{\mu}_\ell\boldsymbol{B}_{\ell\ell}\vec{\boldsymbol{u}}_\ell + \boldsymbol{F_\ell}\boldsymbol{\epsilon_F} - \boldsymbol{H}\acute{\vec{\boldsymbol{y}}}\right)\\
\longrightarrow & \boldsymbol{C}\acute{\vec{\boldsymbol{y}}} = \sigma\left(\boldsymbol{I}_T\otimes\boldsymbol{U_p}\right)\boldsymbol{G}\boldsymbol{H}\vec{\boldsymbol{y}}\\
& \quad + \left(\boldsymbol{I}_T\otimes\boldsymbol{U_p}\right)\left(\boldsymbol{\mu}_\ell\boldsymbol{B}_{\ell\ell}\vec{\boldsymbol{u}}_\ell + \boldsymbol{F_\ell}\boldsymbol{\epsilon_F}\right)\\
& \quad + \left(\boldsymbol{I}_T\otimes\boldsymbol{U}_{\boldsymbol{p}}^{\perp}\right)\boldsymbol{\epsilon_p}, \quad \exists\boldsymbol{\epsilon_p}\in\mathbb{R}^{T(M-1)}\\
\longrightarrow & \acute{\vec{\boldsymbol{y}}} = \boldsymbol{g}\left(\vec{\boldsymbol{y}}\right), \quad \exists\boldsymbol{g}\in\mathcal{G}_\ell
\end{aligned} \tag{14}$$

∎